

# Journal of Medical Internet Research

Impact Factor (2018): 4.945 - ranked #1 medical informatics journal by Impact Factor  
Volume 22 (2020), Issue 10 ISSN: 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

## Contents

### Original Papers

Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study ( <a href="#">e22635</a> ) Daniel Low, Laurie Rumker, Tanya Talkar, John Torous, Guillermo Cecchi, Satrajit Ghosh. . . . .	12
Public Opinion About E-Cigarettes on Chinese Social Media: A Combined Study of Text Mining Analysis and Correspondence Analysis ( <a href="#">e19804</a> ) Di Wang, Joanne Lyu, Xiaoyu Zhao. . . . .	28
Googling for Ticks and Borreliosis in Germany: Nationwide Google Search Analysis From 2015 to 2018 ( <a href="#">e18581</a> ) Cora Scheerer, Melvin R��th, Linda Tizek, Martin K��berle, Tilo Biedermann, Alexander Zink. . . . .	41
Disaster eHealth: Scoping Review ( <a href="#">e18310</a> ) Samaneh Madanian, Tony Norris, Dave Parry. . . . .	194
Assessment of a Personal Interactive Carbon Monoxide Breath Sensor in People Who Smoke Cigarettes: Single-Arm Cohort Study ( <a href="#">e22811</a> ) Jennifer Marler, Craig Fujii, Kristine Wong, Joseph Galanko, Daniel Balbierz, David Utley. . . . .	289
Predictors and Effects of Usage of an Online Mindfulness Intervention for Distressed Cancer Patients: Usability Study ( <a href="#">e17526</a> ) Linda Cillessen, Monique van de Ven, F��lix Compen, Else Bisseling, Marije van der Lee, Anne Speckens. . . . .	313
Twelve-Month Follow-Up to a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention for Rural Adults With Depression Symptoms: Single-Arm Longitudinal Study ( <a href="#">e21336</a> ) Mark Schure, Bernadette McCrory, Kathryn Tuchscherer Franklin, John Greist, Ruth Weissman. . . . .	324
Mechanisms of Action of a Web-Based Intervention With Health Professional Support to Increase Adherence to Nebulizer Treatments in Adults With Cystic Fibrosis: Qualitative Interview Study ( <a href="#">e16782</a> ) Sarah Drabble, Alicia O'Cathain, Alexander Scott, Madelynne Arden, Samuel Keating, Marlene Hutchings, Chin Maguire, Martin Wildman. . . . .	3
Health-Related Quality of Life Improvements in Systemic Lupus Erythematosus Derived from a Digital Therapeutic Plus Tele-Health Coaching Intervention: Randomized Controlled Pilot Trial ( <a href="#">e23868</a> ) Faiz Khan, Nora Granville, Raja Malkani, Yash Chathampally. . . . .	349

Internet-Based Multimodal Pain Program With Telephone Support for Adults With Chronic Temporomandibular Disorder Pain: Randomized Controlled Pilot Trial ( <a href="#">e22326</a> )	
Julia Lam, Peter Svensson, Per Alstergren. . . . .	371
Identifying the Value of an eHealth Intervention Aimed at Cognitive Impairments: Observational Study in Different Contexts and Service Models ( <a href="#">e17720</a> )	
Monika Jurkeviciute, Lex van Velsen, Henrik Eriksson, Svante Lifvergren, Pietro Trimarchi, Ulla Andin, Johan Svensson. . . . .	388
Efficacy of Electronic Acupuncture Shoes for Chronic Low Back Pain: Double-Blinded Randomized Controlled Trial ( <a href="#">e22324</a> )	
Bo-Yan Yeh, Geng-Hao Liu, Tzung-Yan Lee, Alice Wong, Hen-Hong Chang, Yu-Sheng Chen. . . . .	402
24-Month Outcomes of Primary Care Web-Based Depression Prevention Intervention in Adolescents: Randomized Clinical Trial ( <a href="#">e16802</a> )	
Benjamin Van Voorhees, Tracy Gladstone, Kunmi Sobowale, C Brown, David Aaby, Daniela Terrizzi, Jason Canel, Eumene Ching, Anita Berry, James Cantorna, Milton Eder, William Beardslee, Marian Fitzgibbon, Monika Marko-Holguin, Linda Schiffer, Miae Lee, Sarah de Forest, Emily Sykes, Jennifer Suor, Theodore Crawford, Katie Burkhouse, Brady Goodwin, Carl Bell. . . . .	418
Effectiveness of a Peer-Led Web-Based Intervention to Improve General Self-Efficacy in Using Dating Apps Among Young Adults: Randomized Clustered Trial ( <a href="#">e16378</a> )	
William Wong, Wai Sun, Shu Chia, Joseph Tucker, William Mak, Lin Song, Kitty Choi, Stephanie Lau, Eric Wan. . . . .	433
Effects of a Mobile and Web App (Thought Spot) on Mental Health Help-Seeking Among College and University Students: Randomized Controlled Trial ( <a href="#">e20790</a> )	
David Wiljer, Jenny Shi, Brian Lo, Marcos Sanches, Elisa Hollenberg, Andrew Johnson, Alexxa Abi-Jaoudé, Gloria Chaim, Kristin Cleverley, Joanna Henderson, Wanrudee Isaranuwatjai, Andrea Levinson, Janine Robb, Howard Wong, Aristotle Voineskos. . . . .	449
The Association Between Electronic Device Use During Family Time and Family Well-Being: Population-Based Cross-Sectional Study ( <a href="#">e20529</a> )	
Sheng Zhao, Ningyuan Guo, Man Wang, Daniel Fong, Agnes Lai, Sophia Chan, Tai Lam, Daniel Ho. . . . .	462
Patterns of Use and Key Predictors for the Use of Wearable Health Care Devices by US Adults: Insights from a National Survey ( <a href="#">e22443</a> )	
Ranganathan Chandrasekaran, Vipanchi Katthula, Evangelos Moustakas. . . . .	473
Wearable Technology Acceptance in Health Care Based on National Culture Differences: Cross-Country Analysis Between Chinese and Swiss Consumers ( <a href="#">e18801</a> )	
Dong Yang Meier, Petra Barthelmess, Wei Sun, Florian Liberatore. . . . .	484
Brain Tumor Discussions on Twitter (#BTSM): Social Network Analysis ( <a href="#">e22005</a> )	
Josemari Feliciano, Liz Salmi, Charlie Blotner, Adam Hayden, Edjah Nduom, Bethany Kwan, Matthew Katz, Elizabeth Claus. . . . .	499
Experiences of Serving and Ex-Serving Members With the PTSD Coach Australia App: Mixed Methods Study ( <a href="#">e18447</a> )	
Jane Shakespeare-Finch, Karolina Alichniewicz, Esben Strodl, Kelly Brown, Catherine Quinn, Leanne Hides, Angela White, Gabriel Gossage, Loretta Poerio, Dimitri Batras, Samantha Jackson, Jess Styles, David Kavanagh. . . . .	507
COVID-19 and the "Film Your Hospital" Conspiracy Theory: Social Network Analysis of Twitter Data ( <a href="#">e22374</a> )	
Wasim Ahmed, Francesc López Seguí, Josep Vidal-Alaball, Matthew Katz. . . . .	517
Low Testosterone on Social Media: Application of Natural Language Processing to Understand Patients' Perceptions of Hypogonadism and Its Treatment ( <a href="#">e21383</a> )	
Vadim Osadchiy, Tommy Jiang, Jesse Mills, Sriram Eleswarapu. . . . .	525



## Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis ([e21597](#))

Nicolò Gozzi, Michele Tizzani, Michele Starnini, Fabio Ciulla, Daniela Paolotti, André Panisson, Nicola Perra. . . . . 541

## Quantifying Public Interest in Police Reforms by Mining Internet Search Data Following George Floyd's Death ([e22574](#))

John Ayers, Benjamin Althouse, Adam Poliak, Eric Leas, Alicia Nobles, Mark Dredze, Davey Smith. . . . . 559

## Public Interest in Acne on the Internet: Comparison of Search Information From Google Trends and Naver ([e19427](#))

Tae Park, Woo Kim, Suyeon Park, Jaeouk Ahn, Moon Cho, Sooyoung Kim. . . . . 570

## Recommendations From the Twitter Hashtag #DoctorsAreDickheads: Qualitative Analysis ([e17595](#))

Anjana Sharma, Ziva Mann, Roy Cherian, Jan Del Rosario, Janine Yang, Urmimala Sarkar. . . . . 580

## Development and Evaluation of a Digital Intervention for Fulfilling the Needs of Older Migrant Patients With Cancer: User-Centered Design Approach ([e21238](#))

Hande Sungur, Nida Yilmaz, Brittany Chan, Maria van den Muijsenbergh, Julia van Weert, Barbara Schouten. . . . . 591

## Harnessing Telemedicine for the Provision of Health Care: Bibliometric and Scientometric Analysis ([e18835](#))

Ahmed Waqas, Soo Teoh, Luís Lapão, Luiz Messina, Jorge Correia. . . . . 605

## Evaluation of a Heart Failure Telemonitoring Program Through a Microsimulation Model: Cost-Utility Analysis ([e18917](#))

Chris Boodoo, Qi Zhang, Heather Ross, Ana Alba, Audrey Laporte, Emily Seto. . . . . 621

## Evaluating the Implementation of a Remote-Monitoring Program for Chronic Obstructive Pulmonary Disease: Qualitative Methods from a Service Design Perspective ([e18148](#))

Florence van Lieshout, Rebecca Yang, Vess Stamenova, Payal Agarwal, Daniel Cornejo Palma, Aman Sidhu, Katrina Engel, Adam Erwood, R Bhatia, Onil Bhattacharyya, James Shaw. . . . . 641

## Maximizing Telerehabilitation for Patients With Visual Loss After Stroke: Interview and Focus Group Study With Stroke Survivors, Carers, and Occupational Therapists ([e19604](#))

Stephen Dunne, Helen Close, Nicola Richards, Amanda Ellison, Alison Lane. . . . . 652

## Barriers and Enablers to Using a Patient-Facing Electronic Questionnaire: A Qualitative Theoretical Domains Framework Analysis ([e19474](#))

Janet Yamada, Andrew Kouri, Sarah-Nicole Simard, Stephanie Segovia, Samir Gupta. . . . . 661

## Automating the Generation of Antimicrobial Resistance Surveillance Reports: Proof-of-Concept Study Involving Seven Hospitals in Seven Countries ([e19762](#))

Cherry Lim, Thyl Miliya, Vilada Chansamouth, Myint Aung, Abhilasha Karkey, Prapit Teparrukkul, Batra Rahul, Nguyen Lan, John Stelling, Paul Turner, Elizabeth Ashley, H van Doorn, Htet Lin, Clare Ling, Soawapak Hinjoy, Sopon Iamsirithaworn, Susanna Dunachie, Tri Wangrangsimakul, Viriya Hantrakun, William Schilling, Lam Yen, Le Tan, Htay Hlaing, Mayfong Mayxay, Manivanh Vongsouvath, Buddha Basnyat, Jonathan Edgeworth, Sharon Peacock, Guy Thwaites, Nicholas Day, Ben Cooper, Direk Limmathurotsakul. . . . . 673

## Efficiency of Computer-Aided Facial Phenotyping (DeepGestalt) in Individuals With and Without a Genetic Syndrome: Diagnostic Accuracy Study ([e19263](#))

Jean Pantel, Nurulhuda Hajjir, Magdalena Danyel, Jonas Elsner, Angela Abad-Perez, Peter Hansen, Stefan Mundlos, Malte Spielmann, Denise Horn, Claus-Eric Ott, Martin Mensah. . . . . 687

## Stress Tracker—Detecting Acute Stress From a Trackpad: Controlled Study ([e22743](#))

Rahul Goel, Michael An, Hugo Alayrangues, Amirhossein Koneshloo, Emmanuel Lincoln, Pablo Paredes. . . . . 698

## Influenza Screening via Deep Learning Using a Combination of Epidemiological and Patient-Generated Health Data: Development and Validation Study ([e21369](#))

Hyunwoo Choo, Myeongchan Kim, Jiyun Choi, Jaewon Shin, Soo-Yong Shin. . . . . 714

Information Access and Use by Patients With Cancer and Their Friends and Family: Development of a Grounded Theory ( <a href="#">e20510</a> )	
Maclean Thiessen, Shane Sinclair, Patricia Tang, Shelley Raffin Bouchal. . . . .	726
A Web Application About Herd Immunity Using Personalized Avatars: Development Study ( <a href="#">e20113</a> )	
Hina Hakim, Julie Bettinger, Christine Chambers, S Driedger, Eve Dubé, Teresa Gavaruzzi, Anik Giguere, Éric Kavanagh, Julie Leask, Shannon MacDonald, Rita Orji, Elizabeth Parent, Jean-Sébastien Paquette, Jacynthe Roberge, Beate Sander, Aaron Scherer, Martin Tremblay-Breault, Kumanan Wilson, Daniel Reinharz, Holly Witteman. . . . .	741
Patients' Adoption of Electronic Personal Health Records in England: Secondary Data Analysis ( <a href="#">e17499</a> )	
Alaa Abd-Alrazaq, Ali Alalwan, Brian McMillan, Bridgette Bewick, Mowafa Househ, Alaa AL-Zyadat. . . . .	761
Adoption of a Personal Health Record in the Digital Age: Cross-Sectional Study ( <a href="#">e22913</a> )	
Consuela Yousef, Abin Thomas, Ahmed Alenazi, Sumaya Elgadi, Laila Abu Esba, Aeshah AlAzmi, Abrar Alhameed, Ahmed Hattan, Saleh Almekhlouf, Mohammed AlShammary, Nazzal Alanezi, Hani Alhamdan, Manal Eldegeir, Rayf Abulezz, Sahal Khoshhal, Clara Masala, Omaira Ahmed. . . . .	786
Evaluation of the Perceived Persuasiveness Questionnaire: User-Centered Card-Sort Study ( <a href="#">e20404</a> )	
Nienke Beerlage-de Jong, Hanneke Kip, Saskia Kelders. . . . .	802
Exergaming With Beat Saber: An Investigation of Virtual Reality Aftereffects ( <a href="#">e19840</a> )	
Ancret Szpak, Stefan Michalski, Tobias Loetscher. . . . .	816
Patients' Convergence of Mass and Interpersonal Communication on an Online Forum: Hybrid Methods Analysis ( <a href="#">e18303</a> )	
Remco Sanders, Theo Araujo, Rens Vliegthart, Mies van Eenbergen, Julia van Weert, Annemiek Linn. . . . .	830
Responses to Concerning Posts on Social Media and Their Implications for Suicide Prevention Training for Military Veterans: Qualitative Study ( <a href="#">e22076</a> )	
Alan Teo, Wynn Strange, Ricky Bui, Steven Dobscha, Sarah Ono. . . . .	844
Health Outcomes from Home Hospitalization: Multisource Predictive Modeling ( <a href="#">e21367</a> )	
Mireia Calvo, Rubèn González, Núria Seijas, Emili Vela, Carme Hernández, Guillem Batiste, Felip Miralles, Josep Roca, Isaac Cano, Raimon Jané. . . . .	854
Impact of Electronic Health Record Interface Design on Unsafe Prescribing of Ciclosporin, Tacrolimus, and Diltiazem: Cohort Study in English National Health Service Primary Care ( <a href="#">e17003</a> )	
Brian MacKenna, Sebastian Bacon, Alex Walker, Helen Curtis, Richard Croker, Ben Goldacre. . . . .	864
Deep Learning With Electronic Health Records for Short-Term Fracture Risk Identification: Crystal Bone Algorithm Development and Validation ( <a href="#">e22550</a> )	
Yasmeen Almog, Angshu Rai, Patrick Zhang, Amanda Moulaison, Ross Powell, Anirban Mishra, Kerry Weinberg, Celeste Hamilton, Mary Oates, Eugene McCloskey, Steven Cummings. . . . .	873
Integrating the Practical Robust Implementation and Sustainability Model With Best Practices in Clinical Decision Support Design: Implementation Science Approach ( <a href="#">e19676</a> )	
Katy Trinkle, Michael Kahn, Tellen Bennett, Russell Glasgow, Heather Haugen, David Kao, Miranda Kroehl, Chen-Tan Lin, Daniel Malone, Daniel Matlock. . . . .	888
Research Participants' Perspectives on Using an Electronic Portal for Engagement and Data Collection: Focus Group Results From a Large Epidemiologic Cohort ( <a href="#">e18556</a> )	
Erika Rees-Punia, Alpa Patel, Asher Beckwitt, Corinne Leach, Susan Gapstur, Tenbroeck Smith. . . . .	903

## Threats of Bots and Other Bad Actors to Data Quality Following Research Participant Recruitment Through Social Media: Cross-Sectional Questionnaire ([e23021](#))

Rachel Pozzar, Marilyn Hammer, Meghan Underhill-Blazey, Alexi Wright, James Tulsy, Fangxin Hong, Daniel Gundersen, Donna Berry. . . .  
9 1 2

## The Abortion Web Ecosystem: Cross-Sectional Analysis of Trustworthiness and Bias ([e20619](#))

Leo Han, Emily Boniface, Lisa Han, Jonathan Albright, Nora Doty, Blair Darney. . . . . 936

## Parental Online Information Access and Childhood Vaccination Decisions in North America: Scoping Review ([e20002](#))

Sarah Ashfield, Lorie Donelle. . . . . 947

## Risk-Taking Behaviors and Adherence to HIV Pre-Exposure Prophylaxis in Users of Geosocial Networking Apps: Real-World, Multicenter Study ([e22388](#))

Hongyi Wang, Jing Zhang, Zhenxing Chu, Qinghai Hu, Willa Dong, Xiaojie Huang, Yaokai Chen, Hui Wang, Xiaoqing He, Lukun Zhang, Zhili Hu, Rantong Bao, Shangcao Li, Hang Li, Sitong Cui, Xia Jin, Haibo Ding, Wenqing Geng, Yongjun Jiang, Junjie Xu, Hong Shang. . . . . 958

## A Data Visualization and Dissemination Resource to Support HIV Prevention and Care at the Local Level: Analysis and Uses of the AIDSvU Public Data Resource ([e23173](#))

Patrick Sullivan, Cory Woodyatt, Chelsea Koski, Elizabeth Pembleton, Pema McGuinness, Jennifer Taussig, Alexandra Ricca, Nicole Luisi, Eve Mokotoff, Nanette Benbow, Amanda Castel, Ann Do, Ronald Valdiserri, Heather Bradley, Chandni Jaggi, Daniel O'Farrell, Rebecca Filipowicz, Aaron Siegler, James Curran, Travis Sanchez. . . . . 971

## Intersection of the Web-Based Vaping Narrative With COVID-19: Topic Modeling Study ([e21743](#))

Kamila Janmohamed, Abdul-Nasah Soale, Laura Forastiere, Weiming Tang, Yongjie Sha, Jakob Demant, Edoardo Airolidi, Navin Kumar. . . .  
9 8 6

## Building a Digital Tool for the Adoption of the World Health Organization's Antenatal Care Recommendations: Methodological Intersection of Evidence, Clinical Logic, and Digital Technology ([e16355](#))

Samira Haddad, Renato Souza, Jose Cecatti, Maria Barreix, Tigest Tamrat, Carolyn Footitt, Garrett Mehl, Inraini Syah, Anuraj Shankar, Özge Tunçalp. . . . . 1000

## Applying Digital Information Delivery to Convert Habits of Antibiotic Use in Primary Care in Germany: Mixed-Methods Study ([e18200](#))

Regina Poss-Doering, Lukas Kuehn, Martina Kamradt, Katharina Glassen, Michel Wensing. . . . . 1014

## Help-Seeking Behaviors of Transition-Aged Youth for Mental Health Concerns: Qualitative Study ([e18514](#))

Chelsea Stunden, Julie Zasada, Nicole VanHeerwaarden, Elisa Hollenberg, Alexxa Abi-Jaoudé, Gloria Chaim, Kristin Cleverley, Joanna Henderson, Andrew Johnson, Andrea Levinson, Brian Lo, Janine Robb, Jenny Shi, Aristotle Voineskos, David Wiljer. . . . . 1030

## Psychometric Evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): Evaluation Study ([e17757](#))

Saskia Kelders, Hanneke Kip, Japie Greeff. . . . . 1045

## Developing a Process for the Analysis of User Journeys and the Prediction of Dropout in Digital Health Interventions: Machine Learning Approach ([e17738](#))

Vincent Bremer, Philip Chow, Burkhardt Funk, Frances Thorndike, Lee Ritterband. . . . . 1057

## Effectiveness of a Participatory and Interactive Virtual Reality Intervention in Patients With Social Anxiety Disorder: Longitudinal Questionnaire Study ([e23024](#))

Hyun-Jin Kim, Seulki Lee, Dooyoung Jung, Ji-Won Hur, Heon-Jeong Lee, Sungkil Lee, Gerard Kim, Chung-Yean Cho, Seungmoon Choi, Seung-Moo Lee, Chul-Hyun Cho. . . . . 1077

## Effects of Interactivity on Recall of Health Information: Experimental Study ([e14783](#))

Emília Pajor, Sander Eggers, Hein de Vries, Anke Oenema. . . . . 1093

Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study ( <a href="#">e22161</a> )	
Akira-Sebastian Poncette, Daniel Glauert, Lina Mosch, Katarina Braune, Felix Balzer, David Back. . . . .	1109
Clinical Context–Aware Biomedical Text Summarization Using Deep Neural Network: Model Development and Validation ( <a href="#">e19810</a> )	
Muhammad Afzal, Fakhare Alam, Khalid Malik, Ghaus Malik. . . . .	1123
Correlation of Online Physician Rating Subscores and Association With Overall Satisfaction: Observational Study of 212,933 Providers ( <a href="#">e11258</a> )	
Hanson Zhao, Michael Luu, Brennan Spiegel, Timothy Daskivich. . . . .	1141
Effect of Smartphone-Based Lifestyle Coaching App on Community-Dwelling Population With Moderate Metabolic Abnormalities: Randomized Controlled Trial ( <a href="#">e17435</a> )	
So Cho, Jung Lee, Jee-Seon Shim, Hyungseon Yeom, Su Lee, Yong Jeon, Hyeon Kim. . . . .	1152
Development of an Item Bank to Measure Medication Adherence: Systematic Review ( <a href="#">e19089</a> )	
Yu Kwan, Livia Oo, Dionne Loh, Jie Phang, Si Weng, Dan Blalock, Eng Chew, Kai Yap, Corrinne Tan, Sungwon Yoon, Warren Fong, Truls Østbye, Lian Low, Hayden Bosworth, Julian Thumboo. . . . .	1179
Web-Based Patient-Reported Outcome Measures for Personalized Treatment and Care (PROMPT-Care): Multicenter Pragmatic Nonrandomized Trial ( <a href="#">e19685</a> )	
Afaf Girgis, Ivana Durcinoska, Anthony Arnold, Joseph Descallar, Nasreen Kaadan, Eng-Siew Koh, Andrew Miller, Weng Ng, Martin Carolan, Stephen Della-Fiorentina, Sandra Avery, Geoff Delaney. . . . .	1226
Associations Between Patient Health Outcomes and Secure Message Content Exchanged Between Patients and Clinicians: Retrospective Cohort Study ( <a href="#">e19477</a> )	
Dawn Heisey-Grove, Laura McClelland, Cheryl Rathert, Alexander Tartaglia, Kevin Jackson, Jonathan DeShazo. . . . .	1240
Federated Learning on Clinical Benchmark Data: Performance Assessment ( <a href="#">e20891</a> )	
Geun Lee, Soo-Yong Shin. . . . .	1272
Accuracy of Nutrient Calculations Using the Consumer-Focused Online App MyFitnessPal: Validation Study ( <a href="#">e18237</a> )	
Charlotte Evenepoel, Egbert Clevers, Lise Deroover, Wendy Van Loo, Christophe Matthys, Kristin Verbeke. . . . .	1282
Reducing Alert Fatigue by Sharing Low-Level Alerts With Patients and Enhancing Collaborative Decision Making Using Blockchain Technology: Scoping Review and Proposed Framework (MedAlert) ( <a href="#">e22013</a> )	
Paul Wan, Abylay Satybaldy, Lizhen Huang, Halvor Holtskog, Mariusz Nowostawski. . . . .	1291
Raising the Digital Profile of Facial Palsy: National Surveys of Patients' and Clinicians' Experiences of Changing UK Treatment Pathways and Views on the Future Role of Digital Technology ( <a href="#">e20406</a> )	
Ala Szczepura, Nikki Holliday, Catriona Neville, Karen Johnson, Amir Khan, Samuel Oxford, Charles Nduka. . . . .	1306
Opportunities and Challenges Surrounding the Use of Data From Wearable Sensor Devices in Health Care: Qualitative Interview Study ( <a href="#">e19542</a> )	
Ijeoma Azodo, Robin Williams, Aziz Sheikh, Kathrin Cresswell. . . . .	1323
Integrating Genomics and Clinical Data for Statistical Analysis by Using GENome MINing (GEMINI) and Fast Healthcare Interoperability Resources (FHIR): System Design and Implementation ( <a href="#">e19879</a> )	
Julian Gruendner, Nicolas Wolf, Lars Tögel, Florian Haller, Hans-Ulrich Prokosch, Jan Christoph. . . . .	1351

## Application of Big Data Technology for COVID-19 Prevention and Control in China: Lessons and Recommendations (e21980)

Jun Wu, Jian Wang, Stephen Nicholas, Elizabeth Maitland, Qiuyan Fan. . . . . 1364

## Interactive Web-Based Resource for Annotation of Genetic Variants Causing Hereditary Angioedema (HADA): Database Development, Implementation, and Validation (e19040)

Alejandro Mendoza-Alvarez, Adrián Muñoz-Barrera, Luis Rubio-Rodríguez, Itahisa Marcelino-Rodríguez, Almudena Corrales, Antonio Iñigo-Campos, Ariel Callero, Eva Perez-Rodríguez, Jose Garcia-Robaina, Rafaela González-Montelongo, Jose Lorenzo-Salazar, Carlos Flores. . . . . 1380

## CoV-Seq, a New Tool for SARS-CoV-2 Genome Analysis and Visualization: Development and Usability Study (e22299)

Boxiang Liu, Kaibo Liu, He Zhang, Liang Zhang, Yuchen Bian, Liang Huang. . . . . 1413

## Characteristics and Symptoms of App Users Seeking COVID-19–Related Digital Health Information and Remote Services: Retrospective Cohort Study (e23197)

Amichai Perlman, Alina Vodonos Zilberg, Peter Bak, Michael Dreyfuss, Maya Leventer-Roberts, Yael Vurembrand, Howard Jeffries, Eyal Fisher, Yael Steurman, Yinat Namir, Yaara Goldschmidt, Daniel Souroujon. . . . . 1419

## Identification of Risk Factors and Symptoms of COVID-19: Analysis of Biomedical Literature and Social Media Data (e20509)

Jouhyun Jeon, Gaurav Baruah, Sarah Sarabadani, Adam Palanica. . . . . 1430

## Association of Web-Based Physical Education With Mental Health of College Students in Wuhan During the COVID-19 Outbreak: Cross-Sectional Survey Study (e21301)

Cheng-Hu Deng, Jing-Qiang Wang, Li-Ming Zhu, He-Wang Liu, Yu Guo, Xue-Hua Peng, Jian-Bo Shao, Wei Xia. . . . . 1440

## Dynamic Panel Surveillance of COVID-19 Transmission in the United States to Inform Health Policy: Observational Statistical Study (e21955)

James Oehmke, Charles Moss, Lauren Singh, Theresa Oehmke, Lori Post. . . . . 1452

## Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study (e21299)

Nicolas Munsch, Alistair Martin, Stefanie Gruarin, Jama Nateqi, Isselmou Abdarrahmane, Rafael Weingartner-Ortner, Bernhard Knapp. . . . . 1472

## Clinical Predictive Models for COVID-19: Systematic Study (e21439)

Patrick Schwab, August DuMont Schütte, Benedikt Dietz, Stefan Bauer. . . . . 1480

## System-Wide Accelerated Implementation of Telemedicine in Response to COVID-19: Mixed Methods Evaluation (e22146)

Diego Garcia-Huidobro, Solange Rivera, Sebastián Valderrama Chang, Paula Bravo, Daniel Capurro. . . . . 1494

## Topics, Delivery Modes, and Social-Epistemological Dimensions of Web-Based Information for Patients Undergoing Renal Transplant and Living Donors During the COVID-19 Pandemic: Content Analysis (e22068)

Charlotte van Klaveren, Peter de Jong, Renée Hendriks, Franka Luk, Aiko de Vries, Paul van der Boog, Marlies Reinders. . . . . 1506

## Associations of Mental Health and Personal Preventive Measure Compliance With Exposure to COVID-19 Information During Work Resumption Following the COVID-19 Outbreak in China: Cross-Sectional Survey Study (e22596)

Yihang Pan, Meiqi Xin, Changhua Zhang, Willa Dong, Yuan Fang, Wenhui Wu, Mingzhe Li, Jun Pang, Zilong Zheng, Zixin Wang, Jinqiu Yuan, Yulong He. . . . . 1518

## Social Media Use, eHealth Literacy, Disease Knowledge, and Preventive Behaviors in the COVID-19 Pandemic: Cross-Sectional Study on Chinese Netizens (e19684)

Xiaojing Li, Qinliang Liu. . . . . 1535

## Epidemiological Parameters of COVID-19: Case Series Study (e19994)

Shujuan Ma, Jiayue Zhang, Minyan Zeng, Qingping Yun, Wei Guo, Yixiang Zheng, Shi Zhao, Maggie Wang, Zuyao Yang. . . . . 1554

Application of an Artificial Intelligence Trilogy to Accelerate Processing of Suspected Patients With SARS-CoV-2 at a Smart Quarantine Station: Observational Study ( <a href="#">e19878</a> )	
Ping-Yen Liu, Yi-Shan Tsai, Po-Lin Chen, Huey-Pin Tsai, Ling-Wei Hsu, Chi-Shiang Wang, Nan-Yao Lee, Mu-Shiang Huang, Yun-Chiao Wu, Wen-Chien Ko, Yi-Ching Yang, Jung-Hsien Chiang, Meng-Ru Shen. ....	1564
How Data Analytics and Big Data Can Help Scientists in Managing COVID-19 Diffusion: Modeling Study to Predict the COVID-19 Diffusion in Italy and the Lombardy Region ( <a href="#">e21081</a> )	
Davide Tosi, Alessandro Campi. ....	1577
Online Health Information Seeking Using “#COVID-19 Patient Seeking Help” on Weibo in Wuhan, China: Descriptive Study ( <a href="#">e22910</a> )	
Xiaoman Zhao, Ju Fan, Iccha Basnyat, Baijing Hu. ....	1585
Knowledge, Awareness, and Attitudes Relating to the COVID-19 Pandemic Among Different Populations in Central China: Cross-Sectional Survey ( <a href="#">e22628</a> )	
Huifang Xu, Maria Gonzalez Mendez, Lanwei Guo, Qiong Chen, Liyang Zheng, Peipei Chen, Xiaoqin Cao, Shuzheng Liu, Xibin Sun, Shaokai Zhang, Youlin Qiao. ....	1598
COVID-19 Self-Reported Symptom Tracking Programs in the United States: Framework Synthesis ( <a href="#">e23297</a> )	
Tracey Koehlmoos, Miranda Janvrin, Jessica Korona-Bailey, Cathaleen Madsen, Rodney Sturdivant. ....	1609
Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study ( <a href="#">e22624</a> )	
Ranganathan Chandrasekaran, Vikalp Mehta, Tejali Valkunde, Evangelos Moustakas. ....	1620
Physical Distancing Measures and Walking Activity in Middle-aged and Older Residents in Changsha, China, During the COVID-19 Epidemic Period: Longitudinal Observational Study ( <a href="#">e21632</a> )	
Yilun Wang, Yuqing Zhang, Kim Bennell, Daniel White, Jie Wei, Ziyang Wu, Hongyi He, Shaohui Liu, Xianghang Luo, Shuo Hu, Chao Zeng, Guanghua Lei. ....	1632
Investigating the Prevalence of Reactive Online Searching in the COVID-19 Pandemic: Infoveillance Study ( <a href="#">e19791</a> )	
Rafael Badell-Grau, Jordan Cuff, Brendan Kelly, Helen Waller-Evans, Emyr Lloyd-Evans. ....	1644
Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing ( <a href="#">e21801</a> )	
Jose Izquierdo, Julio Ancochea, Savana COVID-19 Research Group, Joan Soriano. ....	1673
Determinants of Patients’ Intention to Use the Online Inquiry Services Provided by Internet Hospitals: Empirical Evidence From China ( <a href="#">e22716</a> )	
Dehe Li, Yinhuan Hu, Holger Pfaff, Liuming Wang, Lu Deng, Chuntao Lu, Shixiao Xia, Siyu Cheng, Ximin Zhu, Xiaoyue Wu. ....	1686
Rapid Serological Assays and SARS-CoV-2 Real-Time Polymerase Chain Reaction Assays for the Detection of SARS-CoV-2: Comparative Study ( <a href="#">e19152</a> )	
Angelo Paradiso, Simona De Summa, Daniela Loconsole, Vito Procacci, Anna Sallustio, Francesca Centrone, Nicola Silvestris, Vito Cafagna, Giuseppe De Palma, Antonio Tufaro, Vito Garrisi, Maria Chironna. ....	1706
Depression, Anxiety, and Lifestyle Among Essential Workers: A Web Survey From Brazil and Spain During the COVID-19 Pandemic ( <a href="#">e22835</a> )	
Raquel De Boni, Vicent Balanzá-Martínez, Jurema Mota, Taiane Cardoso, Pedro Ballester, Beatriz Atienza-Carbonell, Francisco Bastos, Flavio Kapczinski. ....	1716

<b>Real-Life Gait Performance as a Digital Biomarker for Motor Fluctuations: The Parkinson@Home Validation Study (e19068)</b>	
Luc Evers, Yordan Raykov, Jesse Krijthe, Ana Silva de Lima, Reham Badawy, Kasper Claes, Tom Heskes, Max Little, Marjan Meinders, Bastiaan Bloem. ....	1732
<b>Digital Pain Mapping and Tracking in Patients With Chronic Pain: Longitudinal Study (e21475)</b>	
Maria Galve Villa, Thorvaldur S Palsson, Albert Cid Royo, Carsten R Bjarkam, Shellie Boudreau. ....	1751
 <b>Reviews</b>	
<b>Implementation of Electronic Informed Consent in Biomedical Research and Stakeholders' Perspectives: Systematic Review (e19129)</b>	
Evelien De Sutter, Drieda Zaçe, Stefania Boccia, Maria Di Pietro, David Geerts, Pascal Borry, Isabelle Huys. ....	49
<b>The Role of Technology and the Continuum of Care for Youth Suicidality: Systematic Review (e18672)</b>	
Hannah Szlyk, Jia Tan. ....	63
<b>Fitbit-Based Interventions for Healthy Lifestyle Outcomes: Systematic Review and Meta-Analysis (e23954)</b>	
Mickael Ringeval, Gerit Wagner, James Denford, Guy Paré, Spyros Kitsiou. ....	82
<b>Tailored Web-Based Smoking Interventions and Reduced Attrition: Systematic Review and Meta-Analysis (e16255)</b>	
Amika Shah, Michael Chaiton, Dolly Baliunas, Robert Schwartz. ....	103
<b>Determining if Telehealth Can Reduce Health System Costs: Scoping Review (e17298)</b>	
Centaine Snoswell, Monica Taylor, Tracy Comans, Anthony Smith, Leonard Gray, Liam Caffery. ....	117
<b>Accuracy of Mobile Device-Compatible 3D Scanners for Facial Digitization: Systematic Review and Meta-Analysis (e22228)</b>	
Hang-Nga Mai, Du-Hyeong Lee. ....	139
<b>Toward a Digital Platform for the Self-Management of Noncommunicable Disease: Systematic Review of Platform-Like Interventions (e16774)</b>	
Sarah Tighe, Kylie Ball, Finn Kensing, Lars Kayser, Jonathan Rawstorn, Ralph Maddison. ....	152
<b>Acceptability and Effectiveness of NHS-Recommended e-Therapies for Depression, Anxiety, and Stress: Meta-Analysis (e17049)</b>	
Melanie Simmonds-Buckley, Matthew Bennion, Stephen Kellett, Abigail Millings, Gillian Hardy, Roger Moore. ....	172
<b>Primary Prevention of Cardiovascular Disease and Type 2 Diabetes Mellitus Using Mobile Health Technology: Systematic Review of the Literature (e21159)</b>	
Vera Buss, Stuart Leesong, Margo Barr, Marlien Varnfield, Mark Harris. ....	210
<b>The Role of Social Media in Enhancing Clinical Trial Recruitment: Scoping Review (e22810)</b>	
Ida Darmawan, Caitlin Bakker, Tabet Brockman, Christi Patten, Milton Eder. ....	922
<b>Measurement Properties of Existing Patient-Reported Outcome Measures on Medication Adherence: Systematic Review (e19179)</b>	
Yu Kwan, Si Weng, Dionne Loh, Jie Phang, Livia Oo, Dan Blalock, Eng Chew, Kai Yap, Corrinne Tan, Sungwon Yoon, Warren Fong, Truls Østbye, Lian Low, Hayden Bosworth, Julian Thumboo. ....	1200



## The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review (e20346)

Madison Milne-Ives, Caroline de Cock, Ernest Lim, Melissa Shehadeh, Nick de Pennington, Guy Mole, Eduardo Normando, Edward Meinert.  
1 2 5 4

## Elaborating Models of eHealth Governance: Qualitative Systematic Review (e17214)

Anne Ekeland, Line Linstad. . . . . 1335

## Artificial Intelligence for COVID-19: Rapid Review (e21476)

Jiayang Chen, Kay See. . . . . 1661

## Viewpoints

### Private Video Consultation Services and the Future of Primary Care (e19415)

Chris Salisbury, Anna Quigley, Nick Hex, Camille Aznar. . . . . 225

### The Need for Sustainable Teleconsultation Systems in the Aftermath of the First COVID-19 Wave (e21211)

Guido Giunti, Richard Goossens, Antoinette De Bont, Jacob Visser, Mark Mulder, Stephanie Schuit. . . . . 235

### Covidom, a Telesurveillance Solution for Home Monitoring Patients With COVID-19 (e20748)

Youri Yordanov, Agnes Dechartres, Xavier Lescure, Caroline Apra, Pascaline Villie, Jerome Marchand-Arvier, Erwan Debuc, Aurélien Dinh, Patrick Jourdain, On Behalf Of The AP-HP / Universities / Inserm COVID-19 Research Collaboration. . . . . 241

### Portals of Change: How Patient Portals Will Ultimately Work for Safety Net Populations (e16835)

Alejandra Casillas, Anshu Abhat, Anish Mahajan, Gerardo Moreno, Arleen Brown, Sara Simmons, Peter Szilagyi. . . . . 249

### Proposed Implementation of Blockchain in British Columbia's Health Care Data Management (e20897)

Danielle Cadoret, Tamara Kailas, Pedro Velmovitsky, Plinio Morita, Okechukwu Igboeli. . . . . 258

### Digital Micro Interventions for Behavioral and Mental Health Gains: Core Components and Conceptualization of Digital Micro Intervention Care (e20631)

Amit Baumel, Theresa Fleming, Stephen Schueller. . . . . 1165

### Adapting an Outpatient Psychiatric Clinic to Telehealth During the COVID-19 Pandemic: A Practice Perspective (e22523)

Farzan Sasangohar, Major Bradshaw, Marianne Carlson, James Flack, James Fowler, Diana Freeland, John Head, Kate Marder, William Orme, Benjamin Weinstein, Jacob Kolman, Bitu Kash, Alok Madan. . . . . 1404

## Notes From the Field

### Selecting Mobile Health Technologies for Electronic Health Record Integration: Case Study (e23314)

Ryan Shaw, Marissa Stroo, Christopher Fiander, Katlyn McMillan. . . . . 273

## Tutorial

### Maximizing the Potential of Patient-Reported Assessments by Using the Open-Source Concerto Platform With Computerized Adaptive Testing and Machine Learning (e20950)

Conrad Harrison, Bao Loe, Przemysław Lis, Chris Sidey-Gibbons. . . . . 280

## Letters to the Editor

Is a Ratio Scale Assumption for Physician Ratings Justified? Comment on “What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data” ( <a href="#">e18289</a> ) Uwe Konerding. . . . .	1392
Authors’ Reply to: Is a Ratio Scale Assumption for Physician Ratings Justified? Comment on “What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data” ( <a href="#">e21057</a> ) Sonja Bidmon, Ossama Elshiewy, Ralf Terlutter, Yasemin Boztug. . . . .	1394
Comment on “Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial” ( <a href="#">e21505</a> ) Yusuke Saishoji, Akihiro Shiroshita, Yasushi Tsujimoto. . . . .	1396
Authors’ Reply to: Comment on “Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial” ( <a href="#">e24084</a> ) Windy Chan, Angela Leung. . . . .	1398

## Corrigenda and Addendas

Correction: Building a Digital Tool for the Adoption of the World Health Organization’s Antenatal Care Recommendations: Methodological Intersection of Evidence, Clinical Logic, and Digital Technology ( <a href="#">e24891</a> ) Samira Haddad, Renato Souza, Jose Cecatti, Maria Barreix, Tigest Tamrat, Carolyn Footitt, Garrett Mehl, Inraini Syah, Anuraj Shankar, Özge Tunçalp. . . . .	1400
Correction: Twelve-Month Follow-Up to a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention for Rural Adults With Depression Symptoms: Single-Arm Longitudinal Study ( <a href="#">e25146</a> ) Mark Schure, Bernadette McCrory, Kathryn Tuchscherer Franklin, John Greist, Ruth Weissman. . . . .	1402

Original Paper

# Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study

Daniel M Low<sup>1,2,3</sup>, MSc; Laurie Rumker<sup>4,5,6,7\*</sup>, MSc; Tanya Talkar<sup>1,8\*</sup>, MEng; John Torous<sup>9</sup>, MBI, MD; Guillermo Cecchi<sup>10</sup>, PhD; Satrajit S Ghosh<sup>1,3,11</sup>, PhD

<sup>1</sup>Program in Speech and Hearing Bioscience and Technology, Harvard Medical School, Boston, MA, United States

<sup>2</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States

<sup>3</sup>McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, United States

<sup>4</sup>Center for Data Sciences, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

<sup>5</sup>Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

<sup>6</sup>Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, United States

<sup>7</sup>Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

<sup>8</sup>Human Health and Performance Systems, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA, United States

<sup>9</sup>Digital Psychiatry Division, Department of Psychiatry at Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

<sup>10</sup>Thomas J Watson Research Center, IBM, Yorktown Heights, NY, United States

<sup>11</sup>Department of Otolaryngology, Harvard Medical School, Boston, MA, United States

\*these authors contributed equally

## Corresponding Author:

Daniel M Low, MSc

Department of Brain and Cognitive Sciences

Massachusetts Institute of Technology

43 Vassar St

Cambridge, MA, 02139

United States

Phone: 1 6176317230

Email: [dlow@mit.edu](mailto:dlow@mit.edu)

## Abstract

**Background:** The COVID-19 pandemic is impacting mental health, but it is not clear how people with different types of mental health problems were differentially impacted as the initial wave of cases hit.

**Objective:** The aim of this study is to leverage natural language processing (NLP) with the goal of characterizing changes in 15 of the world's largest mental health support groups (eg, r/schizophrenia, r/SuicideWatch, r/Depression) found on the website Reddit, along with 11 non-mental health groups (eg, r/PersonalFinance, r/conspiracy) during the initial stage of the pandemic.

**Methods:** We created and released the Reddit Mental Health Dataset including posts from 826,961 unique users from 2018 to 2020. Using regression, we analyzed trends from 90 text-derived features such as sentiment analysis, personal pronouns, and semantic categories. Using supervised machine learning, we classified posts into their respective support groups and interpreted important features to understand how different problems manifest in language. We applied unsupervised methods such as topic modeling and unsupervised clustering to uncover concerns throughout Reddit before and during the pandemic.

**Results:** We found that the r/HealthAnxiety forum showed spikes in posts about COVID-19 early on in January, approximately 2 months before other support groups started posting about the pandemic. There were many features that significantly increased during COVID-19 for specific groups including the categories "economic stress," "isolation," and "home," while others such as "motion" significantly decreased. We found that support groups related to attention-deficit/hyperactivity disorder, eating disorders, and anxiety showed the most negative semantic change during the pandemic out of all mental health groups. Health anxiety emerged as a general theme across Reddit through independent supervised and unsupervised machine learning analyses. For instance, we provide evidence that the concerns of a diverse set of individuals are converging in this unique moment of history; we discovered that the more users posted about COVID-19, the more linguistically similar (less distant) the mental health support

groups became to r/HealthAnxiety ( $\rho=-0.96$ ,  $P<.001$ ). Using unsupervised clustering, we found the suicidality and loneliness clusters more than doubled in the number of posts during the pandemic. Specifically, the support groups for borderline personality disorder and posttraumatic stress disorder became significantly associated with the suicidality cluster. Furthermore, clusters surrounding self-harm and entertainment emerged.

**Conclusions:** By using a broad set of NLP techniques and analyzing a baseline of prepandemic posts, we uncovered patterns of how specific mental health problems manifest in language, identified at-risk users, and revealed the distribution of concerns across Reddit, which could help provide better resources to its millions of users. We then demonstrated that textual analysis is sensitive to uncover mental health complaints as they appear in real time, identifying vulnerable groups and alarming themes during COVID-19, and thus may have utility during the ongoing pandemic and other world-changing events such as elections and protests.

(*J Med Internet Res* 2020;22(10):e22635) doi:[10.2196/22635](https://doi.org/10.2196/22635)

## KEYWORDS

COVID-19; mental health; psychiatry; infodemiology; infoveillance; infodemic; social media; Reddit; natural language processing; ADHD; eating disorders; anxiety; suicidality

## Introduction

The ongoing outbreak of a novel coronavirus causing the disease COVID-19 is likely to have impacts on mental health as many individuals experience losses of income, social engagement, mobility, physical health, and uncertainty. Characterizing these impacts is critical to motivate and inform the provision of appropriate therapeutic responses. Public commentary posted to mental health support groups on the website Reddit through subfora (subreddits) captures in real time the language used by those sharing and processing their pandemic experiences online. In this study, we apply text processing and machine learning techniques to this data set to analyze COVID-19's impacts on mental health discourse as a potential proxy for changes in mental health needs.

In the setting of a quarantine, incidence rates rise for mood disorders including acute stress disorder, posttraumatic stress disorder (PTSD), major depressive disorder, and generalized anxiety disorder, as do rates of subclinical mental health deterioration [1]. Early data from the present outbreak gathered in China indicates that rates of depressive symptoms (50.4%), anxious symptoms (44.6%), insomnia (34%), and general distress (71.5%) are especially high among health care workers [2]. Young adults with pre-existing mental and physical health conditions indicate worsening symptoms of depression and anxiety, particularly nonbinary and female young adults [3]. Increased alcohol consumption is linked to isolation, income loss, and adjustments to living with children [4]. The scope of likely mental health deterioration during this pandemic provides an unprecedented need to understand how different mental health cohorts are responding to the outbreak to best design patient assessments and allocate resources. The collective mental health care needs of the general population, which are insufficiently addressed even at baseline, are now increased in the setting of the pandemic. In the context of this limited access to care, it is all the more important to identify subpopulations most impacted by the pandemic to triage resource allocation in an informed manner.

Using natural language processing (NLP) to infer mental states of individuals from their social media posts is a growing field (for reviews, see [5,6]). On Reddit, individuals compose

anonymous posts on *subreddit* fora, each with a central topic. Although Reddit posts are not accompanied by formal clinical diagnoses and related covariates, this data set offers several advantages relative to traditional mental health clinical data sets: data is immediately and publicly available, historical data allows for comparisons of multiple time frames, and anonymous and free-form posts create an ecological documentation of vulnerable first-person experiences.

Machine learning models have been created to classify and characterize Reddit posts originating from mental health subreddits [7-10]. Shen and Rudzicz [7] achieved 98% accuracy in separating posts on any of four different anxiety-related subreddits from posts on control subreddits. They used a combination of N-gram language modeling and Linguistic Inquiry and Word Count (LIWC) [11]. LIWC has successfully aided in the detection of depression from Twitter activity with an accuracy of 70% [12] and of bipolar disorder from Reddit posts [8]. These models typically focus on binary classification of posts with respect to a single disorder or subreddit (eg, was the post made on r/Anxiety or a control subreddit).

Visualization and analysis of topics through unsupervised machine learning methods also aid in differentiating between mental health subreddits. Using k-means clustering, analysis of themes present in r/Anxiety, r/Depression, and r/PTSD found that the r/Anxiety and r/PTSD subreddits shared more common terms with each other than with the r/Depression subreddit [13].

Reddit can also help monitor discussions around public health subjects. Using latent dirichlet allocation (LDA), one study identified that Reddit discussions of Ebola focused on health education through posts about best practices and the implications of public health events [14]. LDA uncovers common topics present in text documents, capturing sets of words that typically appear in documents together, which can then be inspected manually to assess common themes across the documents. Combined with an analysis of linguistic changes within mental health subreddits, topic analysis can provide insights into the issues surrounding users during public health crises.

In this paper, we analyze changes and trends in language features during the pandemic to uncover any negative semantic changes, develop machine learning models that successfully classify

posts as originating from a particular mental health subreddit to then explain which features characterize each subreddit, understand concerns shared across Reddit independent of the subreddit's origin using unsupervised machine learning methods, and assess whether subreddits are becoming more similar using supervised dimensionality reduction given the global focus on the pandemic. We expect that quantifying differences between pre- and midpandemic post content and characterizing discussions in the r/COVID19\_support group will yield valuable insight into the impact of COVID-19 on mental health and help design treatment more effectively.

## Methods

### Reddit Mental Health Data set

#### Reddit Users

Demographic information per subreddit is unavailable, but Reddit users collectively are predominantly American (49.9%), male (67%), and young (22%, 18-29 years of age; 14%, 30-49 years of age) [15,16].

#### Data Downloading and Preprocessing

Data was downloaded using the pushshift application programming interface [17]. Posts were extracted from fifteen subreddits focused on specific mental health communities (r/EDAnonymous, r/addiction, r/alcoholism, r/adhd, r/anxiety, r/autism, r/BipolarReddit, r/bpd, r/depression, r/healthanxiety, r/lonely, r/ptsd, r/schizophrenia, r/socialanxiety, and r/SuicideWatch), two broad mental health subreddits (r/mentalhealth and r/COVID19\_support), and 11 nonmental health subreddits (r/conspiracy, r/divorce, r/fitness, r/guns, r/jokes, r/legaladvice, r/meditation, r/parenting, r/personalfinance, r/relationships, and r/teaching). Details of data and preprocessing are provided in [Multimedia Appendix 1 Methods 1.1](#).

#### Feature Extraction

The following features were extracted from posts: LIWC (n=62); sentiment analysis (n=4); basic word and syllable counts (n=8); punctuation (n=1); readability metrics (n=9); term frequency-inverse document frequency (TF-IDF) ngrams (256-1024) to capture words and phrases that characterize specific posts; and manually built lexicons about suicidality (n=1), economic stress (n=1), isolation (n=1), substance use (n=1), domestic stress (n=1), and guns (n=1). Trend analysis used all features except TF-IDF; classification and supervised dimensionality reduction used all features, as did unsupervised clustering, and LDA used TF-IDF features. See [Multimedia Appendix 1 Methods 1.2](#) for more details.

### Classification and Feature Importance

#### Training and Testing

Binary classification was performed on each of the 15 specific mental health subreddits versus a control group (n=2700 each) made of a random sample of the remaining subreddits (first balanced to assure subreddits were equally represented). Only one post per user was used to avoid overfitting. An 80-20 train-test split was used. Two additional test sets were built to

test the prepandemic model. One included midpandemic (from March 11, 2020, to April 20, 2020; mean n=504, SD 54.3 posts combined) to measure potential data set shift. A second test set was composed of posts from r/COVID19\_support and a control group (n=1574 combined) to measure how they would be classified by our prepandemic model. A weighted F1 was used to measure performance.

#### Models

Our goal was to use the model with the lowest complexity to determine feature importance more directly as long as the model is not considerably outperformed by more complex models. We tested three linear models (stochastic gradient descent linear classifier [SGD] with L1 penalty, SGD with elastic net penalty, and linear support vector machine) along with two more complex tree ensemble classifiers (extra trees and gradient boosting trees).

#### Trend Analysis

We first tracked COVID-19-related tokens (see [Multimedia Appendix 1 Table S1](#)) from January 1, 2020, to April 20, 2020, across mental health subreddits. We compared this to the confirmed COVID-19 cases with data obtained from [18]. We then grouped data every 2 days from January 1 to April 20 for every year. For each of the 90 features and 28 subreddits, we fit a linear regression for the average feature value at each time point as a function of time. We applied the Benjamini-Hochberg procedure for multiple hypothesis testing correction with  $\alpha=.05$ . We defined change for each feature and subreddit as the slope  $\times R^2$  (ie, the rate of change weighed by the goodness of fit). We tested the amount of absolute change between years by using a Mann-Whitney *U* test.

### Unsupervised Clustering

Prepandemic posts made in 2019 from 15 mental health subreddits were downsampled to a balanced representation (1500 posts/subreddit). The feature set for these posts previously described, including 1024 TF-IDF ngrams, was reduced to 30 principal component analysis (PCA) components. scikit-learn's SpectralClustering function was used to identify k=20 clusters, using default parameters apart from a nearest neighbors-based (n=10) affinity matrix (selection of k is described in [Multimedia Appendix 1 Figure S8](#)). Wilcoxon rank sum tests (with Bonferroni correction) identified cluster-characteristic features for annotating each cluster with a defining theme. Cluster annotation validity was verified through post review. Hypergeometric tests (with Bonferroni correction) identified enrichment of posts from particular clusters on particular subreddits. Posts from the midpandemic data set were processed through the same pipeline. The resulting clusters were compared on the basis of cluster-characteristic features to clusters defined in the prepandemic data set, and the vast majority of clusters were found to have a close match pairing in each time window, as illustrated in [Multimedia Appendix 1 Table S4](#).

### Topic Modeling With Latent Dirichlet Allocation

The *gensim* library was used to perform LDA model estimation, which determined sets of words that appeared frequently together in posts across mental health subreddits. To balance



across subreddits, 2700 prepandemic posts from 2019 were sampled from each of the 15 focused mental health subreddits to form our model. Multiple models were created to assess topic stability. A final prepandemic model with 10 topics was chosen to ensure distinct and important topics. A separate model was created with 10 topics using 1300 posts from each midpandemic mental health subreddits, spanning January 1, 2020, to April 20, 2020, to assess any shift in topics. Models were then applied to all subreddits to assess the distribution of the posts across the topics. Models were applied to midpandemic posts between March 16, 2020, and April 20, 2020, to capture topic distributions during the acute phase of the pandemic. Comparison of topic distributions between pre- and midpandemic posts was done with a two-sided Wilcoxon signed rank test with the Benjamini-Hochberg procedure ( $\alpha=.05$ ) to test whether the incidence of these topics changed across subreddits as a result of the pandemic. Details are provided in [Multimedia Appendix 1](#) Methods 1.3.

### Measuring Similarity Between Subreddits Over Time With Supervised Dimensionality Reduction

With the goal of measuring the similarity between the 15 mental health subreddits as COVID-19 spread, we first reduced posts' 346D feature vectors to 2D using supervised Uniform Manifold Approximation and Projection (UMAP), a dimensionality reduction technique that can capture nonlinear structures in the data (in comparison to PCA) and that better preserves global structure in comparison to other methods such as t-distributed stochastic neighbor embedding [19]. We then measured the

asymmetric Hausdorff distance between subreddits as time progresses to estimate which subreddits are becoming more similar (ie, less distant) to each other. The Hausdorff distance between two clusters is the greatest of all the distances from a point in one cluster to the closest point in the other cluster and, therefore, considers all points in a cluster instead of just a single point such as the centroid as other distances like Euclidean distance would. We balanced subreddits to 1300 posts each spanning from January 1, 2020, to April 20, 2020, grouped in 15-day time windows. This subsampling and dimensionality reduction was repeated for 50 bootstrapping samples, and we took the median distance value. See [Multimedia Appendix 1](#) Methods 1.4 for a test of the precision of this method on 2019 data.

## Results

### Classification and Feature Importance

Models all performed similarly with mean weighted F1 scores between 0.798-0.857 (see [Multimedia Appendix 1](#) Table S1). The SGD L1 (F1=0.851) was chosen for further analysis given that it had the lowest model complexity. Important features are available in [Table 1](#). We then applied this model to midpandemic data (March 11, 2020, to April 20, 2020), and performance changed in several classes (see [Multimedia Appendix 1](#) Table S3). These binary classification models were applied to r/COVID19\_support posts to characterize them psychologically (in section r/COVID19\_support Characterization).

**Table 1.** Important features for classification (ranked).<sup>a</sup>

Subreddit	Positive coefficients	Negative coefficients
r/EDAnonymous	ed, restrict, purg, bing, calori, LIWC <sup>b</sup> ingestion, fast, recoveri, eat, ate	bpd, anxieti, addict, diagnos, drug, substance use lexicon, ptsd, LIWC work, LIWC health, med
r/addiction	addict, clean, smoke, rehab, sober, drug, weed, relaps, use, guns lexicon	bpd, diagnos, ptsd, adhd, therapi, isolation lexicon, LIWC hear, therapist, post, LIWC work
r/adhd	adhd, adderal, add, vyvans, focu, forget, final, LIWC work, medic, came	bpd, ptsd, hurt, therapi, guess, suicidality lexicon, fear, bodi, suicid, pain
r/alcoholism	sober, alcohol, drink, withdraw, drunk, LIWC nonfluencies, drank, meet, beer	drug, weight, therapi, adhd, medic, isolation lexicon, notic, attack, dure, addict
r/anxiety	anxieti, wa dead, LIWC negative emotion, LIWC money, anxio, LIWC motion, LIWC numbers, panic attack, anxious	ptsd, bpd, adhd, LIWC ingestion, LIWC articles article, addict, kill, substance use lexicon, LIWC body, social
r/autism	autism, autist, spectrum, son, game, diagnos, function, diagnosi, explain, interest	ptsd, bpd, addict, LIWC health, disord, adhd, med, 2, stay, guns lexicon
r/BipolarReddit	bipolar, manic, mania, lithium, mood, episod, psychiatrist, hospit, LIWC money, med	adhd, addict, bpd, ptsd, LIWC ingestion, LIWC anxiety, LIWC work, automated readability index, LIWC future tense
r/bpd	bpd, fp, LIWC numbers, LIWC inclusive, LIWC negative emotion, LIWC sadness, bad, drug, LIWC affective processes, feel	ptsd, adhd, weight, LIWC articles article, LIWC health, addict, LIWC 1st pers, anxious, food, isolation lexicon
r/depression	depress, LIWC sadness, LIWC negations, gunning fog index, LIWC positive emotion, LIWC family, cri, LIWC feel, bed, LIWC total pronouns	bpd, symptom, ptsd, adhd, food, isolation lexicon, LIWC conjunctions, diagnos, addict, n sents
r/healthanxiety	cancer, LIWC biological, health anxieti, LIWC health, health, LIWC body, test, fine, LIWC assent, googl	ptsd, adhd, bpd, addict, emot, disord, LIWC 3rd pers, LIWC social processes, social, mental
r/lonely	lone, loneli, isolation lexicon, messag, LIWC certainty, friend, girl, LIWC positive emotion, sit, LIWC religion	LIWC anxiety, ptsd, addict, bpd, suicidality lexicon, symptom, therapist, abus, suicid, med
r/ptsd	ptsd, trauma, flashback, trigger, nightmar, sexual, domestic stress lexicon, abus, tire, guns lexicon	bpd, addict, drink, adhd, LIWC health, isolation lexicon, LIWC work, disord, LIWC certainty, LIWC sadness
r/schizophrenia	schizophrenia, hallucin, delus, schizophren, voic, paranoid, LIWC religion, hospit, ill, LIWC tentative	adhd, ptsd, bpd, addict, abus, flesch kincaid grade level, isolation lexicon, LIWC body, LIWC biological, LIWC health
r/socialanxiety	social anxieti, nervou, walk, awkward, girl, group, convers, speak, face, anxieti	bpd, adhd, ptsd, LIWC health, addict, diagnos, suicid, LIWC sadness, support
r/SuicideWatch	suicidality lexicon, suicid, LIWC negations, death, kill, want die, LIWC sadness, LIWC friends, LIWC money, plan	bpd, symptom, anxious, substance use lexicon, usual, LIWC 2nd pers, smog index, isolation lexicon, weight, attack

<sup>a</sup>Their presence makes it more (positive) or less (negative) likely the classifier will predict the subreddit. Individual word stems are obtained from term frequency–inverse document frequency.

<sup>b</sup>LIWC: Linguistic Inquiry and Word Count.

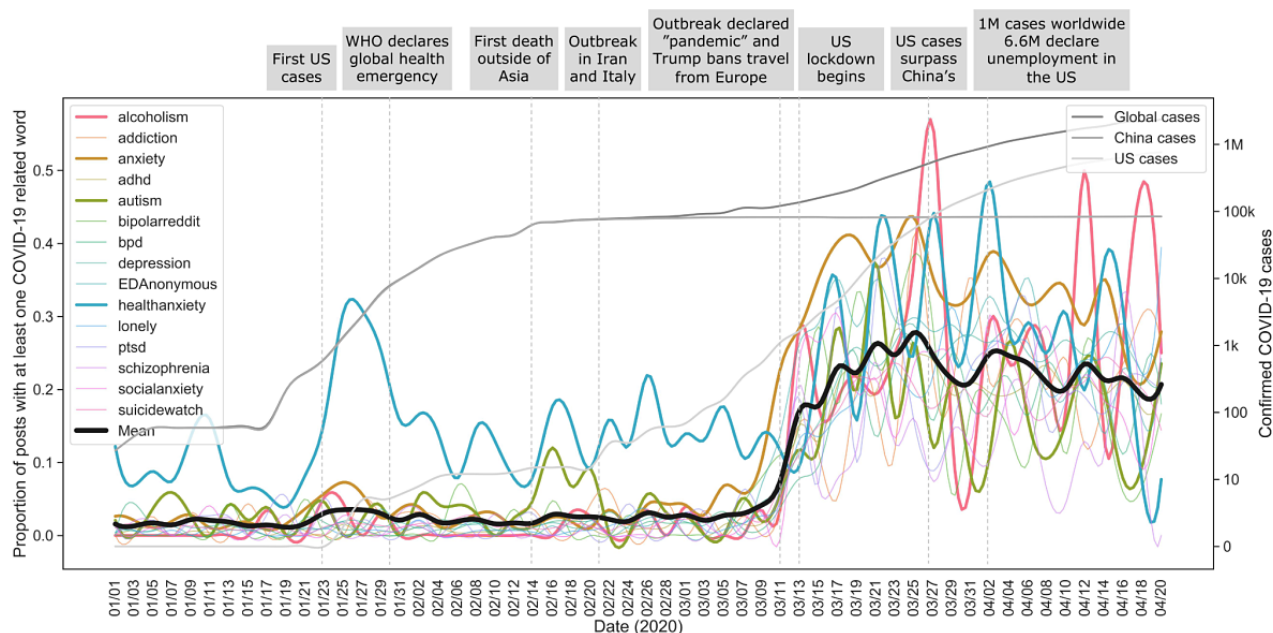
## Trend Analysis

See [Figure 1](#) for the proportion of posts about COVID-19 in each support group. For a measure of how much of each post was about COVID-19, see [Multimedia Appendix 1](#) Figure S3. For an example of how trends were computed for a single feature and subreddit, see [Multimedia Appendix 1](#) Figure S4. [Figure 2](#) shows a subset of the trends and most negative semantic change (ie, sum of change in negative semantic features). Throughout many subreddits, we found significant increases in the use of

tokens related to isolation (eg, “lonely,” “can’t see anyone,” “quarantine”), economic stress (eg, “rent,” “debt,” “pay the bills”), and home (“fridge,” “pet,” “lease”), and a decrease in the lexicon related to motion (eg, “walk,” “visit,” “travel”), all consistent with the type of changes many are facing during the ongoing pandemic. See [Multimedia Appendix 1](#) Figures S5-S7 for full results (90 features in 28 subreddits). See [Textbox 1](#) for examples of posts that score high on important features and display risky behavior.

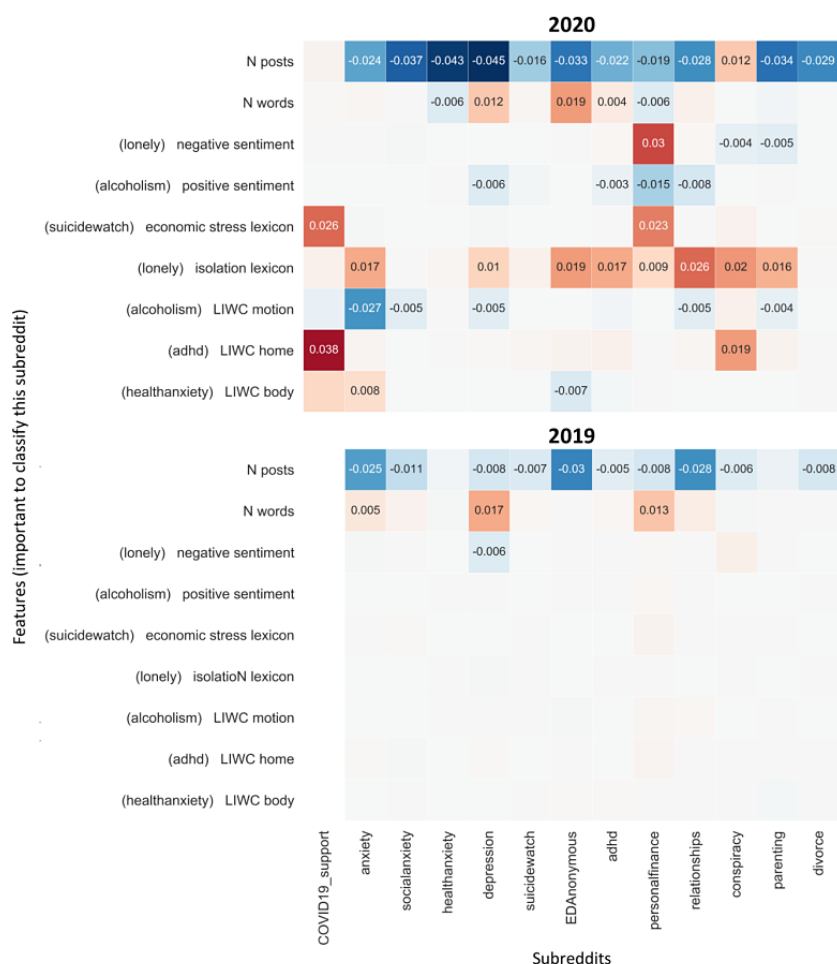


**Figure 1.** Mention of COVID-19–related words across mental health support groups. Timeline landmarks were chosen from NBC News timeline given that US users are the most prevalent across Reddit. Global, China, and US confirmed COVID-19 cases are displayed. The overall acute rise in COVID-19–related words occurs on March 11, 2020. The correlation between the mean proportion of COVID-19–related posts and global COVID-19 cases is  $p=0.83$  ( $P<.001$ ). The health anxiety subreddit has a large increase in COVID-19–related posts almost 2 months before the general increase. *r/alcoholism* has the most amount of posts related to COVID-19 on March 27. *adhd*: attention-deficit/hyperactivity disorder; *bpd*: borderline personality disorder; *EDAnonymous*: Eating Disorders Anonymous; *ptsd*: posttraumatic stress disorder; *WHO*: World Health Organization.



**Figure 2.** Trend analysis of linguistic features over time. A) Significant change in average feature values from January to April 2020 and 2019 across subreddits. COVID19 support subreddit was created in 2020 and, therefore, does not appear in 2019. Features important for classifying a subreddit are added to the y-axis. Change is defined by slope  $\times R^2$  (ie, increases have a positive slope and tend toward red, decreases have a negative slope and tend toward blue). Significant trends after multiple comparison correction on full results are displayed. There is significantly more absolute change in 2020 than in 2019 ( $P < .001$ ) and 2018 ( $P < .001$ ). B) Rank of subreddits by the amount of negative semantic change throughout COVID-19 (January 1, 2020, to April 20, 2020) across significant full results. In bold are the mental health subreddits with the most negative semantic change using the following features with emotional valence: negative sentiment; the lexicons about economic stress, isolation, substance use, guns, domestic stress, and suicidality; LIWC measures of anger, anxiety, death, negations, negative emotion, and sadness; and three positive features inversely weighed, compound sentiment, positive sentiment, and positive emotion. r/ptsd and r/conspiracy decreased in negative semantic features. adhd: attention-deficit/hyperactivity disorder; bpd: borderline personality disorder; EDAnonymous: Eating Disorders Anonymous; LIWC: Linguistic Inquiry and Word Count; ptsd: posttraumatic stress disorder.

A



B

Subreddit	Negative semantic change
personalfinance	0.139
relationships	0.042
fitness	0.029
COVID19_support	0.026
jokes	0.024
<b>adhd</b>	<b>0.023</b>
<b>EDAnonymous</b>	<b>0.019</b>
mentalhealth	0.018
<b>anxiety</b>	<b>0.017</b>
depression	0.014
parenting	0.01
legaladvice	0.007
bpd	0.006
addiction	0.005
alcoholism	0.004
guns	0.004
meditation	0.003
socialanxiety	0
healthanxiety	0
autism	0
bipolarreddit	0
lonely	0
suicidewatch	0
divorce	0
teaching	0
schizophrenia	0
ptsd	-0.005
conspiracy	-0.011

**Textbox 1.** Example of posts that score high on a subset of important features.

#### COVID-19 lexicon

**r/healthanxiety (January 19, 2020):** “I’ve been seeing a lot of news stories today about an incurable, lethal disease originating in China spreading to other countries. I know the media often over exaggerates things like this, but WHO is debating whether or not to declare this an international emergency. Nothing triggers my anxiety worse than things like this, if anyone has any advice on how to cope/can inform me on it I would really appreciate it!”

#### Linguistic Inquiry and Word Count (LIWC) achievement

**r/addiction (March 20, 2020):** “Anyone else relapse during quarantine? Looking for others who are going through this and would like to swap experience, strength and hope?...all my meetings have been canceled”

#### LIWC affective processes

**r/bpd (March 27, 2020):** “...I posted this in another group but I’m really struggling right now. I really wish I could get into a DBT group but now I’m so confused with all of this virus stuff. Does anyone have effective tips for dealing with splitting? Sometimes I will be upset with my S/O for hours or days and I don’t even know why at a certain point. I get so manipulative and I really hate it. Does anyone else get like this?”

#### LIWC ingestion

**r/EDAnonymous (March 19, 2020):** “Quarantine is slowly chipping away at my recovery. I can’t help but feel like this is the perfect time to fast or restrict since no one’s monitoring my meals. Nobody at school noticing. I’m skipping lunch for the 4th day in a row or only eating celery. I was doing so well!!!!...”

#### LIWC money and guns lexicon

**r/SuicideWatch (March 29, 2020):** “Suicide is too expensive with increasing gun prices I’ve been watching gun deals to see if I could afford something that was quick and surefire, but covid has made guns more expensive with greater wait times.”

#### LIWC negations

**r/depression (March 21, 2020):** “I can’t do it If things don’t get back to normal in a few weeks I want to kill myself. The only thing I was excited about was starting a new job and I can’t anymore because of all the coronavirus shut downs. I’m going crazy I don’t have anyone to talk to and I just can’t do this anymore”

#### LIWC see

**r/socialanxiety (April 13, 2020):** “Now that everyone is wearing face masks, I suddenly don’t have trouble making eye contact with people anymore. I might keep wearing a mask out in public after the quarantine is over. I really like the confidence it gives me. I don’t have to worry about what my facial expression is or looks like, I don’t have to worry about smiling, everyone looks the same around me.”

#### Economic stress lexicon

**r/COVID19\_support (March 10, 2020):** “Would I be allowed a temporary paid Leave of Absence by my psychiatrist from my unionized supermarket due to COVID19? It’s causing me major anxiety, everyone else is calm but it sadly compromises my calmness. I am diagnosed with schizoaffective disorder/bipolar. There’s only one confirmed case in my county, but still. School for me, may be canceled. How do I handle working my part time job?”

#### Negative sentiment

**r/adhd (March 20, 2020):** “Drug interactions with COVID-19?? I just read about methylphenidate being directly related to a lower white blood cell count/ worse immune system...AND it can cause high blood pressure. I’m hyper-focused on this virus and i’m terrified that Concerta is putting me more at risk. Anyone else in the same boat?”

#### LIWC health

**r/conspiracy (March 13, 2020):** “...How has the coronavirus been treated? Iodine and vitamin C. How has radiation sickness been treated? Iodine and vitamin C. China and Italy lead the world in radioactive 5G cell towers. 5G operates on the same frequency and wavelength as military grade weapon technology. Connect the dots.”

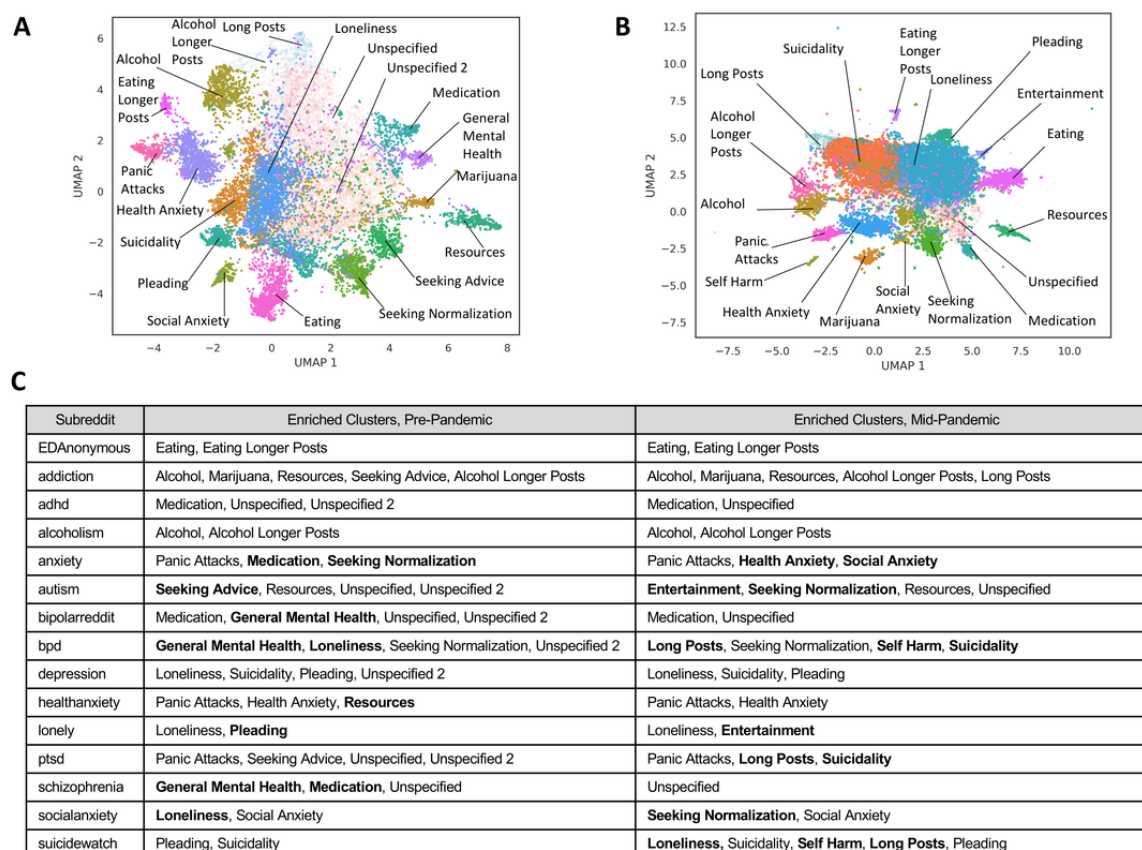
## Unsupervised Analysis

### *Unsupervised Clustering Uncovers Post Language Themes*

Clustering of prepandemic posts ([Multimedia Appendix 1](#) Figure S8, [Figure 3](#)) highlights that language use across subreddits

represents a continuum but contains meaningful axes of variation reflected by the proximity of clusters for closely related conversation topics. Cluster-distinguishing terms ([Multimedia Appendix 1](#) Table S4) used to assign annotations reveal that some clusters are characterized by discussion of particular mental health concerns (eg, “Suicidality”) while others are characterized by post tone (eg, “Seeking Normalization”).

**Figure 3.** Unsupervised clustering reveals post groupings with representation across mental health subreddits. A) Unsupervised clustering of pre-pandemic (year 2019) posts from 15 mental health subreddits, presented in 2D UMAP space. Posts in two thematically-related, adjacent clusters were collapsed into a single “Resources” cluster. Three clusters could not be assigned identifiable themes. Two of these—annotated as “Unspecified”—were the largest clusters in the dataset, containing 6329 and 4272 total posts, respectively, while the next largest cluster contained 1620 posts. The other cluster without an identifiable theme was characterized by very long posts. This “Long Posts” cluster had an average post length of 886 words, while the cluster with the next most lengthy posts had an average post length of 554 words. As a result, this “Long Posts” cluster had an overwhelming number of cluster-characteristic text features, which made any core linguistic theme poorly discernible. The identified clusters were not an approximation of post subreddit of origin, as demonstrated by several metrics quantifying the lack of correspondence between cluster labels and post subreddit of origin: Homogeneity (0.20), Completeness (0.22), V-measure (0.21), and Adjusted Rand-Index (0.08). B) Unsupervised clustering of mid-pandemic posts using the same process resulted primarily in replication of cluster annotations observed in the pre-pandemic data, with a few clusters (e.g., Seeking Advice) detected only in the mid-pandemic clustering. Two clusters increased notably in size in the mid-pandemic clustering: Suicidality (204% increase in number of posts) and Loneliness (233% increase in number of posts). C) Enrichment of clusters on mental health subreddits during the pre-pandemic period and the mid-pandemic period, using clusters detected during each time period, respectively. Associations were assessed with hypergeometric tests, and those displayed here passed strict Bonferroni correction for multiple hypothesis testing. Associations present only for the pre-pandemic or only for the mid-pandemic time period are shown in bold.



We identified 47 cluster-subreddit pairings (Figure 3C) for which posts from the given cluster were enriched on the given subreddit. Expected pairings were recapitulated and additional associations of interest were revealed. For example, r/addiction is enriched for posts from the “Substance Use Alcohol” and “Substance Use Marijuana” clusters, as expected. It is also one of few subreddits enriched for posts from the “Resources” and “Seeking Advice” clusters. Although cluster-subreddit associations are present, every cluster contains posts from  $\geq 12$  subreddits, emphasizing the value of clustering to define categories that span subreddits. Clusters defined by tone and unspecified clusters contain the greatest diversity of post representation across subreddits, quantified by the Shannon Index (Multimedia Appendix 1 Table S5). The distribution of posts across subreddits within four example clusters in Multimedia Appendix 1 Figure S9 illustrates the subreddits with greatest representation in the medication cluster (r/adhd, r/BipolarReddit, r/schizophrenia, and r/anxiety), the pleading

cluster (r/SuicideWatch, r/lonely, r/depression, and r/addiction), the social anxiety cluster (r/socialanxiety, r/anxiety, and r/autism), and the suicidality cluster (r/SuicideWatch, r/depression, and r/bpd).

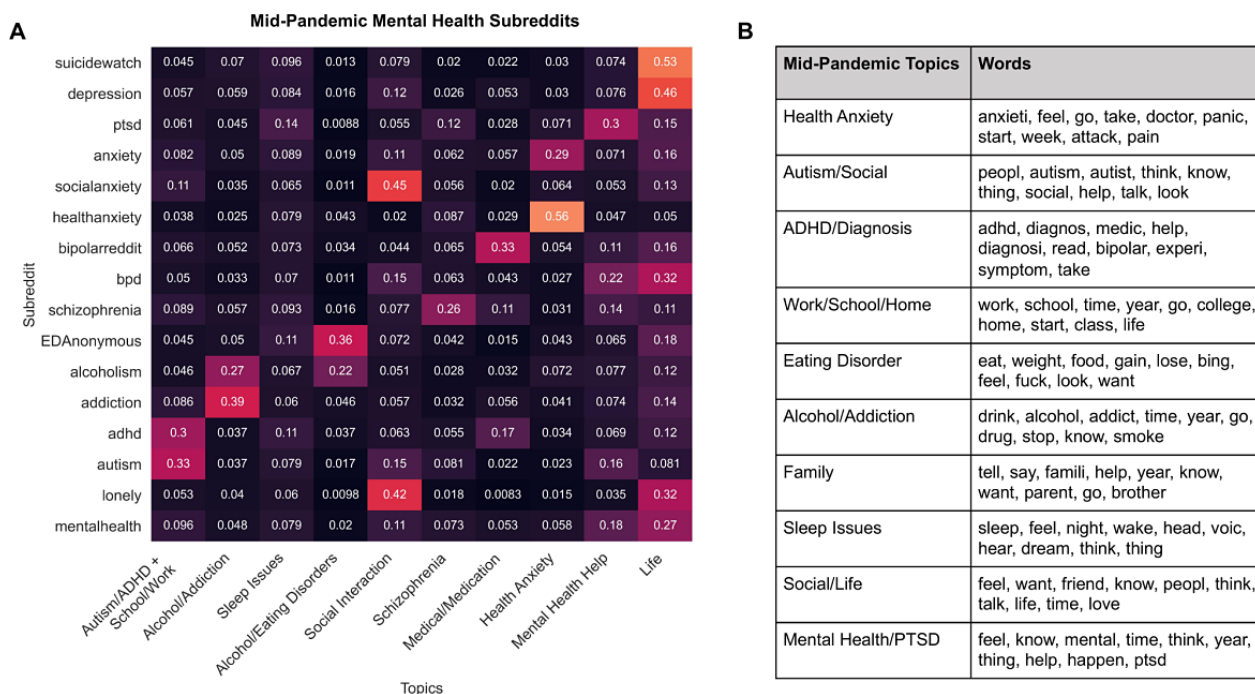
### LDA Topic Modeling on Multiple Time Frames

Figure 4 shows the distribution of posts in midpandemic mental health subreddits over the 10 topics extracted using LDA on prepandemic mental health subreddits as well as the 10 topics extracted using LDA on the midpandemic mental health subreddits. The topic number was chosen to include distinct yet important topics. Topics extracted from the prepandemic LDA model (x-axis of Figure 4A, and Multimedia Appendix 1 Table S6) largely matched the expected topics from the prepandemic subreddits. Topics emerged related to “Alcoholism and Addiction,” “Health Anxiety,” “Alcoholism and Eating Disorder,” “Schizophrenia,” and “ADHD and Autism,” corresponding to specific subreddits. The

attention-deficit/hyperactivity disorder (ADHD) and autism token words were also combined into the same topic with tokens related to school and work, and did not separate even in models with a larger number of topics. More general topics such as “Social Interaction,” “Life,” and “Mental Health Help” also emerged, which captured common topics in subreddits such as r/SuicideWatch, r/depression, r/lonely, and r/mentalhealth. Compared to the prepanemic topic model, the midpanemic LDA model splits the autism and ADHD tokens, includes a topic on family, and includes a topic with a PTSD token. [Multimedia Appendix 1](#) contains the distribution of prepanemic

model topics across mental health subreddits ([Multimedia Appendix 1](#) S10) and across non-mental health subreddits ([Multimedia Appendix 1](#) Figure S11) in both pre- and midpanemic time frames. Wilcoxon signed rank tests of distributions of the pre- and midpanemic posts over prepanemic model topics implies there was an increase in the “Health Anxiety” ( $P=.008$ ) topic and in the “Life” topic ( $P=.01$ ), as well as a decrease in the “Alcoholism and Addiction” topic ( $P=.004$ ), while the remainder of the topics showed no significant change in distribution after multiple comparison correction.

**Figure 4.** Latent dirichlet allocation (LDA) reveals prominent topics in mental health subreddits. A) Distribution of midpanemic posts from 15 mental health subreddits across 10 topics extracted using LDA on prepanemic mental health subreddit posts. Topic distribution was assessed for midpanemic posts between March 16, 2020, and April 20, 2020, to capture the phase of the pandemic right after stay-at-home orders had been announced or enacted for many areas in the United States. Inspection of the topic distribution indicated that there was minimal shift in most topics for all subreddits between the pre- and midpanemic time frames. We tested changes in topic distributions across all 27 subreddits using a Wilcoxon signed rank test (COVID19\_support was not available during 2019). B) Manually labelled topics and the top 10 terms associated with each topic derived from an LDA model created on midpanemic subreddit posts. ADHD: attention-deficit/hyperactivity disorder; PTSD: posttraumatic stress disorder.



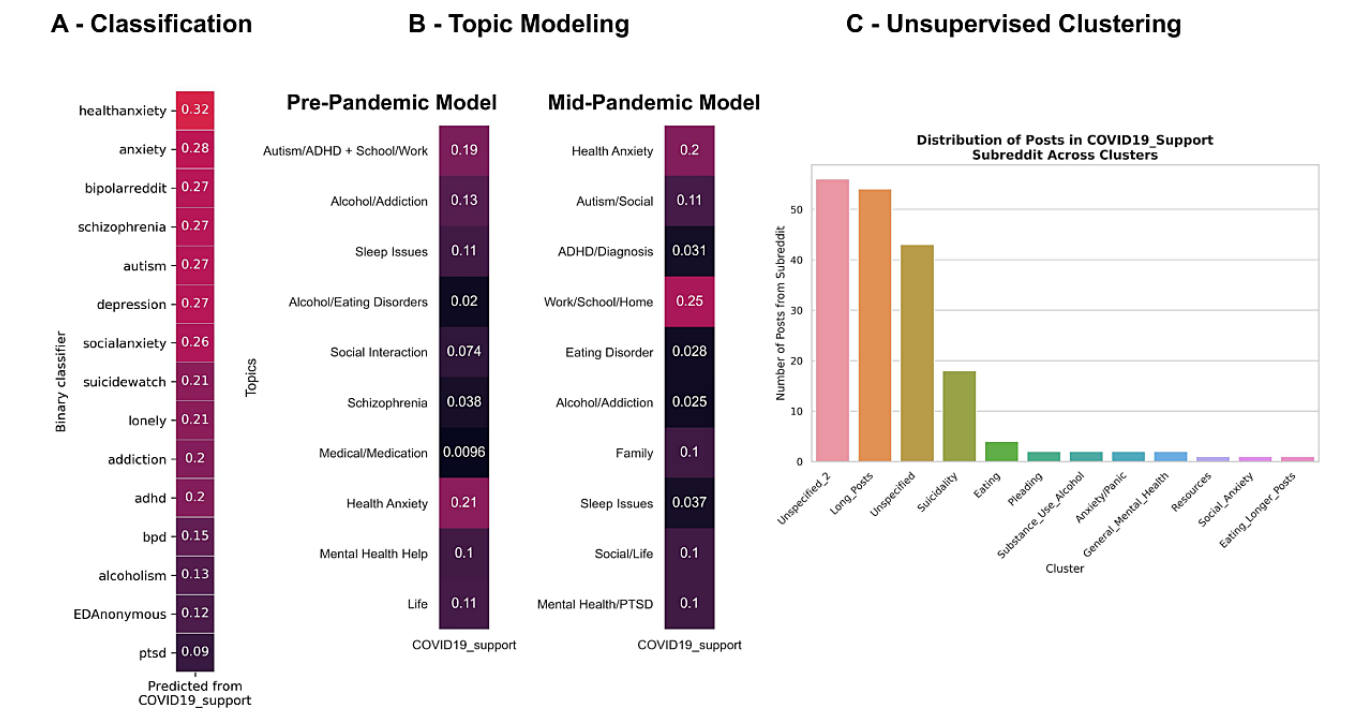
## r/COVID19\_support Characterization

[Figure 5](#) characterizes the new r/COVID19\_support subreddit through supervised and unsupervised methods. Health anxiety

emerged as a major concern through classification (ie, r/healthanxiety was classified most prevalently) and topic modeling, while unsupervised clustering revealed a substantial portion of posts were assigned to the suicidality cluster.



**Figure 5.** Characterization of r/COVID19\_support through supervised and unsupervised methods. A) Proportion of r/COVID19\_support posts (March 11 to April 20, 2020) that each binary classifier trained on prepandemic data detects. B) Distribution of prepandemic model topics (left) and midpandemic model topics (right) for posts in r/COVID19\_support, highlighting prominent topics in the posts, such as health anxiety and issues in school, work, and home scenarios. The distribution of topics indicate common themes of pain points, which could help guide the medium and content of mental health resources. C) Distribution of unsupervised cluster representation among posts from r/COVID19\_support. Although many posts were assigned to unspecified clusters, the substantial portion of posts assigned to the suicidality cluster is notable. ADHD: attention-deficit/hyperactivity disorder; EDanonymous: Eating Disorders Anonymous; PTSD: posttraumatic stress disorder.

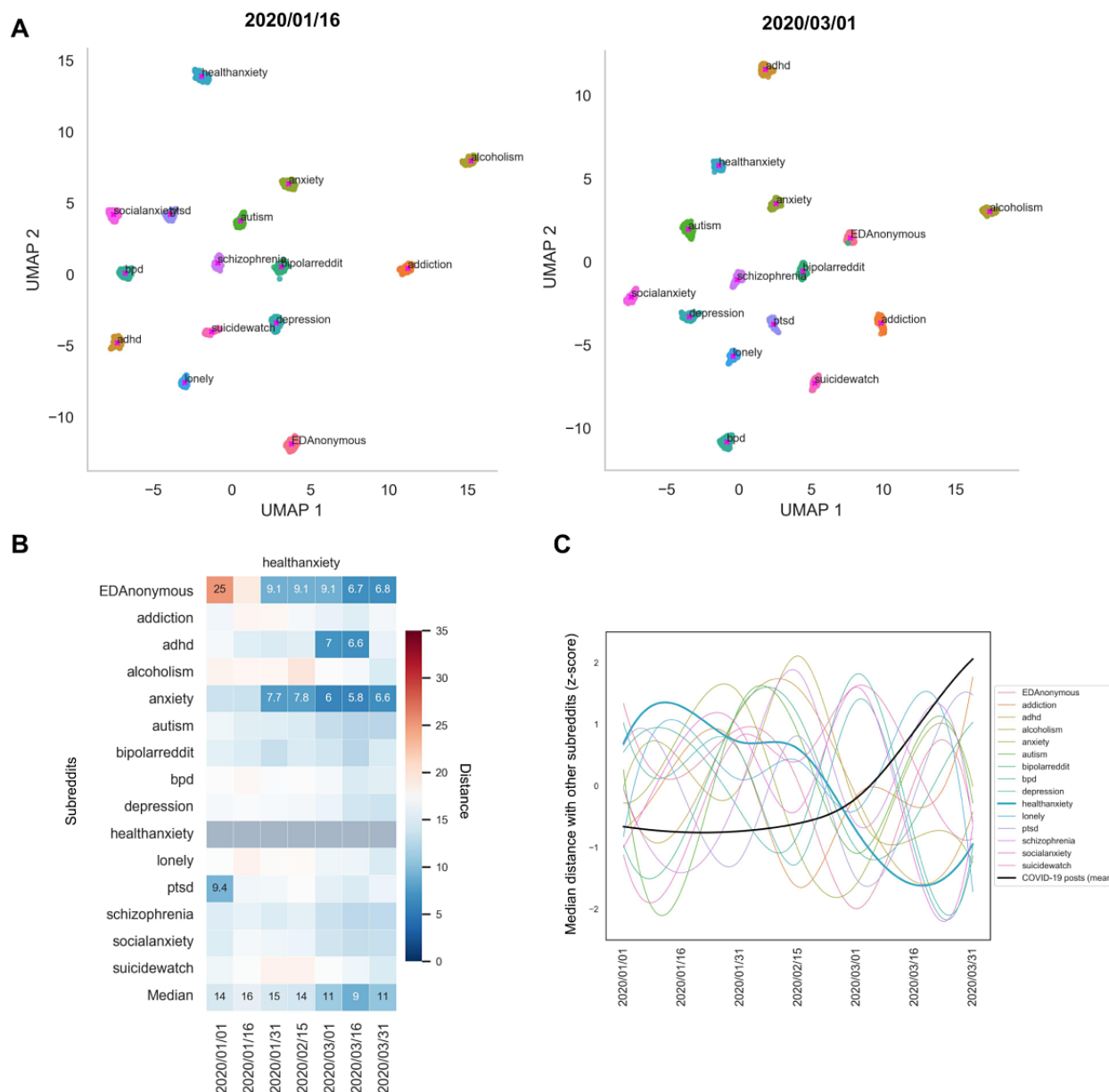


Measuring Similarity Between Subreddits Over Time With Supervised Dimensionality Reduction

The highest silhouette score (0.93) was obtained with n neighbors=200, min dist=0.0, and cosine distance. Using UMAP

and true labels, we reduced feature sets for each post to 2D and measured the directed Hausdorff distances between the clusters to quantify which subreddits are converging or diverging as the pandemic advances (see Figure 6 and Multimedia Appendix 1 Figure S12 for full results).

**Figure 6.** Supervised dimensionality reduction to measure how certain subreddits are becoming more or less similar over time. A) Supervised dimensionality reduction of posts within 15-day time windows with starting day displayed (r/healthanxiety becomes more similar to other subreddits). B) Median pairwise distance with r/healthanxiety for each time window over 50 bootstrapping samples displaying only extreme values with regard to normal 2019 fluctuations (top and bottom 5th percentiles), which indicates they are less likely to be part of normal fluctuations in distance. C) The median distance across all subreddits (last row in B) shows subreddits becoming more similar to r/healthanxiety during the increase in COVID-19-related posts (Figure 2 mean values were split into 7 time windows to match subreddit trends, and the mean was taken for each window). r/healthanxiety is the only trend that significantly correlates with COVID-19 posts after Benjamini-Hochberg multiple comparison correction ( $\rho=-0.96$ ,  $P<.001$ ). adhd: attention-deficit/hyperactivity disorder; bpd: borderline personality disorder; EDAnonymous: Eating Disorders Anonymous; ptsd: posttraumatic stress disorder; UMAP: Uniform Manifold Approximation and Projection.



## Discussion

We applied an array of NLP techniques including statistical analysis of feature trends, supervised learning, interpretability, and unsupervised learning to measure how COVID-19 may have impacted different mental health support groups. Overall, we discuss each analysis in turn, focusing on clinical takeaways.

## Trend Analysis

In our results, r/healthanxiety had several spikes in pandemic-related posts in January before other subreddits were posting about a possible pandemic (see example of post in [Textbox 1](#)). This evidence supports concerns regarding the prolonged stress that people with health anxiety may be experiencing. We then evaluated the nature of linguistic changes in mental health subreddits as the pandemic spread. Strikingly, many linguistic features significantly changed from January to April of 2020 than in the same months of previous years (see



[Multimedia Appendix 1](#) Figure S5 for full results). The amount of posts significantly decreased in multiple subreddits in 2020, as tends to happen during that time period, but the decrease was often more than in previous years. One potential reason is that some users may avoid social media since they perceive increased anxiety from news and discussions, and want to protect their mental health. However, this could be a concern if individuals are not using these subreddits to seek support and have limited access or barriers to other support structures including psychotherapy. Interestingly, the amount of posts significantly grew for r/conspiracy, which is consistent with findings from other analyses [20]. A sign of concern comes from the increase in negative semantic features for certain subreddits, the highest of which belonged to the groups for ADHD, eating disorders, anxiety, and depression within the mental health subreddits and personal finance, relationships, and fitness within the non-mental health subreddits. Regarding ADHD, some parents in France of children and adolescents diagnosed with ADHD reported increased hyperactivity and inattention, while other parents reported symptomatic improvement [21]. There is also evidence that individuals with eating disorders are experiencing worse symptoms and heightened risk for relapse, hypothesized to stem from limited care access, less structure in daily activities, and decreased social support [22], and captured in the example in [Textbox 1](#). We found an increase in the use of body-related words for r/anxiety and a decrease in r/EDAnonymous and r/fitness. Pandemic-related features (motion, isolation, economic stress, home) increased or decreased across some subreddits in the direction that would be expected for a pandemic, and the amount of change provides further evidence of concerns or lack thereof. For instance, although the isolation feature significantly increased for most subreddits, it increased most for r/anxiety and did not increase for r/socialanxiety (ie, users write less about isolation and loneliness; see r/socialanxiety example in [Textbox 1](#)).

### Classification and Feature Importance

We successfully developed binary models to classify 2019 posts originating from a given mental health subreddit as distinct from posts originating from other mental health-related subreddits. We then used these models to classify r/COVID19 support posts and identify the general distribution of complaints as well as identify posts potentially containing at-risk behavior (eg, from r/SuicideWatch). By leveraging interpretable linear methods, we were able to establish features key to distinguishing mental health subreddits from one another (see [Multimedia Appendix 1](#) Tables S4 and S5, and the y-axis of Figure S5), which helps understand how different mental health concerns may manifest in language. Some of the most interesting top important features used to classify each subreddit were gunning fog index (ie, how readable the text is) for r/depression, the tokens “cancer” and “google” for r/healthanxiety, domestic stress and guns lexicons for r/ptsd, LIWC “religion” and number of long words for schizophrenia; LIWC see (eg, “picture,” “screen,” “stare”) for r/socialanxiety, and LIWC money and certainty for r/SuicideWatch (see [Textbox 1](#) for more examples).

### Unsupervised Methods: Topic Modeling and Clustering

We established important conversation topics that span across mental health support subreddits through unsupervised methods (LDA and clustering). Unsupervised clusters such as seeking advice, resources, seeking normalization, suicidality, and medication varied in representation across the mental health subreddits and provided insight into the forms of discussion occurring. This cluster structure has utility for assessing changes in discussion on mental health and for analyzing subreddits like r/COVID19\_support, on which many posts were found to map to the suicidality cluster. r/bpd and r/ptsd became significantly enriched for posts from the suicidality cluster in the midpandemic data set. The r/anxiety subreddit became significantly enriched for posts from the health anxiety cluster, which captures the general theme of heightened health anxiety. Topic modeling found an increase in the distribution of the health anxiety topic across midpandemic posts and highlighted changes in topics between pre- and midpandemic time frames, such as the introduction of the topics social interaction and mental health help. Additionally, a large number of posts in the r/COVID19\_support group were identified to relate to the health anxiety topic and to voice concerns with daily living at home, school, and work. Posts in these topics and clusters could be important for subreddit moderators to track as they seek to cultivate a culture of support and provide effective assistance to authors in crisis (eg, tracking clusters like seeking advice, medication, suicidality, and resources) and understand the chief concerns of their communities (eg, tracking topics like sleep issues and social interaction).

### Measuring Similarity Between Subreddits Over Time With Supervised Dimensionality Reduction

Psychiatric care for patients during the pandemic should be informed by an understanding of possible convergence among some disorders, which could merit a blending of standard treatment approaches, and of possible separation of certain clusters from the rest, which could identify an at-risk population. Overall, the more users were posting about COVID-19, the more similar subreddits became to r/healthanxiety (see [Figure 5C](#)). The r/healthanxiety similarity to other subreddits was the only subreddit that significantly correlated with the rise in COVID-19-related posts. Interestingly, the subreddits that became most negative per trend analyses, r/ADHD, r/EDAnonymous, and r/Anxiety, also became most like r/healthanxiety during the general spike of COVID-19-related posts during March 2020. These results suggest a clinically testable hypothesis that the symptom of health anxiety may have increased most in the psychiatric populations corresponding to these three subreddits.

### Convergent Findings Across Analyses: Health Anxiety and Suicidality

Our findings suggest the pandemic may have induced health anxiety among several mental health and non-mental health communities given that posts on r/COVID19\_support were classified most frequently as belonging to r/healthanxiety, midpandemic posts from the r/anxiety subreddit became significantly enriched for posts from the health anxiety cluster, LDA topic analysis found the health anxiety topic significantly

increased in the midpandemic posts compared to the prepandemic posts, and supervised dimensionality reduction revealed that the more users posted about COVID-19 the more similar subreddits became to r/healthanxiety.

Suicidality was another concerning theme given that unsupervised clustering revealed the suicidality cluster doubled in size and a new cluster surrounding self-harm emerged. Notably, a substantial portion of posts from r/COVID19\_support were assigned to the suicidality cluster, and two subreddits (r/bpd and r/ptsd) became significantly associated with the suicidality cluster during the pandemic. Furthermore, 26% of posts from COVID19\_support were classified by r/SuicideWatch by its binary classifier.

### Limitations

Our study population is not characterized with formally documented clinical diagnoses, although some post authors make diagnostic claims. For example, 5% of posts in our full 2019 data set are authored by individuals who make a diagnostic claim (eg, “I have obsessive compulsive disorder”). If further application of NLP techniques can expand the set of post authors with high likelihood of clinical diagnoses, analyses could be restricted to that subset of authors without loss of sufficient post volume. Furthermore, although linguistic changes occurred during the pandemic, in this study we do not causally link any individual changes to specific events. Finally, in our study, it is possible that post authors who are more stable in their mental health state and would, therefore, have contributed more consistent post content between the prepandemic and midpandemic periods, decrease their participation in these support fora as a means of coping with the stressors of the pandemic. Alternatively, it is also possible that post authors who are more dynamic in their mental health state, and would therefore have contributed to more dramatic differences in post content between the prepandemic and midpandemic periods, decrease their participation in these support fora as a means of coping. Both forms of selection bias are possible and represent a limitation of our observational study design.

### Future Directions

Extremely risky behavior is common content in these posts, including asking for advice for restricting food in r/EDAnonymous or planning suicide in r/SuicideWatch; therefore, more clinical attention is urgently needed to provide effective resources to Reddit users. Furthermore, understanding the nature of posts from subreddits other than r/SuicideWatch (including r/COVID19\_support) that were classified as r/SuicideWatch posts or that belong to the suicidality cluster demands further research. Critically, these posts were made on subreddits without the policies and response systems used by moderators on r/SuicideWatch itself and whose authors may benefit from urgent intervention. NLP applied to Reddit posts could help direct users to tailored resources or more ideal

subreddits where they are more likely to receive support and help with the triage of moderator responses. For example, authors of posts in the seeking normalization cluster may be looking for solidarity whereas authors of posts with high use of the sleep issues LDA topic may benefit from sleep-related advice. These analyses of Reddit could also be followed by related analyses of posts on Twitter, which enables some geolocation but restricts anonymity and post length. Geographic information could, for example, be used to characterize the impact on mental health of specific policy developments like the imposition of statewide lockdowns that varied in time across geographic areas. Taking trend analysis and supervised dimensionality reduction, one can track subreddits that did not change. More research is needed to understand if certain groups were more resilient to the pandemic. Ultimately, we have found many linguistic patterns for specific mental health groups, and these patterns could be studied further in clinical settings that include formal diagnoses and more extensive covariate information (eg, racial or socioeconomic background).

### Conclusions

We performed successful classification among mental health subreddits and identified important features to understand how each mental health problem may manifest in language. We tracked features across time and observed the largest negative semantic changes during the COVID-19 pandemic for r/ADHD, r/EDAnonymous (eating disorders), and r/Anxiety, which were the same groups that become most similar to r/healthanxiety during the rise of COVID-19 posts in March. Understanding linguistic features that distinguish these communities and how language use has changed during the pandemic has generated several important hypotheses for evaluation in clinical settings that may help inform the provision of responsive care. Importantly, different analyses found increases in health anxiety and suicidality. We hope this work will bring attention to these large online communities' needs and concerns since Reddit may be the first path to treatment for many online users, and these subreddits currently lack sufficient resources. We have helped to define some of these concerns through multiple methods such as feature importance (eg, the guns lexicon is important for classifying r/ptsd), trend analysis (eg, decreased use over time of the motion lexicon in r/anxiety), unsupervised clustering (eg, a substantial number of posts on COVID19\_support mapped to a Suicidality post cluster and subreddits show discussion of topics distinct from the overall subreddit theme), topic modeling (eg, the high rate of concerns regarding home, work, and school on r/COVID19\_support). We further hope that insights from this work will deepen our understanding of mental health challenges during the pandemic, inform the provisioning of appropriate therapeutic resources, and inspire greater use of NLP for the inspection of world-changing events such as epidemics, elections, and protests.

### Acknowledgments

All data and code are available through the Open Science Framework [23] and [24]. We thank the course staff of Massachusetts Institute of Technology's (MIT's) Machine Learning for Healthcare class, who provided feedback on the project structure and

analysis, and Kate H Bentley and Mostafa Abdou who provided feedback on the manuscript. DML was supported by the NIH-NIDCD T32 (5T32DC000038-28). TT was supported by the NIH-NIDCD T32 (5T32DC000038). LR was supported by the NIH-NHGRI T32 (5T32HG2295-17). The work was supported by a gift to the McGovern Institute for Brain Research at MIT.

Approved for public release. Distribution is unlimited. This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering. © 2020 Massachusetts Institute of Technology. Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

## Authors' Contributions

DML, LR, and TT conceived the experimental design, performed data analysis, and drafted the manuscript. All authors provided feedback on analyses and reviewed the manuscript.

## Conflicts of Interest

JT reports unrelated research support from Otsuka. None of the authors declare any competing interests.

## Multimedia Appendix 1

Supplementary materials.

[PDF File (Adobe PDF File), 2258 KB - [jmir\\_v22i10e22635\\_app1.pdf](#)]

## References

- Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet* 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30460-8](#)] [Medline: [32112714](#)]
- Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. *Lancet Psychiatry* 2020 Apr;7(4):300-302 [FREE Full text] [doi: [10.1016/S2215-0366\(20\)30073-0](#)] [Medline: [32085840](#)]
- Alonzi S, La Torre A, Silverstein MW. The psychological impact of preexisting mental and physical health conditions during the COVID-19 pandemic. *Psychol Trauma* 2020 Aug;12(S1):S236-S238. [doi: [10.1037/tra0000840](#)] [Medline: [32525380](#)]
- Wardell J, Kempe T, Rapinda KK, Single AN, Bilevicius E, Frohlich JR, et al. Drinking to cope during the COVID-19 pandemic: the role of external and internal stress-related factors in coping motive pathways to alcohol use, solitary drinking, and alcohol problems. *PsyArXiv* 2020 Jun 11:1. [doi: [10.31234/osf.io/8vfp9](#)]
- Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. *Curr Opin Behav Sci* 2017 Dec;18:43-49. [doi: [10.1016/j.cobeha.2017.07.005](#)]
- Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. *Nat Lang Eng* 2017 Jan 30;23(5):649-685. [doi: [10.1017/S1351324916000383](#)]
- Shen JH, Rudzicz F. Detecting anxiety through reddit. 2017 Presented at: The Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality; August 2017; Vancouver, BC p. 58. [doi: [10.18653/v1/w17-3107](#)]
- Sekulic I, Gjurković M, Šnajder J. Not just depressed: bipolar disorder prediction on Reddit. 2018 Presented at: The 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; October 2018; Brussels, Belgium. [doi: [10.18653/v1/w18-6211](#)]
- Yates A, Cohan A, Goharian N. Depression and self-harm risk assessment in online forums. 2017 Presented at: 2017 Conference on Empirical Methods in Natural Language Processing; September 2017; Copenhagen, Denmark. [doi: [10.18653/v1/D17-1322](#)]
- Zirikly A, Resnik P, Uzuner Ö, Hollingshead K. CLPsych 2019 shared task: predicting the degree of suicide risk in Reddit posts. 2019 Presented at: The Sixth Workshop on Computational Linguistics and Clinical Psychology; June 2019; Minneapolis, Minnesota. [doi: [10.18653/v1/w19-3003](#)]
- Pennebaker JW, Booth RJ, Boyd RL, Francis ME. LIWC 2015 Operator's Manual. Austin, TX: Pennebaker Conglomerates Inc; 2015.
- De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Presented at: Seventh International AAAI Conference on Weblogs and Social Media AAAI Publications; 2013; Cambridge, Massachusetts.
- Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. *Comput Human Behav* 2018 Jan;78:98-112 [FREE Full text] [doi: [10.1016/j.chb.2017.09.001](#)] [Medline: [29456286](#)]

14. Park A, Conway M. Tracking health related discussions on Reddit for public health applications. AMIA Annu Symp Proc 2017;2017:1362-1371 [FREE Full text] [Medline: 29854205]
15. Regional distribution of desktop traffic to Reddit.com as of May 2020, by country. Statista. URL: <https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/> [accessed 2020-06-19]
16. Barthel M, Stocking G, Holcomb J, Mitchell A. Reddit news users more likely to be male, young and digital in their news preferences. Pew Research Center. 2016. URL: [https://www.pewresearch.org/wp-content/uploads/sites/8/2016/02/PJ\\_2016\\_02.25\\_Reddit\\_FINAL.pdf](https://www.pewresearch.org/wp-content/uploads/sites/8/2016/02/PJ_2016_02.25_Reddit_FINAL.pdf)
17. pushshift / api. GitHub. 05. URL: <https://github.com/pushshift/api>
18. Coronavirus (COVID-19) Cases. Our World in Data. 05. URL: <https://ourworldindata.org/covid-cases>
19. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. arXiv 2018 Feb 09:1 [FREE Full text]
20. Shahsavari S, Holur P, Tangherlini TR, Roychowdhury V. Conspiracy in the time of corona: automatic detection of Covid-19 conspiracy theories in social media and the news. arXiv 2020 Apr 28:1 [FREE Full text] [doi: 10.21203/rs.3.rs-52079/v1]
21. Bobo E, Lin L, Acquaviva E, Caci H, Franc N, Gamon L, et al. [How do children and adolescents with attention deficit hyperactivity disorder (ADHD) experience lockdown during the COVID-19 outbreak?]. Encephale 2020 Jun;46(3S):S85-S92 [FREE Full text] [doi: 10.1016/j.encep.2020.05.011] [Medline: 32522407]
22. Termorshuizen J, Watson H, Thornton L, Borg S, Flatt R, MacDermid C, et al. Early impact of COVID-19 on individuals with eating disorders: a survey of ~1000 individuals in the United States and the Netherlands. medRxiv 2020 May 29:2020.05.28.20116301. [doi: 10.1101/2020.05.28.20116301] [Medline: 32511555]
23. Low DM, Talkar T. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19. Open Science Framework. 2020 Jun 20. URL: <https://osf.io/7peyq/>
24. danielmlow / reddit. GitHub. 05. URL: <https://github.com/danielmlow/reddit>

## Abbreviations

**ADHD:** attention-deficit/hyperactivity disorder  
**LDA:** latent dirichlet allocation  
**LIWC:** Linguistic Inquiry and Word Count  
**MIT:** Massachusetts Institute of Technology  
**NLP:** natural language processing  
**PCA:** principal component analysis  
**PTSD:** posttraumatic stress disorder  
**SGD:** stochastic gradient descent linear classifier  
**TF-IDF:** term frequency-inverse document frequency  
**UMAP:** Uniform Manifold Approximation and Projection

*Edited by G Eysenbach; submitted 18.07.20; peer-reviewed by C Corcoran, M Fuller-Tyszkiewicz; comments to author 20.08.20; revised version received 05.09.20; accepted 13.09.20; published 12.10.20.*

### *Please cite as:*

Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS

Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study

J Med Internet Res 2020;22(10):e22635

URL: <http://www.jmir.org/2020/10/e22635/>

doi: [10.2196/22635](https://doi.org/10.2196/22635)

PMID: [32936777](https://pubmed.ncbi.nlm.nih.gov/32936777/)

©Daniel M Low, Laurie Rumker, Tanya Talkar, John Torous, Guillermo Cecchi, Satrajit S Ghosh. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org/>), 12.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Original Paper

# Public Opinion About E-Cigarettes on Chinese Social Media: A Combined Study of Text Mining Analysis and Correspondence Analysis

Di Wang<sup>1\*</sup>, PhD; Joanne Chen Lyu<sup>2\*</sup>, PhD; Xiaoyu Zhao<sup>1\*</sup>, MA

<sup>1</sup>Faculty of Humanities and Arts, Macau University of Science and Technology, Taipa, Macao

<sup>2</sup>Center for Tobacco Control Research and Education, University of California, San Francisco, CA, United States

\* all authors contributed equally

**Corresponding Author:**

Joanne Chen Lyu, PhD

Center for Tobacco Control Research and Education

University of California, San Francisco

530 Parnassus Ave

Suite 366

San Francisco, CA, CA 94143-1390

United States

Phone: 1 415 502 4181

Fax: 1 415 514 9345

Email: [chenjoanne.lyu@ucsf.edu](mailto:chenjoanne.lyu@ucsf.edu)

## Abstract

**Background:** Electronic cigarettes (e-cigarettes) have become increasingly popular. China has accelerated its legislation on e-cigarettes in recent years by issuing two policies to regulate their use: the first on August 26, 2018, and the second on November 1, 2019. Social media provide an efficient platform to access information on the public opinion of e-cigarettes.

**Objective:** To gain insight into how policies have influenced the reaction of the Chinese public to e-cigarettes, this study aims to understand what the Chinese public say about e-cigarettes and how the focus of discussion might have changed in the context of policy implementation.

**Methods:** This study uses a combination of text mining and correspondence analysis to content analyze 1160 e-cigarette-related questions and their corresponding answers from Zhihu, China's largest question-and-answer platform and one of the country's most trustworthy social media sources. From January 1, 2017, to December 31, 2019, Python was used to text mine the most frequently used words and phrases in public e-cigarette discussions on Zhihu. The correspondence analysis was used to examine the similarities and differences between high-frequency words and phrases across 3 periods (ie, January 1, 2017, to August 27, 2018; August 28, 2018, to October 31, 2019; and November 1, 2019, to January 1, 2020).

**Results:** The results of the study showed that the consistent themes across time were comparisons with traditional cigarettes, health concerns, and how to choose e-cigarette products. The issuance of government policies on e-cigarettes led to a change in the focus of public discussion. The discussion of e-cigarettes in period 1 mainly focused on the use and experience of e-cigarettes. In period 2, the public's attention was not only on the substances related to e-cigarettes but also on the smoking cessation functions of e-cigarettes. In period 3, the public shifted their attention to the e-cigarette industry and government policy on the banning of e-cigarette sales to minors.

**Conclusions:** Social media are an informative source, which can help policy makers and public health professionals understand the public's concerns over and understanding of e-cigarettes. When there was little regulation, public discussion was greatly influenced by industry claims about e-cigarettes; however, once e-cigarette policies were issued, these policies, to a large extent, set the agenda for public discussion. In addition, media reporting of these policies might have greatly influenced the way e-cigarette policies were discussed. Therefore, monitoring e-cigarette discussions on social media and responding to them in a timely manner will both help improve the public's e-cigarette literacy and facilitate the implementation of e-cigarette-related policies.

(*J Med Internet Res* 2020;22(10):e19804) doi:[10.2196/19804](https://doi.org/10.2196/19804)

**KEYWORDS**

e-cigarettes; public opinion; social media; infodemiology; infoveillance; regulation; China

## Introduction

### Background

Electronic cigarettes (e-cigarettes) are devices that look like real cigarettes but deliver a nicotine-containing aerosol to users by heating a solution typically made up of propylene glycol or glycerol (glycerin), nicotine, and flavoring agents [1,2]. As there are fewer carcinogens in the cartridges or vapor of e-cigarettes, they are often marketed as less harmful products or as a healthier alternative to traditional cigarettes, which can help people to quit smoking and even replace the use of traditional cigarettes [3-6]. Studies have also found that e-cigarette use is associated with less negative cardiovascular effects compared with traditional cigarettes, and there may be beneficial effects on blood pressure regulation and endothelial function for smokers who switch to e-cigarettes [7]. Some proponents have called e-cigarettes a revolutionary product in tobacco harm reduction with the potential to significantly reduce the burden of smoking-related diseases globally [8]. In contrast, opponents argue that there has been no conclusive scientific evidence that e-cigarettes are an effective tool for harm reduction compared with smoking combustible cigarettes [2,9] or that e-cigarettes can promote long-term smoking cessation [10]. On the contrary, studies have found that e-cigarette use causes adverse health effects in many organ systems, such as gastrointestinal, cardiovascular, and pulmonary system symptoms [11-14], and even brings changes in the brain that boost the risk of other drug addictions [15]. Besides the health impact, opponents noted that if an increasing number of people think that e-cigarette products are less harmful than cigarettes, this may motivate young people to use e-cigarettes [16,17]. In particular, some e-cigarettes are being promoted as a lifestyle choice and an identity statement [18]. This may make some people who would otherwise not smoke start smoking and serve as a gateway to traditional cigarettes, further prolonging the tobacco epidemic [19]. Therefore, many countries have begun to formulate policies to regulate and control e-cigarettes [20].

The uniqueness of China's tobacco industry (including e-cigarettes) makes China an especially worthwhile area to study. First, China's tobacco product market is dominated by the state-owned monopoly of China National Tobacco Corporation and State Tobacco Monopoly Administration (STMA), which are one institution with 2 names [21]. Second, China is the principal producer of e-cigarettes, manufacturing approximately 95% of the world's e-cigarettes [22]. Although the rate of e-cigarette use in China is still lower than that in many high-income countries, e-cigarettes have become increasingly popular, particularly among young people [23]. In China, e-cigarette sales reached 4.09 billion RMB (approximately US \$589 million) in 2017, an increase of 25.3% year-on-year in the consumer market [24]. However, before 2018, China had no regulations on e-cigarettes and their manufacture, distribution, sales, health warnings, packaging, or advertising [25]. The first national-level regulation on e-cigarettes in China was released in August 2018, when China's

State Administration for Market Regulation and STMA jointly issued the *Notice on Prohibiting the Sale of Electronic Cigarettes to Minors*. The notice stated that China's Minor Protection Law clearly stipulates that it is forbidden to sell tobacco and alcohol to minors and that e-cigarettes, as a supplement to cigarettes and other traditional tobacco products, have big safety and health risks. Therefore, various market entities must not sell e-cigarettes to minors [26]. On November 1, 2019, one more regulation, *Notice on Further Protecting Minors from Electronic Cigarettes* (Circular No. 1 of 2019), was issued. In addition to stressing the safety risks and health hazards of e-cigarettes, the notice also stipulated that no sales channels are allowed to sell e-cigarettes to minors and that all web-based e-cigarette advertisements should be withdrawn. At the same time, it urged e-cigarette production and sales enterprises and individuals to close e-cigarette sales websites and urged electronic commerce (e-commerce) platforms to take down e-cigarette products and close web-based e-cigarette stores [27]. Therefore, in effect, in China, e-cigarette sales and advertising were banned on the web from November 1, 2019.

### Objectives

On social media, information about e-cigarettes is spreading exponentially among audiences [28]. The public are more likely to rely on information provided by social media, especially when they are uncertain about the long-term consequences of e-cigarettes and where there is a lack of clear policies to regulate e-cigarettes [29]. Therefore, social media have been a popular means for disseminating information [30] and shaping the attitude of the public toward novel health issues [31,32]. Although potential e-cigarette consumers can easily access information on social media, they can also shape discussions around new products [33], which makes social media become both a good platform to gauge public knowledge of health issues [34,35] and an important source for the surveillance of public opinion about e-cigarettes [36-39]. Owing to the accessibility of social media, it is necessary for public health communities and governmental agencies to be aware of information that is being circulated on social media [40]. In particular, after the issuance of policies, social media can provide an efficient platform to access information that may help in understanding how the public perceive the policies; this is critical to regulatory efforts [41]. However, social media studies on public tobacco discussions in China are rare. To the best of our knowledge, there are only 2 relevant studies: one on the nature and extent of e-cigarette discussions on social media in 2013 [42], when China had not yet begun e-cigarette regulation, and one focusing on public reactions to a city-level cigarette control policy in China [43]. Therefore, by examining public conversation about e-cigarettes on social media, this study would be the first to understand what the Chinese public have said about e-cigarettes and how the focus of discussion changed in the context of different policies (ie, in the period before the first e-cigarette regulation in 2018, the period between the first regulation and the issuance of the second in 2019, and the period after the second regulation) to gain insight into how e-cigarette policies

have influenced the reaction of the Chinese public to e-cigarettes. To achieve these goals, specifically, this study will identify the most frequently used words and phrases the public used in their discussion of e-cigarettes on Chinese social media and examine the similarities and differences between the high-frequency words the public used in web-based discussion about e-cigarettes across time.

## Methods

### Combining Text Mining and Correspondence Analysis

This study used a combination of the text mining and correspondence analysis to analyze the Chinese public discussion on e-cigarettes. Text mining was used to study the ranking of high-frequency words and phrases during the 3 periods of web-based discussion. To understand the similarities and differences across time in terms of discussion focus, we also conducted the correspondence analysis to examine the relationships between 2 nominal variables, high-frequency words and periods. The results will show the relative positions of various words listed in the frequency table in the form of a perceptual map so that we can see the relationship between high-frequency words and periods. The ability to deal with frequency data provides the correspondence analysis a methodological strength, and the graphical displays provide the correspondence analysis an interpretive strength [44]. In medical research, it has been used to study how the relative frequencies of headache types change with age and the association between personality types and various medical diagnostic groups [45]. It has also been widely used in education [46], tourism [47], and many other fields.

### Data Source Selection

We chose Zhihu as the social media platform to acquire data. *Zhīhū*, which means “Do you know?” in Chinese, is China’s largest question-and-answer platform where questions are asked, answered, and edited by its community of users [48]. It has cultivated a reputation for being one of the country’s most trustworthy social media platforms [49]. Its motto is “Share your knowledge, experiences and thoughts with the world” [50]. As of January 2019, the number of Zhihu users exceeded 220 million and accumulated more than 130 million answers [51]. Unlike Weibo, the most popular microblogging platform in China, on which the posts spreading tobacco control policies were mostly from professional new media accounts [43], the posts about tobacco issues on Zhihu were mostly individual accounts. Thus, we used Zhihu as the platform to study Chinese public discussions about e-cigarettes [48].

### Data Acquisition and Preprocessing

We used Python (Python Software Foundation), a programming language, to retrieve information about e-cigarettes on Zhihu

in January 2020. In total, 2 Chinese keywords, “Dian Zi Yan” and “Dian Zi Xiang Yan” (both mean e-cigarette in Chinese), were used to identify all the questions related to e-cigarettes and their corresponding answers through to December 31, 2019, when the study was completed. Owing to the limited number of questions and discussions on e-cigarettes before 2017, our sample selection was from January 1, 2017, to December 31, 2019, with a total of 3275 questions and their answers. Questions about the promotion of a specific e-cigarette brand and questions about how to choose an e-cigarette brand were deleted along with their answers. The removal of advertisements and e-cigarette selection strategies resulted in 1160 questions and their corresponding answers. Next, we used *Jieba* word segmentation software to segment Chinese words and phrases, as there are no spaces in Chinese sentences. In addition, we combined the frequencies of synonyms and filtered out meaningless demonstrative pronouns, conjunctions, and degree adverbs. According to the issue time of China’s 2 e-cigarette policies, we divided the sample into 3 periods. The sample size for the 3 periods was 208 questions and their answers for period 1, 473 questions and their answers for period 2, and 479 questions and their answers for period 3.

## Results

### Text Mining and Analysis

According to the issuing time of China’s 2 e-cigarette policies (August 28, 2018, and November 1, 2019) [26,27], we divided the sample into 3 periods (January 1, 2017, to August 27, 2018; August 28, 2018, to October 31, 2019; and November 1, 2019, to December 31, 2019).

We first ran the word frequency for the 3 periods separately and listed the top 50 words for each period (Table 1). Among the top 50 frequently mentioned words, “e-cigarette,” “nicotine,” “tobacco tar,” “cigarette,” “smoking cessation,” “smoking,” “harm,” and “tobacco” appeared in all 3 stages. To further our understanding of the relationships between the keywords, we searched the above words in the 1160 questions and found that 2.9% (6/209) of questions in period 1, 10.8% (51/473) of questions in period 2, and 10.5% (50/478) of questions in period 3 were related to the comparisons between e-cigarettes and traditional cigarettes. At the same time, 22.9% (48/209) of questions in period 1, 20.9% (51/473) of questions in period 2, and 18.2% (50/478) of questions in period 3 asked about the harm of e-cigarettes. In addition, 19.6% (41/209) of questions in period 1, 15.6% (74/473) of questions in period 2, and 3.7% (18/478) of questions in period 3 asked about whether e-cigarettes can effectively help smokers quit smoking. Other common words appearing in the 3 stages were e-cigarette-related substances such as “tobacco tar,” “cartridges,” and “atomizer.”



**Table 1.** Top 50 high-frequency words and phrases in different periods.

Rank	Period 1	Period 2	Period 3
1	e-cigarette <sup>a</sup>	e-cigarette	e-cigarette
2	Nicotine	cigarette	cigarette
3	Tar	nicotine	tobacco
4	Cigarette	smoking cessation	harm
5	smoking cessation	smoking	nicotine
6	smoking	harm	smoking
7	Harm	tobacco	country
8	Tobacco	tar	problem
9	Use	problem	channel
10	Smog	health	product
11	Taste	product	industry
12	Health	country	sale
13	Product	use	offline
14	Problem	smoker	brand
15	Smoker	market	market
16	Market	smog	tar
17	Produce	tar	minor
18	Hong Kong	produce	health
19	Addition	tradition	forbid
20	Flavor	America	Cartridges
21	Atomizer	industry	physical store
22	low voltage e-cigarettes	influence	buy
23	Mouthfeel	body	online
24	Tar	burn	smoking cessation
25	Influence	research	smoker
26	Body	low voltage e-cigarettes	supervise
27	Domestic	China	America
28	Personal	brand	tradition
29	Choose	component	China
30	Content	atomize	domestic
31	Research	heat	online
32	Component	content	policy
33	Essence	cartridge	sales prohibition
34	passive smoking	choose	price
35	Tradition	forbid	influence
36	Burn	domestic	e-commerce <sup>b</sup>
37	Brand	formaldehyde	flavor
38	Atomize	company	company
39	Cartridge	addition	use
40	Advice	data	user
41	Equipment	user	enterprise

Rank	Period 1	Period 2	Period 3
42	Steam	taste	choose
43	Forbid	supervise	produce
44	Glycerinum	personal	announcement
45	Like	material	protect
46	Recommend	addition	teenager
47	Harmless	friend	advice
48	Country	flavor	develop
49	Price	Hong Kong	authentic
50	Friend	harmless	benefit

<sup>a</sup>e-cigarette: electronic cigarette.

<sup>b</sup>e-commerce: electronic commerce.

In addition to the common ground, the focus of discussion also changed in the 3 periods. In period 1, the public discussions were purely focused on the use of e-cigarettes. For example, *flavor* and *mouthfeel* were discussed more frequently in period 1 than in periods 2 and 3. In period 2, words such as *industry* and *company* were frequently mentioned, which showed that with the release of e-cigarette policy in the second period, relevant departments began to increase supervision and the public began to pay attention to the future development of the e-cigarette industry. Despite these regulations, many people still believe that China's e-cigarette market has great potential. In the third period, with the release of the notice on protecting minors from e-cigarettes, discussions about the sales of e-cigarettes and the protection of minors became increasingly heated. Words such as *sales*, *policy*, *minors*, and *teenagers* began to appear frequently. Text mining results showed that as the policy changed, the public discussion on e-cigarettes also changed.

It is worth noting that some geographical terms appeared in different periods. For example, *Hong Kong* appeared in the first stage, whereas *China* and *America* appeared in the second and third stages. In period 1, Hong Kong banned the sale of e-cigarettes, which led to the discussion on "How to evaluate the impact of banning e-cigarettes in Hong Kong?" In period

2, the first case of death related to e-cigarette use occurred in the United States, and several states banned the sale of e-cigarettes, leading to the discussion on the impact of these events. Zhihu users often compared e-cigarette policies between China and other places.

In addition, verbs that appeared frequently are also worth noting. Frequently used verbs across the 3 periods were the "use" of e-cigarettes and how to "choose" e-cigarettes. The finding that "use" ranked ninth in the first period, 13th in the second period, and 39th in the third period reflected the declining public interest in discussing the use of e-cigarettes. In addition, there were some verbs that only appeared in the third stage, such as "forbid," "protect," and "supervise." It showed that in the third stage, the public began to pay attention to the meaning of the e-cigarette policies and the measures of the policies.

### Correspondence Analysis

We used SPSS version 25 for the correspondence analysis. The first step is to enter the frequency (the number of times a word is used in the period/total number of words used in the period) of words and phrases in each period as a contingency table, which is a two-way table with the 50 high-frequency words and phrases as row headings and the 3 periods as column headings. Table 2 is a contingency table.

**Table 2.** The 3×50 contingency table.

Words and phrases	Period 1 frequency, n (%)	Period 2 frequency, n (%)	Period 3 frequency, n (%)
e-cigarette <sup>a</sup>	3308 (4.30)	13015 (4.96)	6266 (5.65)
cigarette	821 (1.07)	4101 (1.56)	1449 (1.31)
nicotine	1185 (1.54)	2933 (1.12)	740 (0.67)
harm	693 (0.90)	2942 (1.12)	772 (0.70)
smoking	803 (1.04)	2532 (0.96)	701 (0.63)
smoking cessation	812 (1.05)	2692 (1.03)	348 (0.31)
tobacco	460 (0.60)	1837 (0.70)	1098 (0.99)
tar	912 (1.18)	1466 (0.56)	575 (0.52)
problem	292 (0.38)	1181 (0.45)	626 (0.56)
product	297 (0.39)	1017 (0.39)	618 (0.56)
health	298 (0.39)	1119 (0.43)	466 (0.42)
country	116 (0.15)	864 (0.33)	690 (0.62)
market	198 (0.26)	711 (0.27)	576 (0.52)
use	352 (0.46)	861 (0.33)	232 (0.21)
smoker	201 (0.26)	801 (0.31)	327 (0.29)
taste	344 (0.45)	656 (0.25)	255 (0.23)
industry	0 (0.00)	569 (0.22)	616 (0.56)
brand	136 (0.18)	491 (0.19)	594 (0.54)
smog	348 (0.45)	665 (0.25)	0 (0.00)
tradition	143 (0.19)	621 (0.24)	301 (0.27)
America	0 (0.00)	616 (0.23)	303 (0.27)
forbid	124 (0.16)	438 (0.17)	457 (0.41)
cartridge	130 (0.17)	458 (0.17)	402 (0.36)
tar	167 (0.22)	654 (0.25)	0 (0.00)
influence	167 (0.22)	537 (0.20)	246 (0.22)
sale	0 (0.00)	0 (0.00)	601 (0.54)
produce	193 (0.25)	633 (0.24)	193 (0.17)
China	0 (0.00)	515 (0.20)	294 (0.27)
low voltage e-cigarette	172 (0.22)	519 (0.20)	0 (0.00)
body	166 (0.22)	532 (0.20)	0 (0.00)
choice	162 (0.21)	444 (0.17)	201 (0.18)
channel	0 (0.00)	0 (0.00)	624 (0.56)
minor	0 (0.00)	0 (0.00)	562 (0.51)
component	150 (0.19)	489 (0.19)	0 (0.00)
content	159 (0.21)	461 (0.18)	0 (0.00)
burn	137 (0.18)	527 (0.20)	0 (0.00)
research	153 (0.20)	523 (0.20)	0 (0.00)
flavor	176 (0.23)	326 (0.12)	241 (0.22)
atomize	133 (0.17)	479 (0.18)	0 (0.00)
company	0 (0.00)	408 (0.16)	238 (0.21)
supervise	0 (0.00)	343 (0.13)	325 (0.29)
user	0 (0.00)	371 (0.14)	235 (0.21)

Words and phrases	Period 1 frequency, n (%)	Period 2 frequency, n (%)	Period 3 frequency, n (%)
addiction	0 (0.00)	406 (0.15)	0 (0.00)
heat	0 (0.00)	465 (0.18)	0 (0.00)
offline	0 (0.00)	0 (0.00)	600 (0.54)
advice	129 (0.17)	0 (0.00)	175 (0.16)
friends	110 (0.14)	327 (0.12)	0 (0.00)
data	0 (0.00)	387 (0.15)	0 (0.00)
atomizer	172 (0.22)	0 (0.00)	0 (0.00)
teenager	0 (0.00)	0 (0.00)	177 (0.16)
sum	76,976 (14.32)	26,2491 (19.8)	11,0895 (20.85)

<sup>a</sup>e-cigarette: electronic cigarette.

On the basis of the number of categories ( $k$ ) in the columns of the contingency table, the correspondence analysis extracts  $k-1$  latent variables, also called *dimensions* [52]. In our study, the correspondence analysis of e-cigarette discussions in the 3 periods showed a two-dimensional solution (Table 3). Similar to the principal component analysis, the first dimension explains

as much variance as possible, and the second dimension is orthogonal to the first and explains as much of the remaining variance as possible [53]. A chi-square test revealed a value of 0.1 ( $df=98$ ), with a  $P$  value of .03, which rejects the null hypothesis of no association between the 2 dimensions.

**Table 3.** Summary table of the correspondence analysis.

Dimension	Singular value	Inertia	Chi-square ( $df$ )	$P$ value	Proportion of inertia	
					Accounted for	Cumulative
1	0.428	0.183	N/A <sup>a</sup>	N/A	0.817	0.817
2	0.203	0.041	N/A	N/A	0.183	1.000
Total	N/A	0.224	0.1 (98)	.03	1.000	1.000

<sup>a</sup>N/A: not applicable.

The singular values can be viewed as the correlation between the rows and columns of the contingency table and are similar to the Pearson correlation coefficient in correlation analysis [53]. They should be greater than 0.2 to be accepted as feasible dimensions [54]. Inertia is an indicator of how much of the variation in the original data is retained in the dimensional solution [55]. The singular value and the inertia are directly related,  $\text{inertia} = \text{singular value}^2$  [44]. For example, the inertia of dimension 1 is  $0.428 \times 0.428 = 0.183$ , which means that the first dimension accounts for 18.3% of the total variability. The second dimension accounts for 4.1% of the total variability, and the total inertia for the whole solution is 2.24%.

The cumulative column shows the proportion of the inertia accounted for by the latent variable. In our case, the first dimension accounts for 81.7% of the total variability, and the 2 dimensions account for 100% of the total variability.

Figure 1 shows the correspondence analysis map that displayed 2 sets of variables, *time periods* and *words and phrases*. The distribution of words was relatively concentrated, and most of them were close to the origin. The closer the words are to the origin, the less distinct they are among the 3 periods. In other words, these words can be considered as the common ground for public discussion of e-cigarettes in the 3 periods. Among them, “e-cigarette,” “health,” “smoker,” “problem,” “tradition,” “influence,” “choice,” “product,” “tobacco,” “harm,” and

“smoking” were closed to the origin, which showed that they appeared in similar frequency in the 3 periods.

Specifically, “e-cigarette,” “tradition,” “tobacco,” and “smoking” were similarly represented in the 3 periods. Combining the results from content analysis, we can see that the comparison between e-cigarettes and traditional cigarettes was a common topic. Meanwhile, “health,” “problem,” “influence,” and “harm” were common in all 3 periods. Combining the results from the content analysis, it showed that the public were continually concerned about how e-cigarettes can harm health. In addition, “choice” and “product” appeared with similar frequency in the 3 periods, which showed that the discussion of how to choose e-cigarette products was another common topic across time.

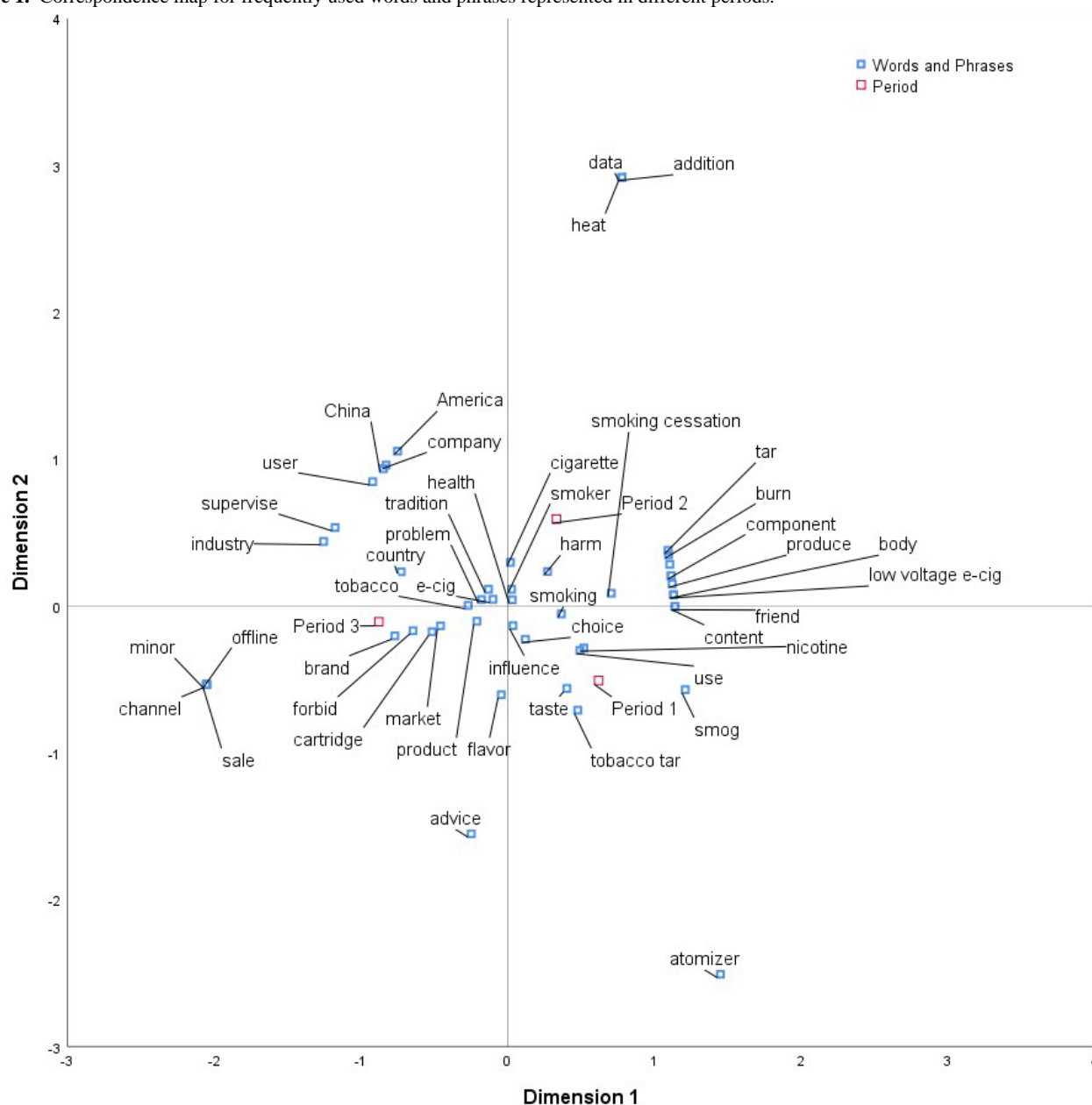
In addition to these similarities, the spots representing the 3 periods appear in 3 completely different directions, revealing differences in the public’s concerns about e-cigarettes in different periods. The further words are from the origin, the more discriminating they are [56]. To interpret the relationship between row and column labels, we need to examine (1) the length of the line connecting the row label to the origin, with longer lines indicating that the association between the row label and some of the column labels is high; (2) the length of the label connecting the column label to the origin, with longer lines again showing that the association between the column label and some of the row labels is high; and (3) the angle

formed between the above mentioned 2 lines, with smaller angles indicating that there is a stronger association between the two, a 90° angle indicating no relationship and angles near 180° indicating a negative association [57]. In our case, the row label is *words and phrases*, and the column label is *period*. If a word or phrase meet the abovementioned 3 standards, that is, the length between row label *period* and the origin is relatively long, the length between that word or phrase is relatively long, and the angle between the 2 lines is relatively small, we can say that the word or phrase is a distinct word or phrase that characterizes that period.

Period 1 appeared in the fourth quadrant. Words that met the abovementioned 3 standards were “atomizer,” “tobacco tar,” “smog,” “taste,” “use,” and “nicotine.” Namely, period 1 was characterized by words that reflect elements related to the “use”

of e-cigarettes. The public in period 1 concentrated their attention on the identified substances found in e-cigarettes and focused on their feelings of using e-cigarettes. Period 2 appeared in the first quadrant. Words such as “addition,” “heat,” “data,” “tar,” “burn,” “smoking cessation,” and “America” met the abovementioned 3 standards, which showed that in period 2, the public’s attention was not only on the identified substances of e-cigarettes but also on the smoking cessation functions of e-cigarettes. Period 3 appeared in the third quadrant. Words that met the abovementioned 3 standards were “channel,” “offline,” “sale,” “minor,” “brand,” “forbid,” “cartridge,” “market,” “industry,” and “supervise.” This showed that the public paid more attention to the e-cigarette industry and government policy on banning the sale of e-cigarettes to minors at this stage, which was quite different from the previous 2 periods.

**Figure 1.** Correspondence map for frequently used words and phrases represented in different periods.



## Discussion

### Principal Findings

The rapid popularity of e-cigarettes and the acceleration of regulation has not only attracted media attention [58] but has also generated more discussion on social media. This study used a text mining approach to identify what the Chinese public opinion about e-cigarettes on Zhihu, a social media platform featuring rich discussion. To gain insight into how e-cigarette policies have influenced the reaction of the Chinese public to e-cigarettes, this study further explored the constant focus of public discussion and how this focus changed in the context of different e-cigarette regulations.

The analysis revealed that when talking about e-cigarettes, the consistent themes were comparisons with traditional cigarettes, health concerns, and how to choose e-cigarette products. This may be closely related to the fact that e-cigarettes were marketed as “healthy alternatives to traditional cigarettes” and “the gospel of smokers” when they entered the Chinese market [22]. Although these claims have changed from earlier assertions to open questions in recent years, the Chinese public still associate e-cigarettes with these claims. Although the high-frequency words used in public discussion in the 3 periods had something in common, the differences were greater. The analyses indicated that the issuance of government policies on e-cigarettes led to public discussion on e-cigarettes. Before the first e-cigarette regulation was released in 2018, the discussion of e-cigarettes mainly revolved around the use and experience of e-cigarettes and what e-cigarettes were. Until this period, although e-cigarettes had been available on the market in China for many years (largely because of the lack of regulations or marketing restrictions on e-cigarettes), public understanding and discussion of e-cigarettes was mainly based on industry claims, such as helping to quit smoking, not containing tar, and providing a similar smoking experience to traditional cigarettes but being less harmful than them [22]. In addition, it is worth noting that many discussions in this period came from e-cigarette users. When Zhihu users raised questions about e-cigarettes, these e-cigarette users actively shared their own feelings and experiences and facilitated public conversation.

The second period (ie, the time between the first release of a ban on the sale of e-cigarettes to minors until the second and more comprehensive e-cigarette ban was issued) had many similarities with the first period. For example, many people still focused on e-cigarettes per se in their discussion. Common topics included whether e-cigarettes were an effective smoking cessation tool. However, the promulgation of the *Notice on Prohibiting the Sale of Electronic Cigarettes to Minors* routed the public's attention to the possible negative effects of e-cigarettes on health, and although “harm” was still a hot topic in this period as in the first period, the core of the harm discussion shifted from the possible harm reduction effect of e-cigarettes in the first period to the possible harm caused by e-cigarettes in the second period. Furthermore, public discussions on e-cigarettes at this stage were more in depth. The notice *triggered* discussion on an important issue, the regulation of e-cigarettes by the government. Although the first

commercially successful e-cigarette was widely considered to be invented by Chinese pharmacist Hon Lik and China is now the largest producer of e-cigarettes [22], e-cigarettes have always lacked a clear definition in China. According to the Law of the People's Republic of China on Tobacco Monopoly, e-cigarettes are not tobacco products; they were not in the regulatory scope of either tobacco monopoly law or the state Food and Drug Administration [59]. As stationery stores near primary and secondary schools were reportedly selling e-cigarettes to students [60], the public thought that it was time to regulate the e-cigarette industry, especially for the young vulnerable population. Meanwhile, they began to reflect on why the e-cigarette legislation was lagging. Some people attributed it to the state-owned monopoly of the tobacco industry and the government's interests from gigantic tobacco taxation [61], including e-cigarette taxation. In this period, the United States also progressed to implementing laws to regulate e-cigarettes, and in July 2019, it launched a new policy requiring an application for deemed *new* tobacco products that were on the market, including e-cigarettes, to be submitted to the Food and Drug Administration [62]. Consequently, the United States easily became a reference that was frequently mentioned when the Chinese public began to talk about the legislation of e-cigarettes in China.

Although there were differences in the focus of discussion among the 3 periods, the biggest difference was between the third period and the former 2 periods. The discussion in the third period jumped far beyond e-cigarettes as a product in itself but put it in a larger social context. In addition to regulatory issues, public discussion also focused on the market, sales, distribution, and other issues regarding the e-cigarette business and industry. This was consistent with the content of the notice and media attention to the policy at that time. The release of the Notice on Further Protecting Minors from Electronic Cigarettes brought forward strict law enforcement, which required various market players to stop selling e-cigarettes to minors and to remove e-cigarette advertisements on the web, and urged e-commerce platforms to close e-cigarette sales channels. In addition, the focus of discussion in this period is very likely connected with mass media reports at that time. A recent analysis of e-cigarette reporting in Chinese newspapers found that an increasing number of news articles referenced policy to frame their reporting in recent years [58]. The findings of this study echoed this analysis, confirming the positive correlation between news coverage on particular issues and public concern about the issues, as found in previous studies [63,64]. A large volume of media reports on the new policies meant that e-cigarettes were no longer just a topic that only e-cigarette users or those who were curious about e-cigarettes had an interest in but a social issue attracting attention from the general public and examined through multiple perspectives. Although the then newly released policy set the agenda for public discussion at that time, media reporting might also have increased the influence of the policy on what the public said about e-cigarettes on social media.

### Limitations

First, it should be cautious to generalize the findings of this study to the whole population of China. On Zhihu, the users



basically *speak* as ordinary netizens not opinion leaders [65], which can facilitate public expression. However, as on any social media platform, what we can see on Zhihu are the discussions of those who would like to express themselves on the web; therefore, their viewpoints cannot fully represent those who keep silent and invisible on social media. Second, the focus of this study is high-frequency words in public discussion; therefore, we did not measure emotion, attitude, or opinions. To fully understand the public's reaction to e-cigarette regulation, future studies that examine the emotional and attitudinal dimensions of public discussion are highly recommended.

## Conclusions

Social media provide an accessible and informative platform to help policy makers and public health professionals understand the public's concerns over and understanding of health issues, especially the issue of e-cigarettes, where the long-term impacts on health are still uncertain and legislation is still in the early stages. Awareness and monitoring of discussions relevant to e-cigarettes on social media in a timely manner is conducive to the identification of areas where policies need to regulate but have not yet regulated. Meanwhile, if public misunderstandings can be discovered from web-based discussion and guidance and education can be conducted quickly through media or education campaigns, this could help improve the public's e-cigarette literacy and facilitate the implementation of e-cigarette-related policies.

## Acknowledgments

This research was funded by Macau Higher Education Fund, Grant ID: HSS-MUST-2020-05 (DW); Macau University of Science and Technology Faculty Research Grants, Grant ID: FRG-20-002-FA (DW); and National Cancer Institute Grant T32 CA 113710 (JL).

## Conflicts of Interest

None declared.

## References

1. Etter J, Bullen C, Flouris AD, Laugesen M, Eissenberg T. Electronic nicotine delivery systems: a research agenda. *Tob Control* 2011 May;20(3):243-248 [FREE Full text] [doi: [10.1136/tc.2010.042168](https://doi.org/10.1136/tc.2010.042168)] [Medline: [21415064](https://pubmed.ncbi.nlm.nih.gov/21415064/)]
2. Grana R, Benowitz N, Glantz SA. E-cigarettes. *Circulation* 2014 May 13;129(19):1972-1986. [doi: [10.1161/circulationaha.114.007667](https://doi.org/10.1161/circulationaha.114.007667)]
3. Cahn Z, Siegel M. Electronic cigarettes as a harm reduction strategy for tobacco control: a step forward or a repeat of past mistakes? *J Public Health Policy* 2011 Feb;32(1):16-31. [doi: [10.1057/jphp.2010.41](https://doi.org/10.1057/jphp.2010.41)] [Medline: [21150942](https://pubmed.ncbi.nlm.nih.gov/21150942/)]
4. Yao T, Jiang N, Grana R, Ling PM, Glantz SA. A content analysis of electronic cigarette manufacturer websites in China. *Tob Control* 2016 Mar;25(2):188-194 [FREE Full text] [doi: [10.1136/tobaccocontrol-2014-051840](https://doi.org/10.1136/tobaccocontrol-2014-051840)] [Medline: [25335902](https://pubmed.ncbi.nlm.nih.gov/25335902/)]
5. Cobb NK, Byron MJ, Abrams DB, Shields PG. Novel nicotine delivery systems and public health: the rise of the 'e-cigarette'. *Am J Public Health* 2010 Dec;100(12):2340-2342. [doi: [10.2105/AJPH.2010.199281](https://doi.org/10.2105/AJPH.2010.199281)] [Medline: [21068414](https://pubmed.ncbi.nlm.nih.gov/21068414/)]
6. Carr ER. E-cigarettes: facts, perceptions, and marketing messages. *Clin J Oncol Nurs* 2014 Feb;18(1):112-116. [doi: [10.1188/14.CJON.112-116](https://doi.org/10.1188/14.CJON.112-116)] [Medline: [24476733](https://pubmed.ncbi.nlm.nih.gov/24476733/)]
7. Kuntic M, Hahad O, Daiber A, Münzel T. Could e-cigarette vaping contribute to heart disease? *Expert Rev Respir Med* 2020;1-9. [doi: [10.1080/17476348.2020.1807332](https://doi.org/10.1080/17476348.2020.1807332)] [Medline: [32757856](https://pubmed.ncbi.nlm.nih.gov/32757856/)]
8. Farsalinos KE, Polosa R. Safety evaluation and risk assessment of electronic cigarettes as tobacco cigarette substitutes: a systematic review. *Ther Adv Drug Saf* 2014;5(2):67-86. [doi: [10.1177/2042098614524430](https://doi.org/10.1177/2042098614524430)] [Medline: [25083263](https://pubmed.ncbi.nlm.nih.gov/25083263/)]
9. Leventhal AM, Strong DR, Kirkpatrick MG, Unger JB, Sussman S, Riggs NR, et al. Association of electronic cigarette use with initiation of combustible tobacco product smoking in early adolescence. *J Am Med Assoc* 2015;314(7):700-707 [FREE Full text] [doi: [10.1001/jama.2015.8950](https://doi.org/10.1001/jama.2015.8950)] [Medline: [26284721](https://pubmed.ncbi.nlm.nih.gov/26284721/)]
10. Marketers of Electronic Cigarettes Should Halt Unproved Therapy Claims. World Health Organization. 2018. URL: <https://www.who.int/mediacentre/news/releases/2008/pr34/en/> [accessed 2020-03-23]
11. Gaur S, Agnihotri R. Health effects of trace metals in electronic cigarette aerosols-a systematic review. *Biol Trace Elem Res* 2019;188(2):295-315. [doi: [10.1007/s12011-018-1423-x](https://doi.org/10.1007/s12011-018-1423-x)] [Medline: [29974385](https://pubmed.ncbi.nlm.nih.gov/29974385/)]
12. Pisinger C, Døssing M. A systematic review of health effects of electronic cigarettes. *Preventive medicine* 2020;69:248-260. [doi: [10.1093/eurpub/cku164.039](https://doi.org/10.1093/eurpub/cku164.039)]
13. Orellana-Barrios MA, Payne D, Mulkey Z, Nugent K. Electronic cigarettes—a narrative review for clinicians. *Am J Med* 2015 Jul;128(7):674-681. [doi: [10.1016/j.amjmed.2015.01.033](https://doi.org/10.1016/j.amjmed.2015.01.033)] [Medline: [25731134](https://pubmed.ncbi.nlm.nih.gov/25731134/)]
14. Layden JE, Ghinai I, Pray I, Kimball A, Layer M, Tenforde MW, et al. Pulmonary illness related to e-cigarette use in Illinois and Wisconsin - final report. *N Engl J Med* 2020 Mar 5;382(10):903-916. [doi: [10.1056/NEJMoa1911614](https://doi.org/10.1056/NEJMoa1911614)] [Medline: [31491072](https://pubmed.ncbi.nlm.nih.gov/31491072/)]

15. Kandel ER, Kandel DB. A molecular basis for nicotine as a gateway drug. *N Engl J Med* 2014 Nov 20;371(21):2038-2039. [doi: [10.1056/NEJMc1411785](https://doi.org/10.1056/NEJMc1411785)] [Medline: [25409384](#)]
16. Barrington-Trimis JL, Berhane K, Unger JB, Cruz TB, Huh J, Leventhal AM, et al. Psychosocial factors associated with adolescent electronic cigarette and cigarette use. *Pediatrics* 2015 Aug;136(2):308-317 [FREE Full text] [doi: [10.1542/peds.2015-0639](https://doi.org/10.1542/peds.2015-0639)] [Medline: [26216326](#)]
17. Pepper JK, Brewer NT. Electronic nicotine delivery system (electronic cigarette) awareness, use, reactions and beliefs: a systematic review. *Tob Control* 2014 Sep;23(5):375-384 [FREE Full text] [doi: [10.1136/tobaccocontrol-2013-051122](https://doi.org/10.1136/tobaccocontrol-2013-051122)] [Medline: [24259045](#)]
18. de Andrade M, Hastings G, Angus K. Promotion of electronic cigarettes: tobacco marketing reinvented? *Br Med J* 2013;f7473. [doi: [10.1136/bmj.f7473](https://doi.org/10.1136/bmj.f7473)] [Medline: [24361526](#)]
19. Cobb NK, Abrams DB. E-cigarette or drug-delivery device? Regulating novel nicotine products. *N Engl J Med* 2011;365(3):193-195. [doi: [10.1056/NEJMp1105249](https://doi.org/10.1056/NEJMp1105249)] [Medline: [21774706](#)]
20. Vaporizers, e-cigarettes, and other Electronic Nicotine Delivery Systems (ENDS). US Food and Drug Association. 2020 Sep 17. URL: <https://www.fda.gov/tobacco-products/products-ingredients-components/vaporizers-e-cigarettes-and-other-electronic-nicotine-delivery-systems-ends> [accessed 2020-09-23]
21. He P, Takeuchi T, Yano E. An overview of the China national tobacco corporation and state tobacco monopoly administration. *Environ Health Prev Med* 2013 Jan;18(1):85-90 [FREE Full text] [doi: [10.1007/s12199-012-0288-4](https://doi.org/10.1007/s12199-012-0288-4)] [Medline: [22696197](#)]
22. Eriksen MP, Mackay J, Ross H. The Tobacco Atlas. Atlanta, GA: American Cancer Society; 2012.
23. Hennelly W. Market Heats Up for E-cigarettes, Largely Produced in China, as US Scrutiny Rises. *China Daily*. 2015 Jun 11. URL: [http://usa.chinadaily.com.cn/opinion/2015-06/11/content\\_20967272.htm](http://usa.chinadaily.com.cn/opinion/2015-06/11/content_20967272.htm) [accessed 2020-04-02]
24. Li L, Zhou NB, Qu XH. Market Development and Legal Supervision of New Tobacco Products. *China Tobacco Journal* 2018;24(02):104-114.
25. Juan S. China Plans to Implement E-cigarette Regulations. *China Daily*. 2015 Mar 21. URL: [http://english.www.gov.cn/state\\_council/ministries/2015/03/21/content\\_281475075002989.htm](http://english.www.gov.cn/state_council/ministries/2015/03/21/content_281475075002989.htm) [accessed 2020-04-02]
26. Notice on Prohibiting the Sale of Electronic Cigarettes to Minors. State Administration for Market Regulation (SAMR). 2018 Aug 31. URL: [http://www.gov.cn/fuwu/2018-08/31/content\\_5317952.htm](http://www.gov.cn/fuwu/2018-08/31/content_5317952.htm) [accessed 2020-04-02]
27. Notice on Further Protecting Minors from Electronic Cigarettes. State Administration for Market Regulation (SAMR). 2019. URL: [http://gkml.samr.gov.cn/nsjg/xyjgs/201911/t20191101\\_308077.html](http://gkml.samr.gov.cn/nsjg/xyjgs/201911/t20191101_308077.html) [accessed 2020-03-23]
28. Gao J, Chapman S, Sun S, Fu H, Zheng P. The growth in newspaper coverage of tobacco control in China, 2000-2010. *BMC Public Health* 2012 Mar 7;12:160 [FREE Full text] [doi: [10.1186/1471-2458-12-160](https://doi.org/10.1186/1471-2458-12-160)] [Medline: [22394417](#)]
29. Lazard AJ, Wilcox GB, Tuttle HM, Glowacki EM, Pikowski J. Public reactions to e-cigarette regulations on Twitter: a text mining analysis. *Tob Control* 2017 Dec;26(e2):e112-e116. [doi: [10.1136/tobaccocontrol-2016-053295](https://doi.org/10.1136/tobaccocontrol-2016-053295)] [Medline: [28341768](#)]
30. Emery SL, Vera L, Huang J, Szczypka G. Wanna know about vaping? Patterns of message exposure, seeking and sharing information about e-cigarettes across media platforms. *Tob Control* 2014 Jul;23(Suppl 3):iii17-iii25 [FREE Full text] [doi: [10.1136/tobaccocontrol-2014-051648](https://doi.org/10.1136/tobaccocontrol-2014-051648)] [Medline: [24935893](#)]
31. Cassa C, Chunara R, Mandl K, Brownstein JS. Twitter as a sentinel in emergency situations: lessons from the Boston marathon explosions. *PLoS Curr* 2013 Jul 2;5:- [FREE Full text] [doi: [10.1371/currents.dis.ad70cd1c8bc585e9470046cde334ee4b](https://doi.org/10.1371/currents.dis.ad70cd1c8bc585e9470046cde334ee4b)] [Medline: [23852273](#)]
32. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. *Am J Infect Control* 2010 Apr;38(3):182-188 [FREE Full text] [doi: [10.1016/j.ajic.2009.11.004](https://doi.org/10.1016/j.ajic.2009.11.004)] [Medline: [20347636](#)]
33. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. *J Med Internet Res* 2015 Oct 27;17(10):e243 [FREE Full text] [doi: [10.2196/jmir.4969](https://doi.org/10.2196/jmir.4969)] [Medline: [26508089](#)]
34. Lachlan KA, Spence PR, Lin X. Expressions of risk awareness and concern through Twitter: on the utility of using the medium as an indication of audience needs. *Comput Hum Behav* 2014 Jun;35:554-559. [doi: [10.1016/j.chb.2014.02.029](https://doi.org/10.1016/j.chb.2014.02.029)]
35. Lazard AJ, Scheinfeld E, Bernhardt JM, Wilcox GB, Suran M. Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. *Am J Infect Control* 2015 Oct 1;43(10):1109-1111. [doi: [10.1016/j.ajic.2015.05.025](https://doi.org/10.1016/j.ajic.2015.05.025)] [Medline: [26138998](#)]
36. Aphinyanaphongs Y, Lulejian A, Brown DP, Bonneau RP. Text Classification for Automatic Detection of E-cigarette Use and Use for Smoking Cessation From Twitter: A Feasibility Pilot. In: Proceedings of the Pacific Symposium. 2016 Presented at: BioComputing'16; January 4-8, 2016; Hawaii, USA. [doi: [10.1142/9789814749411\\_0044](https://doi.org/10.1142/9789814749411_0044)]
37. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. *J Med Internet Res* 2015 Aug 25;17(8):e208 [FREE Full text] [doi: [10.2196/jmir.4392](https://doi.org/10.2196/jmir.4392)] [Medline: [26307512](#)]
38. Kavuluru R, Sabbir A. Toward automated e-cigarette surveillance: spotting e-cigarette proponents on Twitter. *J Biomed Inform* 2016 Jun;61:19-26 [FREE Full text] [doi: [10.1016/j.jbi.2016.03.006](https://doi.org/10.1016/j.jbi.2016.03.006)] [Medline: [26975599](#)]
39. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: social bots, emerging topics, and trends. *JMIR Public Health Surveill* 2017 Dec 20;3(4):e98 [FREE Full text] [doi: [10.2196/publichealth.8641](https://doi.org/10.2196/publichealth.8641)] [Medline: [29263018](#)]

40. van der Tempel J, Noormohamed A, Schwartz R, Norman C, Malas M, Zawertailo L. Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. *Int J Public Health* 2016 Mar;61(2):249-256. [doi: [10.1007/s00038-016-0791-2](https://doi.org/10.1007/s00038-016-0791-2)] [Medline: [26841895](https://pubmed.ncbi.nlm.nih.gov/26841895/)]
41. Lazard AJ, Saffer AJ, Wilcox GB, Chung AD, Mackert MS, Bernhardt JM. E-cigarette social media messages: a text mining analysis of marketing and consumer conversations on Twitter. *JMIR Public Health Surveill* 2016 Dec 12;2(2):e171 [FREE Full text] [doi: [10.2196/publichealth.6551](https://doi.org/10.2196/publichealth.6551)] [Medline: [27956376](https://pubmed.ncbi.nlm.nih.gov/27956376/)]
42. Cui K, Zheng X, Zeng D, Leischow S. An Empirical Analysis on Communications about Electronic Nicotine Delivery Systems (ENDS) in Chinese Social Media. In: International Conference on Smart Health. 2014 Presented at: ICHS'14; July 10-11, 2014; Beijing, China. [doi: [10.1007/978-3-319-08416-9\\_4](https://doi.org/10.1007/978-3-319-08416-9_4)]
43. Wen W, Zhang Z, Li Z, Liang J, Zhan Y, Zeng DD, et al. Public reactions to the cigarette control regulation on a Chinese microblogging platform: empirical analysis. *J Med Internet Res* 2020 Apr 27;22(4):e14660 [FREE Full text] [doi: [10.2196/14660](https://doi.org/10.2196/14660)] [Medline: [32338615](https://pubmed.ncbi.nlm.nih.gov/32338615/)]
44. Askeff-Williams HMJ. A correspondence analysis of child-care students' and medical students' knowledge about teaching and learning. *International Education Journal* 2004;5(2):176-204.
45. Greenacre M. Correspondence analysis in medical research. *Stat Methods Med Res* 1992;1(1):97-117. [doi: [10.1177/096228029200100106](https://doi.org/10.1177/096228029200100106)] [Medline: [1341654](https://pubmed.ncbi.nlm.nih.gov/1341654/)]
46. Beishuizen JJ, Hof E, van Putten CM, Bouwmeester SJ, Asscher JJ. Students' and teachers' cognitions about good teachers. *Br J Educ Psychol* 2001 Jun;71(Pt 2):185-201. [doi: [10.1348/000709901158451](https://doi.org/10.1348/000709901158451)] [Medline: [11449932](https://pubmed.ncbi.nlm.nih.gov/11449932/)]
47. Tang L, Choi S, Morrison AM, Lehto XY. The many faces of Macau: a correspondence analysis of the images communicated by online tourism information sources in English and Chinese. *J Vacat Mark* 2009 Jan;15(1):79-94. [doi: [10.1177/1356766708098173](https://doi.org/10.1177/1356766708098173)]
48. Bischoff P. A look inside Zhihu, China's Answer to Quora. *Techinasia*. 2014 Jun 09. URL: <https://www.techinasia.com/zhihu-chinas-answer-quora> [accessed 2020-03-23]
49. Shu C. How Zhihu Has Become One of China's Biggest Hubs for Experts. *Techcrunch*. 2019 Sep 03. URL: <https://techcrunch.com/2019/09/03/how-zhihus-become-one-of-chinas-biggest-hubs-for-experts/> [accessed 2020-03-23]
50. Dudarenok AG. Econsultancy. 2018. URL: <https://econsultancy.com/zhihu-china-s-q-a-social-platform-that-s-ripe-for-brands/> [accessed 2020-03-23]
51. Yang S. Zhouyuan, Ceo of Zhihu: the Number of Users Has Exceeded 220 Million, Exploring Different Liquidation Paths. *Sohu*. 2019 Apr 28. URL: [https://www.sohu.com/a/310809463\\_202972](https://www.sohu.com/a/310809463_202972) [accessed 2020-03-23]
52. Hoffman DL, Franke GR. Correspondence analysis: graphical representation of categorical data in marketing research. *J Mark Res* 1986 Aug;23(3):213-217. [doi: [10.2307/3151480](https://doi.org/10.2307/3151480)]
53. SPSS Categories Reference Guide. Eleventh Edition. Chicago: SPSS Inc; 2001.
54. Hair JF, Anderson RE, Tatham RL, Black WC. *Multivariate Data Analysis*. Upper Sadle River, NJ: Prentice-Hall; 1998.
55. Bendixen M. A practical guide to the use of correspondence analysis in marketing research. *Mark Res On-Line* 1996;1(1):16-36. [doi: [10.1201/b14416-8](https://doi.org/10.1201/b14416-8)]
56. Bock T. How to Interpret Correspondence Analysis Plots (It Probably Isn't the Way You Think). *Displayr: Analysis and Reporting Software for Survey Data*. URL: <https://www.displayr.com/interpret-correspondence-analysis-plots-probably-isnt-way-think/> [accessed 2020-04-24]
57. Bock T. How Correspondence Analysis Works (A Simple Explanation). URL: <https://www.displayr.com/how-correspondence-analysis-works/> [accessed 2020-04-24]
58. Lyu J, Wang D, Mao Z, Ling P. Evolution of Media Frames About E-cigarettes From 2004 to 2019: A Content Analysis of Newspapers in China. In: The 70th Annual Conference of International Communication Association. 2020 Presented at: ICA'20; May 21-25, 2020; Gold Coast, Australia.
59. Liu Y. 'Ruyan' Status Leads to Embarrassment of Supervision, Claiming to 'fill Legal Gap'. *Beijing Times*. 2006. URL: [https://www.dnb.nl/binaries/Supervision%20of%20Behaviour%20and%20Culture\\_tcm46-334417.pdf](https://www.dnb.nl/binaries/Supervision%20of%20Behaviour%20and%20Culture_tcm46-334417.pdf) [accessed 2020-10-01]
60. Wan J. The State Administration of Market Regulation and the State Tobacco Monopoly Administration Issued a Notice Banning the Sale of Electronic Cigarettes to Minors. *Legal Daily*. 2018 Sep 05. URL: [http://www.xinhuanet.com/legal/2018-09/05/c\\_129947671.htm](http://www.xinhuanet.com/legal/2018-09/05/c_129947671.htm) [accessed 2020-04-05]
61. Goodchild M, Zheng R. Early assessment of China's 2015 tobacco tax increase. *Bull World Health Organ* 2018 May 14;96(7):506-512. [doi: [10.2471/blt.17.205989](https://doi.org/10.2471/blt.17.205989)]
62. Sharpless N. How FDA is Regulating E-Cigarettes. *US Food and Drug Association*. 2019 Sep 10. URL: <https://www.fda.gov/news-events/fda-voices-perspectives-fda-leadership-and-experts/how-fda-regulating-e-cigarettes> [accessed 2020-04-14]
63. McCombs M, Shaw D. The agenda-setting function of mass media. *Public Opinion Q* 1972;36(2):176-187. [doi: [10.1086/267990](https://doi.org/10.1086/267990)]
64. Smith KA. Effects of newspaper coverage on community issue concerns and local government evaluations. *Communication Research* 2016;14(4):379-395. [doi: [10.1177/009365087014004001](https://doi.org/10.1177/009365087014004001)]
65. Chen M, Huang R. Where has opinion leader gone? Analysis of the discussion of Hainan arbitration case on Weibo, Wechat and Zhihu. *Shanghai J Rev* 2018;425(7):63-74.

---

**Abbreviations**

**e-cigarette:** electronic cigarette

**e-commerce:** electronic commerce.

**STMA:** State Tobacco Monopoly Administration

---

*Edited by G Eysenbach; submitted 01.05.20; peer-reviewed by A Sanders-Jackson, S Song; comments to author 27.07.20; revised version received 04.09.20; accepted 07.09.20; published 14.10.20.*

*Please cite as:*

Wang D, Lyu JC, Zhao X

Public Opinion About E-Cigarettes on Chinese Social Media: A Combined Study of Text Mining Analysis and Correspondence Analysis  
*J Med Internet Res* 2020;22(10):e19804

URL: <http://www.jmir.org/2020/10/e19804/>

doi: [10.2196/19804](https://doi.org/10.2196/19804)

PMID: [33052127](https://pubmed.ncbi.nlm.nih.gov/33052127/)

©Di Wang, Joanne Chen Lyu, Xiaoyu Zhao. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Googling for Ticks and Borreliosis in Germany: Nationwide Google Search Analysis From 2015 to 2018

Cora Scheerer<sup>1</sup>, MD; Melvin R  th<sup>1</sup>; Linda Tizek<sup>1</sup>, MPH, PhD; Martin K  berle<sup>1</sup>, PhD; Tilo Biedermann<sup>1</sup>, Prof Dr, MD; Alexander Zink<sup>1</sup>, MD, MPH, PhD

Department of Dermatology and Allergy, Technical University of Munich, Munich, Germany

**Corresponding Author:**

Alexander Zink, MD, MPH, PhD

Department of Dermatology and Allergy

Technical University of Munich

Biedersteinerstr.29

Munich, 80802

Germany

Phone: 49 08941400

Email: [alexander.zink@tum.de](mailto:alexander.zink@tum.de)

## Abstract

**Background:** Borreliosis is the most frequently transmitted tick-borne disease in Europe. It is difficult to estimate the incidence of tick bites and associated diseases in the German population due to the lack of an obligation to register across all 16 federal states of Germany.

**Objective:** The aim of this study is to show that Google data can be used to generate general trends of infectious diseases on the basis of borreliosis and tick bites. In addition, the possibility of using Google AdWord data to estimate incidences of infectious diseases, where there is inconsistency in the obligation to notify authorities, is investigated with the perspective to facilitate public health studies.

**Methods:** Google AdWords Keyword Planner was used to identify search terms related to ticks and borreliosis in Germany from January 2015 to December 2018. The search volume data from the identified search terms was assessed using Excel version 15.23. In addition, SPSS version 24.0 was used to calculate the correlation between search volumes, registered cases, and temperature.

**Results:** A total of 1999 tick-related and 542 borreliosis-related search terms were identified, with a total of 209,679,640 Google searches in all 16 German federal states in the period under review. The analysis showed a high correlation between temperature and borreliosis ( $r=0.88$ ), and temperature and tick bite ( $r=0.83$ ), and a very high correlation between borreliosis and tick bite ( $r=0.94$ ). Furthermore, a high to very high correlation between Google searches and registered cases in each federal state was observed (Brandenburg  $r=0.80$ , Mecklenburg-West Pomerania  $r=0.77$ , Saxony  $r=0.74$ , and Saxony-Anhalt  $r=0.90$ ; all  $P<.001$ ).

**Conclusions:** Our study provides insight into annual trends concerning interest in ticks and borreliosis that are relevant to the German population exemplarily in the data of a large internet search engine. Public health studies collecting incidence data may benefit from the results indicating a significant correlation between internet search data and incidences of infectious diseases.

(*J Med Internet Res* 2020;22(10):e18581) doi:[10.2196/18581](https://doi.org/10.2196/18581)

## KEYWORDS

Google; infodemiology; infoveillance; public health; seasonal health trend; medical internet research; tick-borne disease; tick bites, borreliosis; Lyme disease

## Introduction

Borreliosis is the most frequently transmitted tick-borne disease in Europe. In 80%-90% of all cases, the disease presents with visible skin manifestations [1-3]. However, there is limited accurate data for the incidence of tick bites and borreliosis as

associated diseases in Germany. Tick bites and associated diseases are important public health concerns because of their high incidence with no clear increasing or decreasing trend in Germany with regional variation [4]. Tick bite protection, correct and prompt tick removal, and medical consultation



should be promoted by physicians and health authorities to facilitate early diagnosis and treatment of associated diseases.

In 9 out of 16 federal states in Germany (Bavaria, Berlin, Brandenburg, Mecklenburg-West Pomerania, Rhineland-Palatinate, Saarland, Saxony, Saxony-Anhalt, and Thuringia), it is mandatory to report diagnosed borreliosis to the German federal government agency and research institute for disease control and prevention (Robert-Koch Institute). Thus, epidemiologic data for tick bites and associated diseases are based on measured, as well as estimated, values.

Google search analysis is a powerful tool to reflect the German population's interest in specific topics because of its 94% market share [5]. Additionally, the general public favors search engines like *Google* over specialized websites when searching for primary health information online [6-9]. Previous studies have already demonstrated that analysis of internet search volume, which is one of the methods of the fields of infodemiology and infoveillance, is a valid method for assessing medical topics [10-15]. The internet's emerging role as a main, or at least primary, source of health advice for the general public has prompted a corresponding increase in its value in the medical field. Huang et al [16], for example, reported a minor association between online cancer-related information searches and skin cancer incidence. Additionally, Wehner et al [17] established that internet search volume positively correlates with the incidence and mortality rates of common cancers in the United States. Regarding infectious diseases, Ginsberg et al [18] showed an accurate estimation of weekly influenza activity in the United States correlated with queries in online search engines. They suggested that internet research could help physicians diagnose influenza earlier to prevent epidemics [18].

Ticks are only active when ambient air temperature is 4-10 °C, so average temperature should be an important factor influencing tick-related queries. Therefore, weather data should provide insight into seasonal patterns [19]. Previous studies have shown that there are seasonal patterns in Google search volumes but have not found significant correlation between mean monthly temperature and internet searches for "tick" [20].

This study aims to investigate the interest of the German population in tick bites and borreliosis by analyzing Google searches. Furthermore, this study aims to explore correlations between searches and whether that could provide information about real life tick bite occurrences, as well as associated diseases.

## Methods

### Study Design

In this retrospective study, Google AdWords Keyword Planner was used to measure the search volume of terms related to tick bites and borreliosis across Germany from January 2015 to December 2018. The Keyword Planner is often used by

advertisers to improve Google marketing campaigns and provides monthly search volumes estimated by Google. The term *search volume* applies to the number of searches for a topic or search term. To assess search volume within a specific field, words are initially entered into the Keyword Planner; thereupon, the program provides keywords that are most relevant to the topic. This process may be used both to answer scientific questions and for medical research [10,11].

In addition, search terms related to tick bites were identified using a keyword cluster for the German words for "tick bite" ("Zeckenbiss") and "borreliosis" ("Borreliose"). Based on this cluster, Google AdWords Keyword Planner determined search terms to be analyzed. This data included only Google users with a German internet protocol address who used the German language. Furthermore, the German Climate Data Centre [21] was used to relate Google search volume to weather data by analyzing mean monthly temperature in degrees Celsius. Due to seasonal differences in tick activity as well as tick bite incidence, we defined summer months as April to September and winter months as October to March.

In 9 of 16 federal states in Germany (Bavaria, Berlin, Brandenburg, Mecklenburg-West Pomerania, Rhineland-Palatinate, Saarland, Saxony, Saxony-Anhalt, and Thuringia) covering 42% of the total German population, mandatory notification for the three most common Lyme borreliosis manifestations (erythema migrans, acute neuroborreliosis, and Lyme arthritis) has been achieved since 2013.

To assess whether the Google search volume correlates with registered cases of borreliosis, all registered cases from the federal states of Brandenburg, Mecklenburg-West Pomerania, Saxony, and Saxony-Anhalt were considered in the analysis, as complete statistics of registered data were only available for these on the website of the German federal government agency and research institute for disease control and prevention (Robert-Koch Institut).

### Statistical Analysis

The search volume data of the identified search terms was assessed using Excel version 15.23 (Microsoft Corporation). To describe the relationship between the investigated variables, we used SPSS version 24.0 (IBM Corp) to calculate the Pearson correlation coefficient (*r*) [22].

## Results

In total, Google AdWords Keyword Planner identified 1999 search terms related to tick bites with a search volume of 26,080,530 in Germany from January 2015 to December 2018. The most frequently searched terms were "tick sting" ("Zeckenstich"; *n*=2,821,800, 10.82%), "tick" ("Zecke"; *n*=2,387,500, 9.15%), and "tick bite" ("Zeckenbiss"; *n*=178,850, 0.69%; [Textbox 1](#)).



**Textbox 1.** Top five key terms for tick bite (Zeckenbiss) and borreliosis (Borreliose).

**Tick bite (Zeckenbiss; German translation in parenthesis)**

- “tick-bite” (“Zeckenbiss”)
- “tick sting” (“Zeckenstich”)
- “tick” (“Zecke”)
- “borreliosis” (“Borreliose”)
- “borreliosis symptoms” (“Borreliose Symptome”)

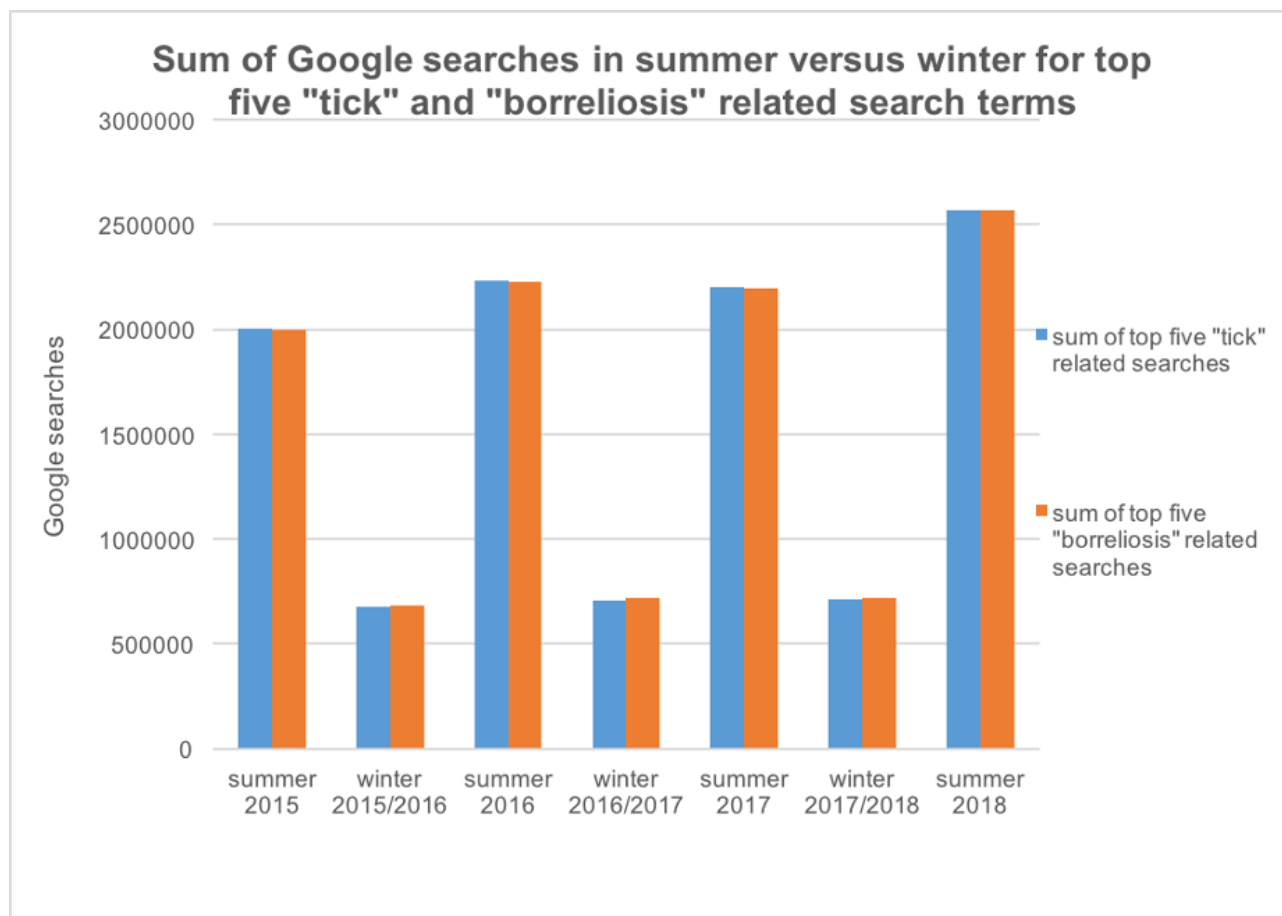
**Borreliosis (Borreliose; German translation in parenthesis)**

- “borreliosis” (“Borreliose”)
- “borreliosis symptoms” (“Borreliose Symptome”)
- “tick- bite” (“Zeckenbiss”)
- “tick” (“Zecke”)
- “symptoms borreliosis” (“Symptome Borreliose”)

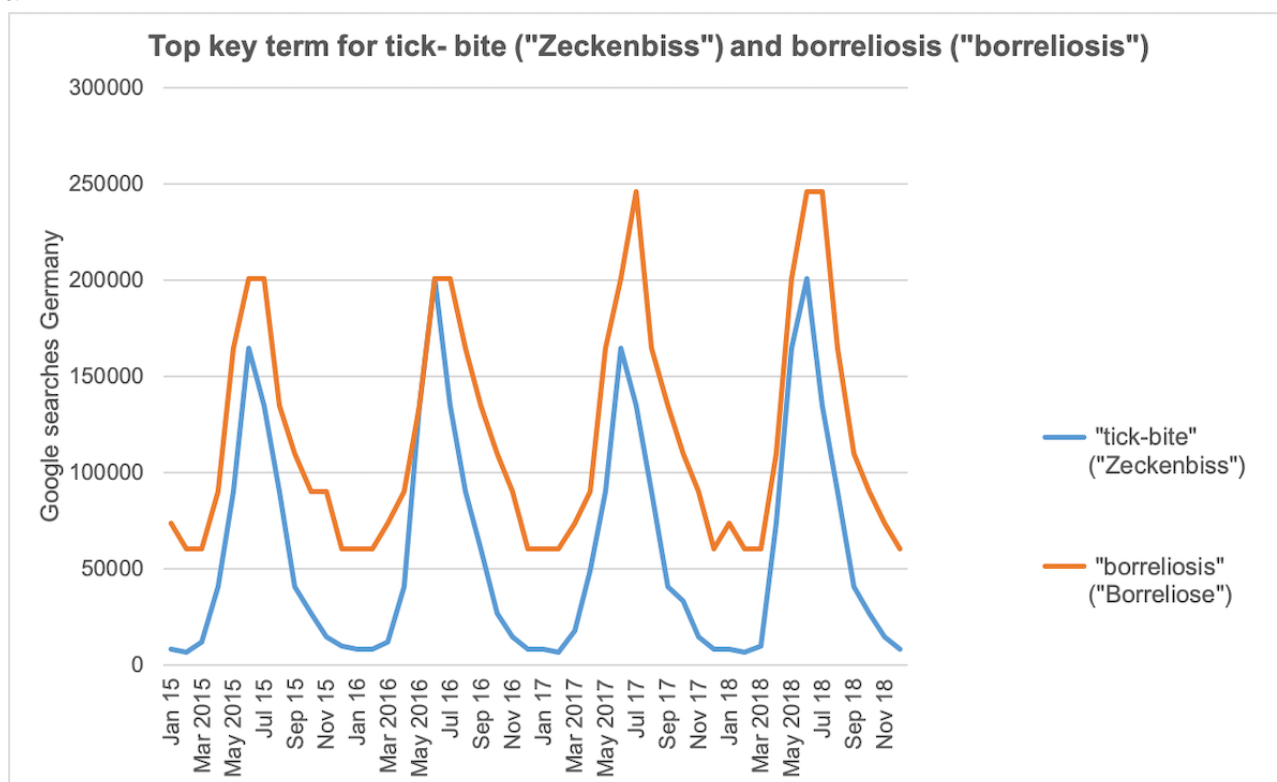
Every year, an increase in search volume during summer months was observed (Figure 1). The month with the highest overall search volume was June 2018 with 1,571,330 searches. The search volume of tick bite and borreliosis showed similar trends in search volume, with “borreliosis” being more frequently

searched. Annual peaks of search volume for tick bite were seen every year in June. Annual peaks of search volume for borreliosis happened in June and July of 2015 and 2016 (n=201,000 searches each), July 2017 (n=246,000 searches), and June and July 2018 (n=246,000 searches each; Figure 2).

**Figure 1.** Google searches in Germany in summer (April to September) vs winter months (October to March) for the top five tick- and borreliosis-related search terms in 2015-2018.

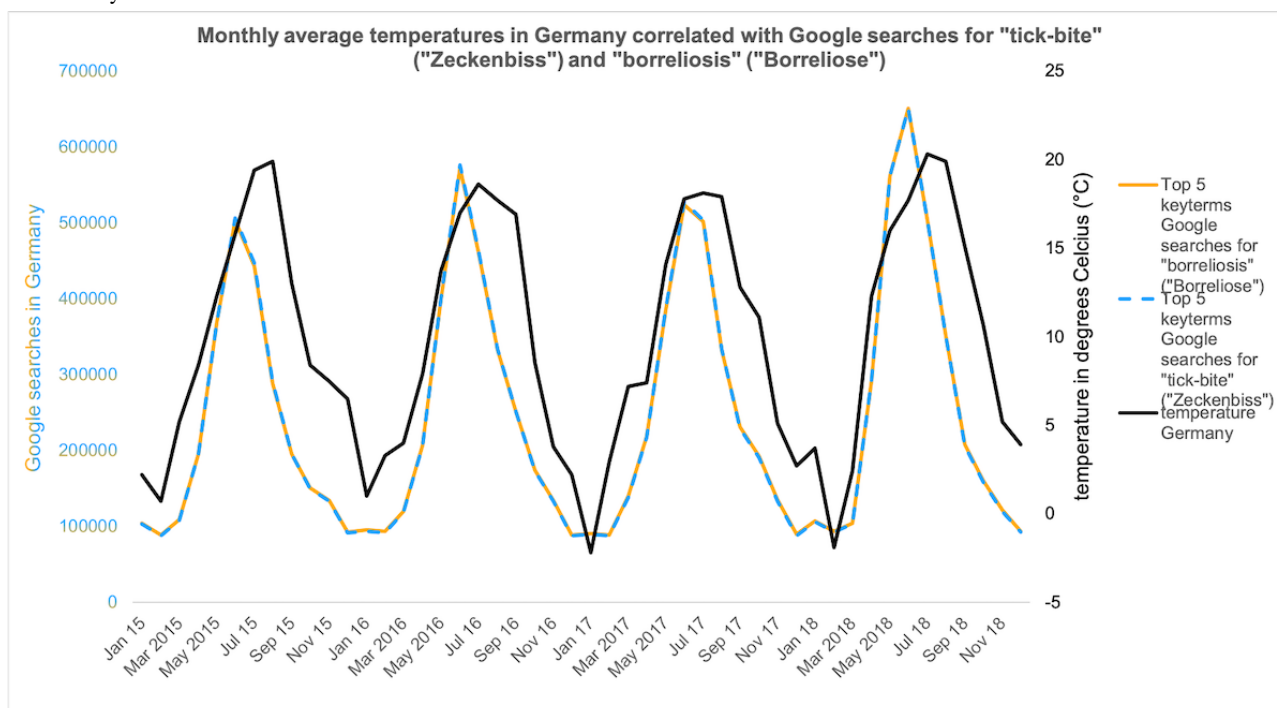


**Figure 2.** Seasonal variation of the two most common keywords searched for in Germany, “tick-bite” and “borreliosis,” from January 2015 to December 2018.



The analysis revealed a high correlation between temperature and borreliosis ( $r=0.88$ ,  $P<.001$ ) as well as between temperature and tick bite ( $r=0.83$ ,  $P<.001$ ; Figure 3). The very high correlation between borreliosis and tick bite ( $r=0.94$ ,  $P<.001$ ) depicts the seasonal- and temperature-dependent interest in the key terms.

**Figure 3.** Google searches in Germany for "tick bite" and "borreliosis" correlated with the monthly average temperature in Germany in Celsius degrees between January 2015 to December 2018.

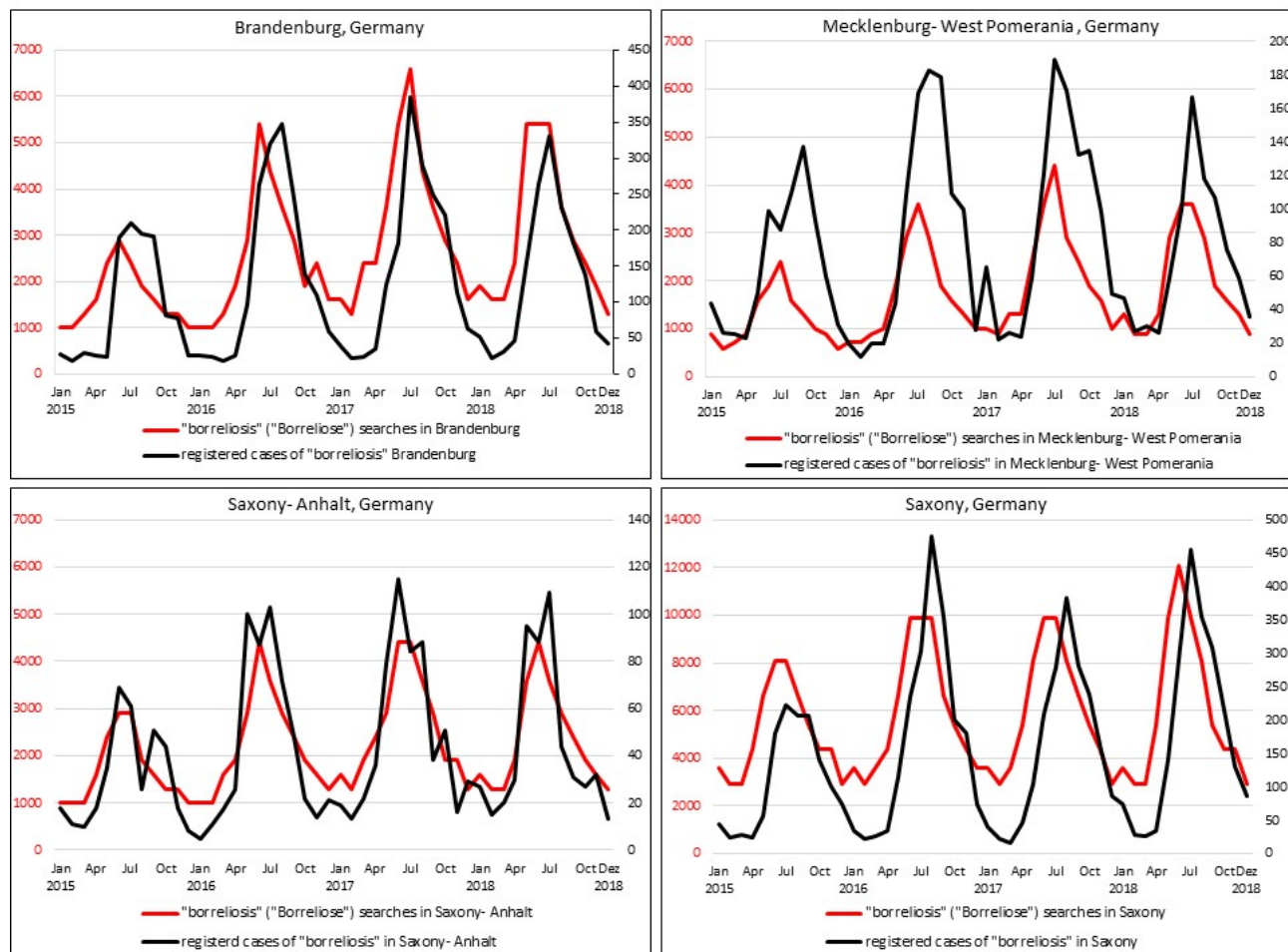


Furthermore, a high and very high correlation between google searches and registered cases in the referred federal states Brandenburg, Mecklenburg-West Pomerania, Saxony, and Saxony-Anhalt were detected (Figure 4; Table 1). Saxony, the

federal state with the largest population, had the highest number of Google searches ( $n=273,800$  searches), as well as the highest number of registered cases of borreliosis ( $n=7387$ ). Accordingly, a high correlation was found ( $r=0.74$ ,  $P<.001$ ). However, the

highest correlation was found in Saxony-Anhalt ( $r=0.90$ ,  $P<.001$ ; Table 1 and Figure 4).

**Figure 4.** Number of Google searches for "borreliosis" with registered cases of borreliosis in Brandenburg, Mecklenburg-West Pomerania, Saxony, and Saxony-Anhalt between January 2015 to December 2018.



**Table 1.** Pearson correlation coefficients for Google searches for borreliosis and registered cases of borreliosis from January 2015 to December 2018 in Brandenburg, Mecklenburg-West Pomerania, Saxony, and Saxony-Anhalt.

Federal state	Registered cases, n	Google searches, n	Pearson correlation	P value
Brandenburg	6107	124,000	0.80	<.001
Mecklenburg-West Pomerania	3720	82,480	0.77	<.001
Saxony	7387	273,800	0.74	<.001
Saxony-Anhalt	2016	104,700	0.90	<.001

## Discussion

### Principal Findings

The analysis of Google search volume related to tick bite and borreliosis identified an annual pattern that people tended to search more frequently during summer months. Therefore, a high correlation with average temperature was observed. Furthermore, a high correlation between registered cases of borreliosis in four German federal states was revealed.

One of the top five key terms of tick bite was borreliosis and vice versa. Therefore, we compared the two most common lay terms, tick bite and borreliosis. Interestingly, the search volume for the latter was higher. Especially in 2017 and 2018, a greater

divergence between the keywords was observed, which can be explained by a greater awareness of associated diseases. This might be because of celebrities diagnosed with Lyme disease, such as Bastian Schweinsteiger and Justin Bieber, or because health education programs have taught people to make more accurate searches. Furthermore, media like smartphone apps and video games significantly improve knowledge of the disease and preventive measures [23,24]. However, a Finnish survey showed that, regarding knowledge, attitudes, and practice toward ticks and tick-borne disease, 65% of participants relied on newspapers and magazines as the main source of information [25]. Pharmacy health magazines, radio, or TV shows start media coverage of ticks and tick-borne diseases in late spring when average temperatures rise and ticks begin to appear.

However, we did not find data to support this well-known approach.

Correlating weather data with the Google search volume showed seasonal trends, which were described in previous works concerning pruritus and identified inhabitants' needs [12]. Especially during the German winter months October 2016 to March 2017 and October 2017 to March 2018, the Google search volume showed a distinctive increase for tick bite and borreliosis compared to data from October 2015 to March 2016. The first hypothesis was that winters get milder in Germany due to climate change so that ticks have a longer active period. Ticks are active when average temperatures are between 4-10 °C (median 7 °C) [19]. However, as during each winter, there were a comparable number of months below this temperature. This does not explain the recognizable increase in search volume that we can see. Potentially, the increase was due to the awareness of tick bites and associated diseases, as well as media campaigns starting earlier in those years.

In some German federal states, it is mandatory to report borreliosis cases. Comparing numbers from Google searches and registered borreliosis cases shows a discrepancy. For example, in Brandenburg, the highest number of registered borreliosis cases in the reviewed years was 1743 in 2017. In comparison, the 2017 Google search volume for "borreliosis" in Brandenburg was 38,200, which is 22 times higher. This might be because the number of tick bites are much greater than the development of borreliosis symptoms. Additionally, not only affected people but also their relatives might search for information online, which explains a considerably higher number of search queries.

Walker [20] posed the question of whether Google trends can be used to study parasitic (ie, tick-borne) diseases [20]. Their results showed seasonal patterns in search volume but no significant correlation between mean monthly temperature and internet searches for "tick." Additionally, they tried to use the internet search volume to estimate parasitic occurrence. However, there was no apparent relationship between the annual number of tick-borne encephalitis cases and mean annual internet searches for either tick or tick-borne encephalitis [20].

We found statistically high correlations between registered borreliosis cases and Google search volume in four federal states. Previous studies identified Google data as a predictor of infectious disease outbreaks [26,27]. Nevertheless, to the best of our knowledge, this is the first work that shows a high correlation between incidence of an infectious disease and Google search volume of the implied disease. These results could help estimate incidences of borreliosis in German federal states where registration is not mandatory. Furthermore, Google search volume could be used to estimate incidences of diseases that are not required to be reported.

### Limitations

This study has some limitations. In Germany, Google accounts for 95% of search engine use, so Google data can depict the interests of the population as a whole. To transfer our findings to other countries, different market shares of Google over alternative search engines need to be taken into account. Although it is common among the whole population to make health-related searches, younger people tend to use the internet more often [28]. Furthermore, the automatic completion of search terms by Google may influence people's search behavior. It may promote an understanding of the health problem and the need to seek necessary medical help; however, priming by autocomplete has the potential to make incorrect associations [29]. Another limitation to our study is that we solely used German key terms.

### Conclusion

Our study provides insight into terms and fields of interest associated with tick bites and borreliosis, relevant to the German population. We found statistically high correlations between Google searches for borreliosis and registered cases of borreliosis across four German federal states. Accordingly, these results could help to estimate the incidence of borreliosis in the remaining 12 German federal states where it is not mandatory to report borreliosis. Furthermore, this approach could aid in the development and implementation of effective and sustainable awareness campaigns.

### Acknowledgments

This work was supported by the department of Dermatology and Allergy, Technical University of Munich, School of Medicine, Munich, Germany.

### Conflicts of Interest

None declared.

### References

1. Glatz M, Resinger A, Semmelweis K, Ambros-Rudolph C, Müllegger RR. Clinical spectrum of skin manifestations of Lyme borreliosis in 204 children in Austria. *Acta Derm Venereol* 2015 May;95(5):565-571 [FREE Full text] [doi: [10.2340/00015555-2000](https://doi.org/10.2340/00015555-2000)] [Medline: [25366035](https://pubmed.ncbi.nlm.nih.gov/25366035/)]
2. Huppertz HI, Böhme M, Standaert SM, Karch H, Plotkin SA. Incidence of Lyme borreliosis in the Würzburg region of Germany. *Eur J Clin Microbiol Infect Dis* 1999 Oct;18(10):697-703. [doi: [10.1007/s100960050381](https://doi.org/10.1007/s100960050381)] [Medline: [10584895](https://pubmed.ncbi.nlm.nih.gov/10584895/)]
3. Hofmann H, Fingerle V, Hunfeld KP, Huppertz HI, Krause A, Rauer S, Consensus group. Cutaneous Lyme borreliosis: guideline of the German Dermatology Society. *Ger Med Sci* 2017;15:Doc14. [doi: [10.3205/000255](https://doi.org/10.3205/000255)] [Medline: [28943834](https://pubmed.ncbi.nlm.nih.gov/28943834/)]



4. Enkelmann J, Böhmer M, Fingerle V, Sifczyk C, Werber D, Littmann M, et al. Incidence of notified Lyme borreliosis in Germany, 2013-2017. *Sci Rep* 2018 Oct 08;8(1):14976. [doi: [10.1038/s41598-018-33136-0](https://doi.org/10.1038/s41598-018-33136-0)] [Medline: [30297731](https://pubmed.ncbi.nlm.nih.gov/30297731/)]
5. Welche Suchmaschinen haben Sie in den letzten 4 Wochen genutzt? Statista. URL: <https://de.statista.com/prognosen/999767/umfrage-in-deutschland-zu-beliebten-suchmaschinen> [accessed 2019-06-25]
6. Mazloomdoost D, Kanter G, Chan RC, Deveaneau N, Wyman AM, Von Bargaen EC, et al. Social networking and internet use among pelvic floor patients: a multicenter survey. *Am J Obstet Gynecol* 2016 Nov;215(5):654.e1-654.e10. [doi: [10.1016/j.ajog.2016.06.011](https://doi.org/10.1016/j.ajog.2016.06.011)] [Medline: [27319368](https://pubmed.ncbi.nlm.nih.gov/27319368/)]
7. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google is here to stay but health care professionals are still valued: an analysis of health care consumers' internet navigation support preferences. *J Med Internet Res* 2017 Jun 14;19(6):e210 [FREE Full text] [doi: [10.2196/jmir.7489](https://doi.org/10.2196/jmir.7489)] [Medline: [28615156](https://pubmed.ncbi.nlm.nih.gov/28615156/)]
8. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ* 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: [10.1136/bmj.324.7337.573](https://doi.org/10.1136/bmj.324.7337.573)] [Medline: [11884321](https://pubmed.ncbi.nlm.nih.gov/11884321/)]
9. Sharpe JD, Hopkins RS, Cook RL, Striley CW. Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian change point analysis: a comparative analysis. *JMIR Public Health Surveill* 2016 Oct 20;2(2):e161 [FREE Full text] [doi: [10.2196/publichealth.5901](https://doi.org/10.2196/publichealth.5901)] [Medline: [27765731](https://pubmed.ncbi.nlm.nih.gov/27765731/)]
10. Seidl S, Schuster B, Rütth M, Biedermann T, Zink A. What do Germans want to know about skin cancer? A nationwide Google search analysis from 2013 to 2017. *J Med Internet Res* 2018 May 02;20(5):e10327 [FREE Full text] [doi: [10.2196/10327](https://doi.org/10.2196/10327)] [Medline: [29698213](https://pubmed.ncbi.nlm.nih.gov/29698213/)]
11. Zink A, Schuster B, Rütth M, Pereira M, Philipp-Dormston WG, Biedermann T, et al. Medical needs and major complaints related to pruritus in Germany: a 4-year retrospective analysis using Google AdWords Keyword Planner. *J Eur Acad Dermatol Venereol* 2019 Jan;33(1):151-156. [doi: [10.1111/jdv.15200](https://doi.org/10.1111/jdv.15200)] [Medline: [30067284](https://pubmed.ncbi.nlm.nih.gov/30067284/)]
12. Tizek L, Schielein M, Rütth M, Ständer S, Pereira MP, Eberlein B, et al. Influence of climate on Google internet searches for pruritus across 16 German cities: retrospective analysis. *J Med Internet Res* 2019 Jul 12;21(7):e13739 [FREE Full text] [doi: [10.2196/13739](https://doi.org/10.2196/13739)] [Medline: [31301128](https://pubmed.ncbi.nlm.nih.gov/31301128/)]
13. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. *J Med Internet Res* 2015 Apr 29;17(4):e106 [FREE Full text] [doi: [10.2196/jmir.4126](https://doi.org/10.2196/jmir.4126)] [Medline: [25925943](https://pubmed.ncbi.nlm.nih.gov/25925943/)]
14. Mavragani A. Infodemiology and infoveillance: scoping review. *J Med Internet Res* 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: [10.2196/16206](https://doi.org/10.2196/16206)] [Medline: [32310818](https://pubmed.ncbi.nlm.nih.gov/32310818/)]
15. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009 Mar 27;11(1):e11 [FREE Full text] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
16. Huang X, Baade P, Youlden DR, Youl PH, Hu W, Kimlin MG. Google as a cancer control tool in Queensland. *BMC Cancer* 2017 Dec 04;17(1):816 [FREE Full text] [doi: [10.1186/s12885-017-3828-x](https://doi.org/10.1186/s12885-017-3828-x)] [Medline: [29202718](https://pubmed.ncbi.nlm.nih.gov/29202718/)]
17. Wehner MR, Nead KT, Linos E. Correlation among cancer incidence and mortality rates and internet searches in the United States. *JAMA Dermatol* 2017 Sep 01;153(9):911-914 [FREE Full text] [doi: [10.1001/jamadermatol.2017.1870](https://doi.org/10.1001/jamadermatol.2017.1870)] [Medline: [28658470](https://pubmed.ncbi.nlm.nih.gov/28658470/)]
18. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Mar 19;457(7232):1012-1014. [doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634)] [Medline: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)]
19. Bouchard C, Dibbernardo A, Koffi J, Wood H, Leighton P, Lindsay L. N Increased risk of tick-borne diseases with climate and environmental changes. *Can Commun Dis Rep* 2019 Apr 04;45(4):83-89. [doi: [10.14745/ccdr.v45i04a02](https://doi.org/10.14745/ccdr.v45i04a02)] [Medline: [31285697](https://pubmed.ncbi.nlm.nih.gov/31285697/)]
20. Walker M. Can Google be used to study parasitic disease? Internet searching on tick-borne encephalitis in Germany. *J Vector Borne Dis* 2018;55(4):327-329 [FREE Full text] [doi: [10.4103/0972-9062.256571](https://doi.org/10.4103/0972-9062.256571)] [Medline: [30997896](https://pubmed.ncbi.nlm.nih.gov/30997896/)]
21. Wetterdienst D. Offenbach, Germany: Deutscher Wetterdienst Climate Data Center. URL: <https://www.dwd.de/DE/leistungen/klimadatendeutschland/klarchivtagmonat.html?nn=480164> [accessed 2019-05-01]
22. Mukaka M. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 2012;24(3):69-71.
23. Antonise-Kamp L, Beaujean DJMA, Crutzen R, van Steenbergen JE, Ruwaard D. Prevention of tick bites: an evaluation of a smartphone app. *BMC Infect Dis* 2017 Dec 04;17(1):744 [FREE Full text] [doi: [10.1186/s12879-017-2836-4](https://doi.org/10.1186/s12879-017-2836-4)] [Medline: [29202704](https://pubmed.ncbi.nlm.nih.gov/29202704/)]
24. Beaujean DJMA, Gassner F, Wong A, Steenbergen JE, Crutzen R, Ruwaard D. Education on tick bite and Lyme borreliosis prevention, aimed at schoolchildren in the Netherlands: comparing the effects of an online educational video game versus a leaflet or no intervention. *BMC Public Health* 2016 Nov 16;16(1):1163 [FREE Full text] [doi: [10.1186/s12889-016-3811-5](https://doi.org/10.1186/s12889-016-3811-5)] [Medline: [27852247](https://pubmed.ncbi.nlm.nih.gov/27852247/)]
25. Zöldi V, Turunen T, Lyytikäinen O, Sane J. Knowledge, attitudes, and practices regarding ticks and tick-borne diseases, Finland. *Ticks Tick Borne Dis* 2017 Oct;8(6):872-877. [doi: [10.1016/j.ttbdis.2017.07.004](https://doi.org/10.1016/j.ttbdis.2017.07.004)] [Medline: [28778675](https://pubmed.ncbi.nlm.nih.gov/28778675/)]



26. Simonsen L, Gog JR, Olson D, Viboud C. Infectious disease surveillance in the big data era: towards faster and locally relevant systems. *J Infect Dis* 2016 Dec 01;214(suppl\_4):S380-S385 [FREE Full text] [doi: [10.1093/infdis/jiw376](https://doi.org/10.1093/infdis/jiw376)] [Medline: [28830112](https://pubmed.ncbi.nlm.nih.gov/28830112/)]
27. Verma M, Kishore K, Kumar M, Sondh AR, Aggarwal G, Kathirvel S. Google search trends predicting disease outbreaks: an analysis from India. *Healthc Inform Res* 2018 Oct;24(4):300-308 [FREE Full text] [doi: [10.4258/hir.2018.24.4.300](https://doi.org/10.4258/hir.2018.24.4.300)] [Medline: [30443418](https://pubmed.ncbi.nlm.nih.gov/30443418/)]
28. Ybarra ML, Suman M. Help seeking behavior and the Internet: a national survey. *Int J Med Inform* 2006 Jan;75(1):29-41. [doi: [10.1016/j.ijmedinf.2005.07.029](https://doi.org/10.1016/j.ijmedinf.2005.07.029)] [Medline: [16129659](https://pubmed.ncbi.nlm.nih.gov/16129659/)]
29. Loh LC. Autocomplete: Dr Google's "helpful" assistant? *Can Fam Physician* 2016 Aug;62(8):622-623. [Medline: [27521382](https://pubmed.ncbi.nlm.nih.gov/27521382/)]

*Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by S Salzmann, C Liao; comments to author 20.04.20; revised version received 06.05.20; accepted 20.05.20; published 16.10.20.*

*Please cite as:*

Scheerer C, Rüth M, Tizek L, Köberle M, Biedermann T, Zink A

*Googling for Ticks and Borreliosis in Germany: Nationwide Google Search Analysis From 2015 to 2018*

*J Med Internet Res* 2020;22(10):e18581

URL: <http://www.jmir.org/2020/10/e18581/>

doi: [10.2196/18581](https://doi.org/10.2196/18581)

PMID: [33064086](https://pubmed.ncbi.nlm.nih.gov/33064086/)

©Cora Scheerer, Melvin Rüth, Linda Tizek, Martin Köberle, Tilo Biedermann, Alexander Zink. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 16.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Review

# Implementation of Electronic Informed Consent in Biomedical Research and Stakeholders' Perspectives: Systematic Review

Evelien De Sutter<sup>1</sup>, MSc; Drieda Zace<sup>2</sup>, MD; Stefania Boccia<sup>2,3</sup>, PhD; Maria Luisa Di Pietro<sup>2</sup>, MD; David Geerts<sup>4</sup>, PhD; Pascal Borry<sup>5</sup>, PhD; Isabelle Huys<sup>1</sup>, PhD

<sup>1</sup>Clinical Pharmacology and Pharmacotherapy, Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, Leuven, Belgium

<sup>2</sup>Section of Hygiene, University Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Roma, Italy

<sup>3</sup>Department of Woman and Child Health and Public Health, Fondazione Policlinico Universitario A Gemelli IRCCS, Roma, Italy

<sup>4</sup>Meaningful Interactions Lab, KU Leuven, Leuven, Belgium

<sup>5</sup>Centre for Biomedical Ethics and Law, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

**Corresponding Author:**

Evelien De Sutter, MSc

Clinical Pharmacology and Pharmacotherapy

Department of Pharmaceutical and Pharmacological Sciences

KU Leuven

O&N II – Herestraat 49 – Box 521

Leuven, 3000

Belgium

Phone: 32 16 37 98 82

Email: [evelien.desutter@kuleuven.be](mailto:evelien.desutter@kuleuven.be)

## Abstract

**Background:** Informed consent is one of the key elements in biomedical research. The introduction of electronic informed consent can be a way to overcome many challenges related to paper-based informed consent; however, its novel opportunities remain largely unfulfilled due to several barriers.

**Objective:** We aimed to provide an overview of the ethical, legal, regulatory, and user interface perspectives of multiple stakeholder groups in order to assist responsible implementation of electronic informed consent in biomedical research.

**Methods:** We conducted a systematic literature search using Web of Science (Core collection), PubMed, EMBASE, ACM Digital Library, and PsycARTICLES. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were used for reporting this work. We included empirical full-text studies focusing on the concept of electronic informed consent in biomedical research covering the ethical, legal, regulatory, and user interface domains. Studies written in English and published from January 2010 onward were selected. We explored perspectives of different stakeholder groups, in particular researchers, research participants, health authorities, and ethics committees. We critically appraised literature included in the systematic review using the Newcastle-Ottawa scale for cohort and cross-sectional studies, Critical Appraisal Skills Programme for qualitative studies, Mixed Methods Appraisal Tool for mixed methods studies, and Jadad tool for randomized controlled trials.

**Results:** A total of 40 studies met our inclusion criteria. Overall, the studies were heterogeneous in the type of study design, population, intervention, research context, and the tools used. Most of the studies' populations were research participants (ie, patients and healthy volunteers). The majority of studies addressed barriers to achieving adequate understanding when using electronic informed consent. Concerns shared by multiple stakeholder groups were related to the security and legal validity of an electronic informed consent platform and usability for specific groups of research participants.

**Conclusions:** Electronic informed consent has the potential to improve the informed consent process in biomedical research compared to the current paper-based consent. The ethical, legal, regulatory, and user interface perspectives outlined in this review might serve to enhance the future implementation of electronic informed consent.

**Trial Registration:** PROSPERO International Prospective Register of Systematic Reviews CRD42020158979; [https://www.crd.york.ac.uk/prospere/display\\_record.php?RecordID=158979](https://www.crd.york.ac.uk/prospere/display_record.php?RecordID=158979)

(*J Med Internet Res* 2020;22(10):e19129) doi:[10.2196/19129](https://doi.org/10.2196/19129)

**KEYWORDS**

Informed consent; systematic review; biomedical research; user interface; research ethics; digital health

**Introduction**

Obtaining informed consent is a fundamental ethical practice in biomedical research. It is the process of providing meaningful information to the potential participant in order to enable an autonomous well-informed decision on whether or not they wish to participate in the research study [1-3]. Moreover, informed consent may serve as one of the legal grounds for processing personal data, as described in articles 6 and 9 of the General Data Protection Regulation [4]. The primary goal of informed consent is to truly inform potential research participants or their representatives about different aspects such as the study design, study procedures, the risks and benefits, treatment options, participants' responsibilities, and the right to withdraw as well as information regarding data processing [1,3,4]. Therefore, information must be available in lay terminology and in a language understandable to the participants [3]. Long and cumbersome paper-based informed consent documents are the result of the increasing complexity of clinical research and the multitude of legal and regulatory requirements to satisfy informed consent needs [5-7]. Regulatory requirements refer to those related to the regulatory approval of medicines [8]. Available evidence has shown that research participants lack understanding of the key concepts of research studies [7,9]. For this reason, many attempts have been made to improve the understanding of research participants [10].

Owing to innovations in information technology, different strategies to consent have been developed, ranging from involving multimedia to the implementation of quizzes [10]. Research on the use of different multimedia formats to present information and improve research participants' understanding is gaining popularity [10-12]. Recently, the US Food and Drug Administration (FDA), in collaboration with the Office for Human Research Protections, issued guidance [13] concerning electronic informed consent in order to provide a shared and harmonized approach. Electronic informed consent refers to electronic systems which may incorporate multimedia in order to convey information and to obtain informed consent. In this guidance [13], recommendations are described covering several aspects related to electronic informed consent such as the presentation of information, the use of electronic signatures, identity verification, FDA inspections, and the review process by ethics committees. The development of an electronic informed consent platform, enabling participants to give and manage their electronic informed consent, could offer several opportunities. First, it could facilitate long-term interaction with research participants in cases where reconsenting for follow-up studies is required or for providing research results. Second, it may truly inform research participants in an interactive, tailored approach based on the individual's information needs [14].

Considerable research has been devoted to single aspects important for electronic informed consent; however, rather less attention has been paid to integrating information from several important scientific domains such as ethical, legal, regulatory, and user interface domains. It is vital to balance the relevant

domains in order to create an electronic informed consent platform that better informs, empowers, and engages research participants [15,16]. In the field of research, the ethics committee plays an important role as it is responsible for reviewing study protocols to ensure that they meet the ethical, legal, and regulatory requirements of the country where the research is being conducted, paying attention to the applicability of international norms and standards [3]. However, it remains unclear to what extent ethics committees are familiar with electronic informed consent and how they will handle electronic informed consent. Moreover, the involvement of research participants in the design of an electronic informed consent platform is of utmost importance as they fulfill a central role in biomedical research. Personalized human-centered design enables understanding of the participants' experiences and incorporation of their feedback to facilitate a participant-centered electronic informed consent platform [17].

Despite an increasing number of studies relevant for electronic informed consent, a comprehensive overview of these studies across the ethical, legal, regulatory, and user interface domains is lacking. Hence, the primary outcome of this systematic review was to provide a descriptive overview of the perspectives of different stakeholder groups (ie, researchers, research participants, health authorities, and ethics committee members) in these different domains with regard to electronic informed consent in biomedical research. The secondary outcome aimed to provide recommendations to assist responsible implementation. Insights of this review may serve as the foundation to design an electronic informed consent platform, thereby taking scientific steps forward in view of the international state-of-the-art.

**Methods**

This systematic review was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [18]. The corresponding review protocol is registered in the International Prospective Register of Systematic Reviews (PROSPERO CRD42020158979).

**Search Strategy**

The electronic databases of Web of Science (Core collection), PubMed, EMBASE, ACM Digital Library, and PsycARTICLES were searched in order to retrieve potential eligible studies. The searches for all databases were performed November 14, 2019. A search string for PubMed was developed and consisted of Medical Subject Headings terms and free-text words. The search string was adjusted for use in the other electronic databases. All search strings were verified by a health sciences librarian. These search strings consisted of keywords such as *electronic (informed) consent*, *dynamic (informed) consent*, *e-consent*, *digital (informed) consent*, *interactive (informed) consent*, *online (informed) consent*, *multimedia*, and *telemedicine*. The search was restricted to studies published in English after January 1, 2010. This timeframe was justified by the fact that electronic informed consent in biomedical research gained popularity only

in the last decade and that technology has evolved quite rapidly. The full search strategy for all databases can be found in [Multimedia Appendix 1](#).

### Study Selection and Criteria

All study types (ie, qualitative, quantitative, and mixed methods) that discussed the concept of electronic informed consent in biomedical research covering the ethical, legal, regulatory, and user interface domains were included. Perspectives of stakeholders, in particular research participants (ie, patients and healthy volunteers), researchers, ethics committee members, and health authorities were considered relevant. We excluded nonempirical studies and abstracts. All studies retrieved from the search strategy were imported to EndNote X9 (Clarivate Analytics) and duplicates were removed. The remaining studies were uploaded to Rayyan (Qatar Computing Research Institute) software. Two researchers (EDS and DZ) independently performed the first screening based on titles and abstracts. In a second step, studies with full texts available were carefully reviewed by two researchers (EDS and DZ) and disagreements were resolved by consensus. When the full texts were not available, the corresponding authors were contacted. The reference lists of the included studies were hand searched for additional studies.

### Data Extraction and Analysis

Data extraction was performed independently by two researchers (EDS and DZ) and was subsequently checked. A dedicated Excel (Microsoft Inc) data extraction form was used retrieving the following information for each eligible study: study identification (first author, title, publication year); study characteristics (study period, country, design, objective, scenario); stakeholder group (research participants, researchers, ethics committee members, health authorities); tool used to collect information (survey, focus groups, interviews); intervention (a description of the electronic informed consent platform); domain being assessed (ethical, legal, regulatory, or user interface); and ethical, legal, regulatory or user interface perspectives regarding electronic informed consent

Two researchers (EDS and DZ) carried out data analysis together using Excel. A combination of deductive and inductive thematic analysis of the ethical, legal, regulatory, and user interface perspectives was used, reporting different concepts and the main findings associated with them. Thematic analysis was conducted

according to the six-phase approach described by Braun and Clarke [19]. During the first phase, notes were created on the ethical, legal, regulatory, or user interface perspectives found in literature. These notes were valuable for the creation of initial codes in the second step. During the third step, we clustered these codes to generate broad concepts. Thereafter, the concepts were thoroughly reviewed and were defined in the fourth and fifth step. In the sixth and last step of this approach, we provided a descriptive overview to summarize the concepts found in literature [19]. Within these concepts, studies reporting similar findings were grouped together in order to provide a concise overview of results.

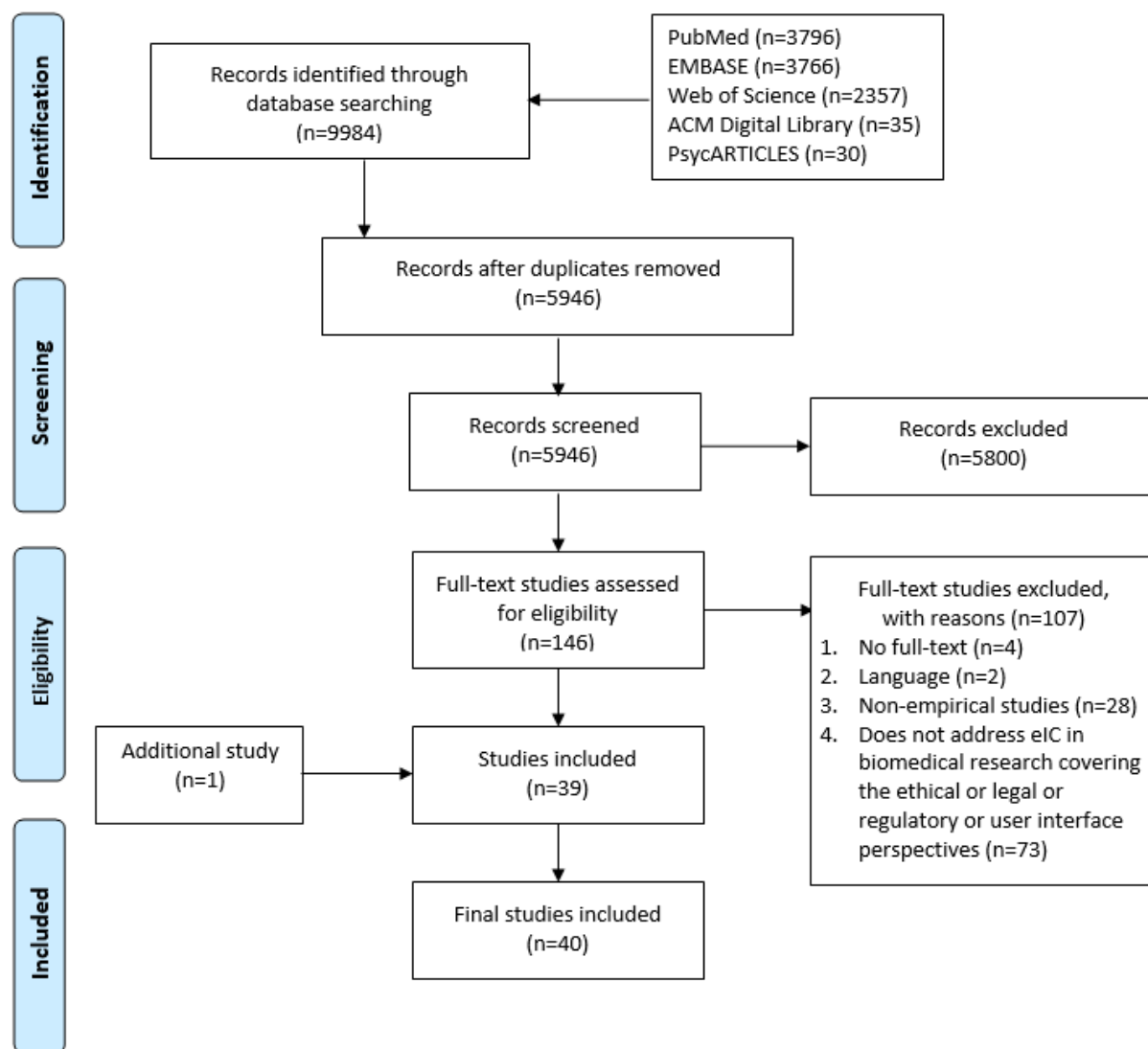
### Quality Assessment

Two researchers (EDS and DZ) assessed the quality of all included studies. Based on the study design, the Critical Appraisal Skills Programme (CASP) for qualitative studies [20], Mixed Methods Appraisal Tool (MMAT) for mixed methods studies [21], and Jadad tool for randomized controlled trials [22] were used. MMAT and CASP were used to criticize study aspects such as the aim and the methodology of the study [20,21]. These aspects were evaluated using a categorical scale (yes, indeterminate, or no), and thereafter, we converted the number of positive assessments into percentages. The quality of randomized controlled trials was assessed by considering the randomization, blinding, dropouts, and withdrawals. Randomized controlled trials could receive up to 5 points using the Jadad tool [22]. The Newcastle-Ottawa scale for cohort studies and an adapted version of this scale for cross-sectional studies were applied. Using the Newcastle-Ottawa scale, cohort and cross-sectional studies were evaluated using the following quality parameters: selection of study groups, comparability of study groups, and ascertainment of the outcome. Moreover, this scale also assessed the statistical test used in cross-sectional studies. Scores ranged from 0 to 9 for cohort studies and from 0 to 10 for cross-sectional studies [23,24].

## Results

### Search Results

Our search strategy produced a total of 9984 studies. After the screening process ([Figure 1](#)), 40 studies were included in the final analysis.

**Figure 1.** PRISMA flowchart of the systematic review. eIC: electronic informed consent.

### Characteristics of Included Literature

[Multimedia Appendix 2](#) summarizes the main characteristics for each study. The majority of studies were conducted in the United States (27 studies), followed by the United Kingdom (3 studies). There was 1 study each in Ireland, Korea, Canada, Australia, Spain, Germany, and Denmark; 3 studies reported results from several countries. The designs of the studies included 10 randomized controlled trials, 10 mixed methods studies, 10 cross-sectional studies, 9 qualitative studies, and 1 cohort study. Most stakeholder groups involved in the studies were represented by research participants (26 studies), followed by researchers (2 studies), and ethics committee members (1 study). In 11 studies, more than 1 of these categories, including health authorities, were involved. As for the characteristics of the research participants, 22 studies included different age categories (range 18-88 years), 3 studies each were conducted in young and elderly populations, and only 1 study was conducted in children. No information regarding age was reported in 5 studies involving research participants. When the information was provided, we noticed that, in 12 studies, the majority of research participants had a college degree or more,

in 8 studies the majority of research participants had less than a college degree, and in 2 studies there was an equal distribution in education level of the research participants. Meanwhile, in 12 studies there was no information regarding the education level of the research participants. Studies were conducted in both female and male populations, except 4 studies that only included either women or men. The tools used to collect relevant information varied, based, also, on the study design. [Multimedia Appendix 3](#) shows an overview of the intervention, scenario, and tools used.

### Quality Assessment

The quality of the included studies varied widely. Among the randomized controlled trials, only 3 studies [25-27] received a quality score of 3. Quality scores of 2 and 1 were provided to 1 study [28] and 6 studies [29-34], respectively. The qualitative studies ranged from satisfying 80% or more of the quality criteria (6 studies) [35-40] to satisfying 50% to 60% of the quality criteria (3 studies) [41-43]. Moreover, 3 mixed methods studies met 80% or more of the quality criteria [16,44,45], 6 studies met from 57% to 71% of the quality criteria [15,46-50], while 1 met only 43% of the quality criteria [51]. The only



cohort study had a quality score of 7 [52], while the quality scores of cross-sectional studies were assessed to be 4 [53-56], 5 [57-61], and 6 [62].

## Domain Perspectives

Perspectives of the ethical, legal, regulatory, and user interface domains were reported using 6 concepts: format, impact on understanding, acceptability, security and trust, storage, and content. Due to the cross-disciplinary nature of these perspectives, they were not reported for each domain individually as each of these concepts touches on 2 or more domains.

### Format

A range of aspects related to the format were identified in the literature review. The majority of research participants, researchers, members of the ethics committee as well as health authorities believed that the ability to incorporate audio, video, or graphics in electronic informed consent is a distinct benefit, especially for vulnerable groups [15,16,28,33,39,43,45,47-49,57]. Patients who were older adults, for example, expressed the usefulness of graphics and audio [45]. Research participants mentioned in a focus group that video and graphics may be more effective than written text in conveying information [16]. Moreover, patients involved in a study of Simon et al [39] indicated that audio narration could be of help for research participants with poor eyesight or limited literacy. However, participants argued for caution because background music and sound effects could be an added distraction [50]. In contrast, video was considered a multimedia element that might hold participants' attention more than a paper-based informed consent could [33]. Important to consider is the length of the video, since research participants highlighted that a 5-minute video was too long [33]. The use of graphics, including icons, was appreciated by all involved stakeholder groups [15,16,28,39,43,45,47-49,57]. For instance, the implementation of a progress bar was advised by researchers to indicate the different steps of the electronic informed consent form [43]. A pilot study [48] involving children highlighted the entertainment of graphics, making the electronic informed consent platform a pleasure to use. On the contrary, adult patients with fragile X syndrome, a genetic disorder causing intellectual impairment, had difficulty understanding aspects of clinical research such as blinding and randomization explained by several animations [47].

Furthermore, hyperlinks were identified as an important aspect of the format by participants and researchers [16,43,56]. Hyperlinks could be used as a video link between researchers and participants in order to combine an online with a face-to-face consent process [56]. Moreover, patients engaged in focus groups elicited hyperlinks as an encouraging way to seek additional information [16]. Nevertheless, it should be noted that only 20 out of 491 patients (4.1%) involved in a randomized controlled trial clicked on 1 or more hyperlinks [26]. Members of the ethics committee and health authorities advised avoiding the use of hyperlinks to webpages that may modify their content [15].

Researchers and ethics committee members as well as research participants criticized having extensive content during usability analysis of electronic informed consent platforms [34,54]. Research participants advised using a bullet point format with access to additional information if desired [28,34,50]. Simple, concise language should be used to encourage a sufficient level of understanding [15,28,40,50]. The possibility of marking information that participants did not understand was considered useful and was viewed as an opportunity to enhance the discussion with the research staff [15]. Nonetheless, research participants and researchers emphasized not losing the personal connection between researcher and participant [15,50].

Moreover, electronic informed consent offers the possibility to give information in several languages by using subtitles or translated text. Although many factors influence participant recruitment, researchers and research participants believed that electronic informed consent may facilitate recruitment, particularly in rare disease research [35]. By translating the content of electronic informed consent into different languages, information is available for a large number of individuals regardless of their geographical setting [35,53].

Little information was provided about the choice of the device when using an electronic informed consent platform. Two studies, both of them using touchscreen formats, reported incorrect end user input [51,53]. Researchers entered incorrect medical record numbers and patient names containing spelling errors [53]. A similar case with research participants was reported; in this particular study [51], several participants accidentally removed their signature, which contributed to the majority of errors when using touchscreen devices. Moreover, research participants who were older adults reported that they would need training to be able to use a tablet-based consent process [45].

### Impact on Understanding

Electronic informed consent may have an impact on the understanding of research participants by implementing a quiz, before signing the consent, to assess the participants' level of comprehension [28,35,39,44,55]. However, research participants argue for caution (to not make it feel like a test) [50]. In addition, interactive technology and a printed consent form offer the opportunity to review information giving participants the time to learn about the research [15,16,37,39,50]. Patients considered a paper-based form more permanent and suggested implementing a printout option in the electronic informed consent technology [39]. In addition, Vanaken et al [15] reported that some research participants preferred to take a printed version of the informed consent document home. Hence, the pressure to give consent immediately was decreased [16]. An extensive amount of studies investigated barriers related to the use of electronic informed consent which could prevent research participants from achieving a reasonable level of understanding [15,25-28,31-34,39,40,44,45,48,49,52,53,55,60,61]. For the majority of studies, no barriers were reported [25-28,31-34,44,48,52]. For example, patients with mental disorders, of whom the majority had a primary level of education, were able to make well-informed decisions by using electronic informed consent [52]. Moreover, patients with

schizophrenia had better understanding of disclosed information when using electronic informed consent compared to when using paper-based informed consent [32]. In a few studies, barriers impeding adequate comprehension were reported by research participants, researchers as well as ethics committee members regarding people with limited computer literacy, visual or auditory impairment, and regarding the lack of access to computers or internet [15,39,40,45,53,60,61]. Other reported barriers to achieving adequate understanding were the attitudes of research participants looking for additional information after reading a very concise informed consent and the use of graphics that may be unclear [47,49,55].

### Acceptability

Some researchers had concerns regarding approval of electronic informed consent as an alternative consent process by ethics committees [16,61]. In addition, ethics committees themselves expressed uncertainty about the impact of the audio-visual aspects on the ethical review process and the review duration [15,59,61]. The ability to confirm the identity of research participants consenting remotely may be challenging, especially for researchers [57]. In general, varying perceptions were reported regarding the use of electronic signatures. Although the majority of ethics committee chairpersons enrolled in a study by Kane et al [59] did not encounter submissions of informed consent documents containing electronic signatures, a large number of involved chairs would approve it. However, researchers and members of ethics committees were unsure about compliance with local regulations [15,59]. Some representatives of health authorities supported the use of electronic signatures while others were concerned about data privacy [15]. With respect to the research participants, a study by Haussen et al [60], involving legal authorized representatives of whom 21 out of 53 representatives (40%) had a low educational status, stated that for these participants there is an increased chance of preferring a paper-based informed consent.

### Security and Trust

Research participants stressed the importance of trust in the authenticity of electronic informed consent to share health data and to agree to take part in the study [36]. This went hand-in-hand with security of the electronic informed consent platform, which was a main concern for all stakeholders [15,34,36,39,40,42,50,61]. A secure platform may enable the transfer of files, which researchers considered an important factor in biomedical research [57]. Chhin et al [53] reported that the electronic informed consent platform that they developed could only be accessed by using individual user accounts and passwords. Moreover, researchers mentioned the need for providing sufficient information on privacy aspects to potential research participants [43]. According to research participants, electronic informed consent may enhance trust in research because of the possibility of returning research information by using innovative technology [38]. Nevertheless, Harle et al [26] conducted a randomized controlled trial with patients receiving standard, interactive-only, or interactive trust-enhanced electronic informed consent. Trust-enhanced messages were implemented containing additional information on data protections, regulations, and training of the research

staff. They indicated there was no effect from the inclusion of these trust-enhanced messages on data sharing, satisfaction, and understanding [26].

### Storage

According to multiple stakeholders, electronic storage of consent details constitutes a notable benefit by enabling researchers to have a trustworthy and traceable overview of the consent status of participants [15,35]. For example, electronic informed consent can support online withdrawal, together with documentation of the reasons for withdrawal [41,46]. Moreover, online storage improves version control which might reduce the number of adverse inspection findings by the health authorities. Nevertheless, health authorities voiced concerns that informed consent forms may be inaccessible during an inspection [15]. An important feature for researchers and the ethics committee is to have the ability to control access for specific consent documents [57]. According to research participants, access to their own health information is a key benefit of participation [46]. Research participants requested more transparency regarding the use of their data [38,58]. They expressed the right to control the sharing and use of their private health information [37,42]. Moreover, electronic informed consent enables researchers to update participants frequently with information about preliminary results, follow-up studies, and main outcomes [35]. Nevertheless, the majority of patients with mental disorders, involved in a study by Sundby et al [62], stated that they would prefer direct contact with the research staff for receiving genomic information concerning serious or life-threatening conditions.

### Content

The usefulness of additional content elements such as definitions was articulated in a focus group involving research participants who were older adults [45]. Moreover, exposing research participants to social annotation, such as comments generated by end users on several aspects of the electronic informed consent form, was considered important to feel adequately informed [29]. However, the emotional force communicated in social annotations has an influence on research participants' perceptions with regard to information given in electronic informed consent. Research participants exposed to positive valence annotations indicated feeling less informed than participants receiving a combination of positive and negative valence annotations [30]. Various studies [15,37,58,62] provided insights into a personalized approach of an electronic informed consent platform. Research participants stated that they would like to receive personalized elements and tailored information such as the display of their name in the electronic informed consent form or the impact of their contribution on a specific research question [15,37]. Moreover, Kim et al [58] conducted a study using an electronic informed consent platform in which research participants were allowed to modify their preferences for data sharing. In this study [58], research participants indicated that a personalized approach could enable participants' eagerness for data sharing for research purposes. Furthermore, researchers appreciated the possibility of research participants indicating what kind of information they would like to obtain [62].

## Discussion

### Importance of Understanding in Electronic Informed Consent

The majority of studies paid particular attention to understanding, considering that it is seen as a crucial point in enabling participants to give their informed consent. In some studies [28,33,34,39,44-46,48,55,60,62], the majority of participants had a high level of education. It should also be noted that a number of studies [27,29,51] assessing comprehension did not report the education level of their participants. Therefore, there might be additional barriers for less educated participants to achieving an adequate level of understanding. If research participants are not adequately informed about a research study, they may be disappointed due to misconceptions of the benefits. As a result, researchers may face increased dropout. Moreover, research participants may distrust ethics committees when they do not put the research participant at center when reviewing an electronic informed consent. Literature reported that electronic informed consent could improve understanding through the opportunity to check participants' level of comprehension by using quizzes before electronically signing the consent form [28,35,39,44,55]. The implementation of a quiz may prevent research participants from immediately agreeing to participate in research, not allowing the content to be reviewed thoroughly. Nevertheless, meta-analysis of different informed consent interventions conducted by Nishimura et al [10] showed a significant higher understanding for paper-based enhanced consent forms including simplified text and facilitated reading level. A non-significant increase of understanding was observed for multimedia interventions. These enhanced consent forms and multimedia approaches were compared with a control consent process that consisted of a paper-based informed consent or an already enhanced informed consent [10]. It would, therefore, seem that further research is needed to explore the effect on understanding of an electronic informed consent platform including all of these aspects, considering the education level, age, health status, and health literacy of the participant.

### Particular Attention to Specific Population Groups

Several studies [15,39,40,45,60,61] included in this review reported concerns about access to electronic informed consent for specific population groups. From an ethical point of view, different population groups need to have the opportunity to be represented in biomedical research. The possibility for several population groups to take part in a research study may, first, broaden their access to treatments, and second, positively impact the generalizability of research results. Adequate support is required for participants with, for example, no or limited computer literacy. Obtaining informed consent in people with mental disorders remains a challenge that may be overcome with electronic informed consent. Health authorities believe that vulnerable populations, which are often underrepresented in clinical research, might benefit [15]. However, it is important to highlight that only a limited number of studies [25,31,32,45,47,48,62] included vulnerable groups who may have particular requirements. In only 1 pilot study, perceptions

of children regarding the use of graphics were evaluated [48]. Recommendations cannot be inferred from this pilot study because they require verification in further research. Generally, electronic informed consent platforms are intended to be used by multiple dissimilar target groups. The target population varies in literacy, education, age, health condition, and many other factors. Therefore, preferences for designing a usable interface may deviate across the type of end users. Visual factors, such as the font size or the use of graphics, will differ for research studies involving older adults, children, or visually impaired participants.

### Personal Connection

Attention needs to be paid to not losing the personal connection between research participants and research staff. Electronic informed consent has the opportunity to supplement the existing paper-based consent process but is not meant to replace it. For example, research participants may prefer the paper-based consent document or electronic informed consent facilitated by a discussion with the research staff to enable an informed decision. To enhance trust in research or in the authenticity of an electronic informed consent platform, the personal connection may play an important role. Moreover, electronic informed consent has the potential to inform research participants about early findings, but participants involved in a study by Sundby et al [62] preferred having direct contact with the research team if this information was related to life-threatening conditions. In addition, the online consent form should provide participants with a link that they can enter to withdraw their consent any time they desire. Nevertheless, for several types of medication, it is dangerous to abruptly stop [63]. For this reason, the interaction between the research participant and the research team is of utmost importance to prevent withdrawal symptoms by giving all necessary information. In general, a lack of face-to-face communication may lead to misunderstanding with regard to several aspects of a research study.

### Guidance Framework

Acceptability of electronic informed consent by the health authorities was unclear [15,59]. Therefore, analysis of the legal framework regarding different aspects of electronic informed consent is required to determine the restrictions in certain jurisdictions. More research is needed to explore the legal aspects related to electronic informed consent across several countries. In addition, uncertainty exists about the review process of electronic informed consent by ethics committees [15,59,61]. Due to the lack of experience, it remains unclear how the evaluation process of ethics committees will be impacted. It would thus be of interest to investigate which aspects of the electronic informed consent process ethics committees would consider for evaluation. General consensus on the review process of ethics committees is required to facilitate harmonized review for multicentric studies in which multiple ethics committees are involved. In addition, researchers need to receive clarification on which materials they need to submit to ethics committees. A guidance framework could reduce the burden of researchers and ethics committees concerning the preparation and review of electronic informed



consent and could ensure protection of the rights, safety, and welfare of human participants.

### Privacy by Design

Notable concerns were expressed by several stakeholders regarding security of electronic informed consent platforms, and thus data privacy [15,34,36,39,40,42,50,61]. Because electronic informed consent establishes the opportunity to give consent remotely, capture of an electronic signature and proof of identity are challenging. Researchers and ethics committees raised concerns about low compliance of electronic signature with local regulations. These concerns may strengthen reluctance to implement and use electronic informed consent platforms in biomedical research. In order to secure the platform, potential security threats need to be identified to counter them in security software. It is of utmost importance to implement the highest standards in security and privacy design to prevent fraudulent use of participants' health data. Moreover, clarifying the integrated security and privacy aspects to the research participants is valuable to reduce these concerns. A breach of privacy could, for research participants, lead to job loss or consequences related to their insurance. Moreover, researchers could be held responsible for misconduct. Ethics committees and health authorities could also be held responsible, as they did not look into the security of the electronic informed consent platform when reviewing or inspecting. In previous years, the failure to obtain and document the consent of research participants was part of frequent FDA inspection failures [64]. Owing to online storage, electronic informed consent may contribute to, for example, better version control and documentation of the consent process [15,35]. Nevertheless, controlled access systems need to be implemented to restrict access for stakeholders for different types of content in order to respect the privacy of the participants.

### Usability

Noteworthy is the discrepancy between quantitative and qualitative usability testing. Despite the mention of hyperlinks as an important feature of the user interface in a qualitative focus group study [16], this feature was barely used in a randomized controlled trial evaluating different electronic informed consent platforms [26]. These results suggest using an iterative design cycle starting with an evaluation of what participants would consider as useful in the user interface, after which a usability analysis should be performed. It is crucial to involve research participants when designing an electronic informed consent platform to create a user-friendly and acceptable system. This will support researchers in ensuring an adequate number of participants in order to detect possible important clinical findings. Moreover, although research participants indicated preferring concise information in consent forms [15,28,40,50], attention needs to be paid to participants who are not motivated to seek additional information. Specific measures are needed if input of researchers or research participants is required when using electronic informed consent in order to prevent errors. Potential errors are related to devices

used by biomedical research stakeholders and the amount of manual input required. The use of touchscreen devices might invoke more inadvertent actions from accidental touches on the display, compared to those from the use of computers. The implementation of a quality assurance procedure is highly recommended in order to avoid incorrect end-user input.

### Inclusivity

Language aspects may complicate obtaining informed consent, and thus recruitment. It requires additional effort to involve research participants whose primary language is not frequently used in the informed consent process. To make the informed consent process more inclusive, subtitles or translated text can be used in electronic informed consent. Additionally, intercultural mediation can be considered to reduce the adverse outcomes of language barriers. An intercultural mediator can be part of the electronic informed consent process to assist the linguistic interpretation of information in the electronic informed consent [65].

### Toward Personalized Electronic Informed Consent

An electronic informed consent platform provides an important means for modifying consent preferences. Electronic informed consent can be tailored to participants' needs to address their preferences. In this way, participants may change different aspects such as which information they would like to receive, how they prefer to be contacted, and how often they wish to be contacted. Research participants indicated that access to their personal health data is an important motivation to participate in research studies. By enabling a personalized approach in the electronic informed consent platform, research participants can indicate whether they would like to receive this information. Moreover, electronic informed consent can be personalized for different kind of diseases. Patient preferences can vary for disease-specific informed consents. Participants with delicate health conditions may, for example, require additional information related to the study and may not be eager to share their data. Personalization may improve research participants' engagement and understanding. However, attention needs to be paid to the review process of such personalized electronic informed consent in order to avoid a time-consuming process. Further research that considers several types of end users is recommended, as electronic informed consent must be tailored to different population needs. Moreover, electronic informed consent provides the possibility of updating research participants with information on an ongoing research study. Nevertheless, it needs to be investigated how an electronic informed consent platform can be personalized and how longitudinal interaction can be assured.

### Recommendations

The ethical, legal, regulatory, and user interface perspectives were converted into recommendations to facilitate implementation of electronic informed consent in biomedical research. The recommendations are based upon the 6 concepts reported in the results and are shown in Table 1.

**Table 1.** Recommendations to guide implementation of electronic informed consent in biomedical research.

Concept	Description
Format	<ul style="list-style-type: none"> <li>Implement audio, video, graphics (ie, icons, progress bar) and hyperlinks. Note: do not add audio that may distract participants, consider the length of the video, create understandable graphics and avoid hyperlinks to webpages with dynamic content</li> <li>Use simple, concise language and implement a bullet point format. Note: provide access to additional information if desired</li> <li>Give the possibility to research participants to highlight information that is difficult to understand in order to facilitate the discussion with the research staff</li> <li>Depending on the research study, make electronic informed consent available in multiple languages by using subtitles or translated text</li> <li>Implement a quality assurance process to check the input of the end user</li> </ul>
Impact on understanding	<ul style="list-style-type: none"> <li>Pay attention to the personal connection between the research participants and the research staff (also refers to security and trust)</li> <li>Implement quizzes to assess the participants' level of comprehension. Note: do not let the quizzes feel like an evaluation</li> <li>Give the possibility to review information by using interactive technology or the printed electronic informed consent form</li> <li>Guarantee adequate support for people with limited computer literacy, visual/auditory impairment and people who do not have access to internet or computers</li> </ul>
Acceptability	<ul style="list-style-type: none"> <li>Collaborate with health authorities and ethics committees to create a framework for reviewing and implementing electronic informed consent</li> </ul>
Security and trust	<ul style="list-style-type: none"> <li>Implement controlled access systems for several stakeholder groups and pay attention to a secured electronic informed consent platform. Note: provide sufficient information to potential research participants about privacy aspects of the platform</li> <li>Make sure that the secure transfer of files is possible between stakeholder groups</li> </ul>
Storage	<ul style="list-style-type: none"> <li>Provide online storage of the informed consent</li> <li>Support online withdrawal with documentation of the reasons for withdrawal</li> <li>Pay attention to transparency regarding the use of participants' health information and their right to control the sharing and use of this information</li> <li>Implement the possibility to update research participants frequently with information about preliminary results, follow-up studies and main outcomes</li> </ul>
Content	<ul style="list-style-type: none"> <li>Implement definitions</li> <li>Implement social annotations but mind the emotional force</li> <li>Implement a personalized approach (eg, by letting the participants indicate what kind of information they would like to receive)</li> </ul>

## Limitations

Our systematic review only included literature published in English. As a consequence, selection bias could have been introduced because it is possible that information specific to the ethical, legal, regulatory, or user interface domain of electronic informed consent in languages other than English was not identified. Another limitation was that the methodological quality of the included studies was rated overall as moderate which needs to be considered when interpreting the reported findings. Most of the studies had small sample sizes, which could be an issue for the generalizability of results. The studies included in the review had a high level of heterogeneity among them, which is the reason why a descriptive analysis was conducted. The differences in study designs, research fields and contexts, populations, and tools

used to assess the results may have impacted the findings reported by each study.

## Conclusions

This systematic review highlights different opportunities and challenges for responsible implementation of electronic informed consent in biomedical research. Electronic informed consent provides the possibility of enforcing a personalized approach and supporting a longitudinal interaction with research participants. Findings suggest that electronic informed consent may have a beneficial impact on the consent process as long as some requirements are fulfilled. Special attention needs to be paid to specific population groups and to personal interaction with research staff. Future high-quality research, especially using randomized controlled trials, is required to provide information that may encourage the use of electronic informed consent for vulnerable groups.



## Acknowledgments

This research was funded by *KU Leuven* and by the European Network Staff Exchange for Integrating Precision Health in the Health Care Systems project (Marie Skłodowska-Curie Research and Innovation Staff Exchange No. 823995). Pascal Borry is funded through a European Union's Horizon 2020 research and innovation programme under grant agreement No. 825903 (euCanShare).

The authors wish to thank Thomas Vandendriessche, Magdalena Jans, and Krizia Tuand, the biomedical reference librarians of the *KU Leuven Libraries 2Bergen-learning Centre Désiré Collen*, for their help in conducting the systematic literature search.

## Authors' Contributions

EDS, DG, PB, and IH designed the study. EDS and DZ conducted the search, selection, screening, and quality assessment of the studies. A first draft of the paper was written by EDS and DZ and was critically revised by DG, PB, SB, MLDP, and IH. All authors approved the final paper.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Search strategy.

[DOC File, 31 KB - [jmir\\_v22i10e19129\\_app1.doc](#)]

### Multimedia Appendix 2

Main characteristics of the 40 included studies.

[DOCX File, 30 KB - [jmir\\_v22i10e19129\\_app2.docx](#)]

### Multimedia Appendix 3

Intervention, scenario, and tools of the 40 included studies.

[DOCX File, 19 KB - [jmir\\_v22i10e19129\\_app3.docx](#)]

## References

1. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC. European Commission. URL: [https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg\\_2014\\_536/reg\\_2014\\_536\\_en.pdf](https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf) [accessed 2020-09-26]
2. The Nuremberg Code (1947). *BMJ* 1996 Dec 07;313(7070):1448-1448. [doi: [10.1136/bmj.313.7070.1448](https://doi.org/10.1136/bmj.313.7070.1448)]
3. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013 Nov 27;310(20):2191-2194. [doi: [10.1001/jama.2013.281053](https://doi.org/10.1001/jama.2013.281053)] [Medline: [24141714](https://pubmed.ncbi.nlm.nih.gov/24141714/)]
4. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Access to European Union Law. 2016. URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> [accessed 2020-09-27]
5. Dankar FK, Gergely M, Dankar SK. Informed Consent in Biomedical Research. *Comput Struct Biotechnol J* 2019;17:463-474 [FREE Full text] [doi: [10.1016/j.csbj.2019.03.010](https://doi.org/10.1016/j.csbj.2019.03.010)] [Medline: [31007872](https://pubmed.ncbi.nlm.nih.gov/31007872/)]
6. Getz KA, Campo RA. Trial watch: Trends in clinical trial design complexity. *Nat Rev Drug Discov* 2017 May;16(5):307. [doi: [10.1038/nrd.2017.65](https://doi.org/10.1038/nrd.2017.65)] [Medline: [28417986](https://pubmed.ncbi.nlm.nih.gov/28417986/)]
7. Nathe JM, Krakow EF. The challenges of informed consent in high-stakes, randomized oncology trials: a systematic review. *MDM Policy Pract* 2019;4(1):2381468319840322. [doi: [10.1177/2381468319840322](https://doi.org/10.1177/2381468319840322)] [Medline: [30944886](https://pubmed.ncbi.nlm.nih.gov/30944886/)]
8. Van Norman GA. Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs. *JACC Basic Transl Sci* 2016 Apr;1(3):170-179 [FREE Full text] [doi: [10.1016/j.jacbs.2016.03.002](https://doi.org/10.1016/j.jacbs.2016.03.002)] [Medline: [30167510](https://pubmed.ncbi.nlm.nih.gov/30167510/)]
9. Falagas ME, Korbila IP, Giannopoulou KP, Kondilis BK, Peppas G. Informed consent: how much and what do patients understand? *Am J Surg* 2009 Sep;198(3):420-435. [doi: [10.1016/j.amjsurg.2009.02.010](https://doi.org/10.1016/j.amjsurg.2009.02.010)] [Medline: [19716887](https://pubmed.ncbi.nlm.nih.gov/19716887/)]
10. Nishimura A, Carey J, Erwin PJ, Tilburt JC, Murad MH, McCormick JB. Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized control trials. *BMC Med Ethics* 2013;14:28 [FREE Full text] [doi: [10.1186/1472-6939-14-28](https://doi.org/10.1186/1472-6939-14-28)] [Medline: [23879694](https://pubmed.ncbi.nlm.nih.gov/23879694/)]
11. Kraft SA, Constantine M, Magnus D, Porter KM, Lee SS, Green M, et al. A randomized study of multimedia informational aids for research on medical practices: Implications for informed consent. *Clin Trials* 2017 Feb;14(1):94-102 [FREE Full text] [doi: [10.1177/1740774516669352](https://doi.org/10.1177/1740774516669352)] [Medline: [27625314](https://pubmed.ncbi.nlm.nih.gov/27625314/)]

12. Synnot A, Ryan R, Prictor M, Fetherstonhaugh D, Parker B. Audio-visual presentation of information for informed consent for participation in clinical trials. *Cochrane Database Syst Rev* 2014 May 09(5):CD003717 [FREE Full text] [doi: [10.1002/14651858.CD003717.pub3](https://doi.org/10.1002/14651858.CD003717.pub3)] [Medline: [24809816](#)]
13. Use of electronic informed consent: questions and answers. Guidance for institutional review boards, investigators, and sponsors. US Food and Drug Administration (FDA). 2016. URL: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-electronic-informed-consent-clinical-investigations-questions-and-answers> [accessed 2020-09-27]
14. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. *Eur J Hum Genet* 2015 Feb;23(2):141-146 [FREE Full text] [doi: [10.1038/ejhg.2014.71](https://doi.org/10.1038/ejhg.2014.71)] [Medline: [24801761](#)]
15. Vanaken H, Masand SN. Awareness and collaboration across stakeholder groups important for eConsent achieving value-driven adoption. *Ther Innov Regul Sci* 2019 Nov;53(6):724-735. [doi: [10.1177/2168479019861924](https://doi.org/10.1177/2168479019861924)] [Medline: [31319717](#)]
16. Anderson EE, Newman SB, Matthews AK. Improving informed consent: stakeholder views. *AJOB Empir Bioeth* 2017;8(3):178-188 [FREE Full text] [doi: [10.1080/23294515.2017.1362488](https://doi.org/10.1080/23294515.2017.1362488)] [Medline: [28949896](#)]
17. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-centered design study: enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users. *JMIR Mhealth Uhealth* 2017 May 30;5(5):e71 [FREE Full text] [doi: [10.2196/mhealth.7046](https://doi.org/10.2196/mhealth.7046)] [Medline: [28559227](#)]
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](#)]
19. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
20. CASP Qualitative Checklist. Critical Appraisal Skills Programme. URL: <https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf> [accessed 2020-09-27]
21. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. *J Clin Epidemiol* 2019 Jul;111:49-59.e1 [FREE Full text] [doi: [10.1016/j.jclinepi.2019.03.008](https://doi.org/10.1016/j.jclinepi.2019.03.008)] [Medline: [30905698](#)]
22. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996 Mar;17(1):1-12. [doi: [10.1016/0197-2456\(95\)00134-4](https://doi.org/10.1016/0197-2456(95)00134-4)] [Medline: [8721797](#)]
23. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomized studies in meta-analysis. Ottawa Hospital Research Institute. URL: [http://www.ohri.ca/programs/clinical\\_epidemiology/nosgen.pdf](http://www.ohri.ca/programs/clinical_epidemiology/nosgen.pdf) [accessed 2020-09-27]
24. Herzog R, Álvarez-Pasquín MJ, Díaz C, Del Barrio JL, Estrada JM, Gil A. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. *BMC Public Health* 2013 Feb 19;13:154 [FREE Full text] [doi: [10.1186/1471-2458-13-154](https://doi.org/10.1186/1471-2458-13-154)] [Medline: [23421987](#)]
25. Warriner AH, Foster PJ, Mudano A, Wright NC, Melton ME, Sattui SE, et al. A pragmatic randomized trial comparing tablet computer informed consent to traditional paper-based methods for an osteoporosis study. *Contemp Clin Trials Commun* 2016 Aug 15;3:32-38 [FREE Full text] [doi: [10.1016/j.conctc.2016.02.003](https://doi.org/10.1016/j.conctc.2016.02.003)] [Medline: [29736454](#)]
26. Harle CA, Golembiewski EH, Rahmanian KP, Brumback B, Krieger JL, Goodman KW, et al. Does an interactive trust-enhanced electronic consent improve patient experiences when asked to share their health records for research? A randomized trial. *J Am Med Inform Assoc* 2019 Jul 01;26(7):620-629 [FREE Full text] [doi: [10.1093/jamia/ocz015](https://doi.org/10.1093/jamia/ocz015)] [Medline: [30938751](#)]
27. Bobb MR, Van Heukelom PG, Faine BA, Ahmed A, Messerly JT, Bell G, et al. Telemedicine Provides Noninferior Research Informed Consent for Remote Study Enrollment: A Randomized Controlled Trial. *Acad Emerg Med* 2016 Jul;23(7):759-765 [FREE Full text] [doi: [10.1111/acem.12966](https://doi.org/10.1111/acem.12966)] [Medline: [26990899](#)]
28. Simon CM, Klein DW, Schartz HA. Interactive multimedia consent for biobanking: a randomized trial. *Genet Med* 2016 Jan;18(1):57-64 [FREE Full text] [doi: [10.1038/gim.2015.33](https://doi.org/10.1038/gim.2015.33)] [Medline: [25834945](#)]
29. Balestra M, Shaer O, Okerlund J, Ball M, Nov O. The effect of exposure to social annotation on online informed consent beliefs and behavior. Presented at: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; 2016; San Francisco, California, USA. [doi: [10.1145/2818048.2820012](https://doi.org/10.1145/2818048.2820012)]
30. Balestra M, Shaer O, Okerlund J, Westendorf L, Ball M, Nov O. Social Annotation Valence: The Impact on Online Informed Consent Beliefs and Behavior. *J Med Internet Res* 2016 Jul 20;18(7):e197 [FREE Full text] [doi: [10.2196/jmir.5662](https://doi.org/10.2196/jmir.5662)] [Medline: [27439320](#)]
31. Ham DY, Choi WS, Song SH, Ahn Y, Park HK, Kim HG, et al. Prospective Randomized Controlled Study on the Efficacy of Multimedia Informed Consent for Patients Scheduled to Undergo Green-Light High-Performance System Photoselective Vaporization of the Prostate. *World J Mens Health* 2016 Apr;34(1):47-55 [FREE Full text] [doi: [10.5534/wjmh.2016.34.1.47](https://doi.org/10.5534/wjmh.2016.34.1.47)] [Medline: [27169129](#)]

32. Harmell AL, Palmer BW, Jeste DV. Preliminary study of a web-based tool for enhancing the informed consent process in schizophrenia research. *Schizophr Res* 2012 Nov;141(2-3):247-250 [[FREE Full text](#)] [doi: [10.1016/j.schres.2012.08.001](https://doi.org/10.1016/j.schres.2012.08.001)] [Medline: [22939457](#)]
33. Rothwell E, Wong B, Rose NC, Anderson R, Fedor B, Stark LA, et al. A randomized controlled trial of an electronic informed consent process. *J Empir Res Hum Res Ethics* 2014 Dec;9(5):1-7 [[FREE Full text](#)] [doi: [10.1177/1556264614552627](https://doi.org/10.1177/1556264614552627)] [Medline: [25747685](#)]
34. Rowbotham MC, Astin J, Greene K, Cummings SR. Interactive informed consent: randomized comparison with paper consents. *PLoS One* 2013;8(3):e58603 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0058603](https://doi.org/10.1371/journal.pone.0058603)] [Medline: [23484041](#)]
35. Budin-Ljøsne I, Teare HJA, Kaye J, Beck S, Bentzen HB, Caenazzo L, et al. Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. *BMC Med Ethics* 2017 Jan 25;18(1):4 [[FREE Full text](#)] [doi: [10.1186/s12910-016-0162-9](https://doi.org/10.1186/s12910-016-0162-9)] [Medline: [28122615](#)]
36. Wood F, Kowalczyk J, Elwyn G, Mitchell C, Gallacher J. Achieving online consent to participation in large-scale gene-environment studies: a tangible destination. *J Med Ethics* 2011 Aug;37(8):487-492. [doi: [10.1136/jme.2010.040352](https://doi.org/10.1136/jme.2010.040352)] [Medline: [21478424](#)]
37. Teare HJ, Morrison M, Whitley EA, Kaye J. Towards 'Engagement 2.0': Insights from a study of dynamic consent with biobank participants. *Digit Health* 2015;1:2055207615605644 [[FREE Full text](#)] [doi: [10.1177/2055207615605644](https://doi.org/10.1177/2055207615605644)] [Medline: [29942545](#)]
38. Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient Perspectives on Sharing Anonymized Personal Health Data Using a Digital System for Dynamic Consent and Research Feedback: A Qualitative Study. *J Med Internet Res* 2016 Apr 15;18(4):e66 [[FREE Full text](#)] [doi: [10.2196/jmir.5011](https://doi.org/10.2196/jmir.5011)] [Medline: [27083521](#)]
39. Simon CM, Schartz HA, Rosenthal GE, Eisenstein EL, Klein DW. Perspectives on Electronic Informed Consent From Patients Underrepresented in Research in the United States: A Focus Group Study. *J Empir Res Hum Res Ethics* 2018 Oct;13(4):338-348. [doi: [10.1177/1556264618773883](https://doi.org/10.1177/1556264618773883)] [Medline: [29790410](#)]
40. Harle CA, Golembiewski EH, Rahmanian KP, Krieger JL, Hagmajer D, Mainous AG, et al. Patient preferences toward an interactive e-consent application for research using electronic health records. *J Am Med Inform Assoc* 2018 Mar 01;25(3):360-368 [[FREE Full text](#)] [doi: [10.1093/jamia/ocx145](https://doi.org/10.1093/jamia/ocx145)] [Medline: [29272408](#)]
41. Chen C, Turner SP, Sholle ET, Brown SW, Blau VLI, Brouwer JP, et al. Evaluation of a REDCap-based workflow for supporting federal guidance for electronic informed consent. *AMIA Jt Summits Transl Sci Proc* 2019;2019:163-172 [[FREE Full text](#)] [Medline: [31258968](#)]
42. Rowan W, O'Connor Y, Lynch L, Heavin C. Exploring User Behaviours when Providing Electronic Consent on Health Social Networks: A 'Just Tick Agree' Approach. *Procedia Computer Science* 2017;121:968-975. [doi: [10.1016/j.procs.2017.11.125](https://doi.org/10.1016/j.procs.2017.11.125)]
43. Iwaya LH, Li J, Fischer-Hübner S, Åhlfeldt R, Martucci LA. E-Consent for Data Privacy: Consent Management for Mobile Health Technologies in Public Health Surveys and Disease Surveillance. *Stud Health Technol Inform* 2019 Aug 21;264:1223-1227. [doi: [10.3233/SHTI190421](https://doi.org/10.3233/SHTI190421)] [Medline: [31438120](#)]
44. Abujarad F, Alfano S, Bright TJ, Kanno S, Grant N, Gueble M, et al. Building an Informed Consent Tool Starting with the Patient: The Patient-Centered Virtual Multimedia Interactive Informed Consent (VIC). *AMIA Annu Symp Proc* 2017;2017:374-383 [[FREE Full text](#)] [Medline: [29854101](#)]
45. Jayasinghe N, Moallem BI, Kakoullis M, Ojie M, Sar-Graycar L, Wyka K, et al. Establishing the Feasibility of a Tablet-Based Consent Process with Older Adults: A Mixed-Methods Study. *Gerontologist* 2019 Jan 09;59(1):124-134 [[FREE Full text](#)] [doi: [10.1093/geront/gny045](https://doi.org/10.1093/geront/gny045)] [Medline: [29757375](#)]
46. Doerr M, Maguire Truong A, Bot BM, Wilbanks J, Suver C, Mangravite LM. Formative Evaluation of Participant Experience With Mobile eConsent in the App-Mediated Parkinson mPower Study: A Mixed Methods Study. *JMIR Mhealth Uhealth* 2017 Feb 16;5(2):e14 [[FREE Full text](#)] [doi: [10.2196/mhealth.6521](https://doi.org/10.2196/mhealth.6521)] [Medline: [28209557](#)]
47. Furberg RD, Ortiz AM, Moultrie RR, Raspa M, Wheeler AC, McCormack LA, et al. A Digital Decision Support Tool to Enhance Decisional Capacity for Clinical Trial Consent: Design and Development. *JMIR Res Protoc* 2018 Jun 06;7(6):e10525 [[FREE Full text](#)] [doi: [10.2196/10525](https://doi.org/10.2196/10525)] [Medline: [29875084](#)]
48. Tait AR, Voepel-Lewis T, McGonegal M, Levine R. Evaluation of a prototype interactive consent program for pediatric clinical trials: a pilot study. *J Am Med Inform Assoc* 2012 Jun;19(e1):e43-e45 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000253](https://doi.org/10.1136/amiajnl-2011-000253)] [Medline: [21803924](#)]
49. Ramos SR. User-Centered Design, Experience, and Usability of an Electronic Consent User Interface to Facilitate Informed Decision-Making in an HIV Clinic. *Comput Inform Nurs* 2017 Nov;35(11):556-564 [[FREE Full text](#)] [doi: [10.1097/CIN.0000000000000356](https://doi.org/10.1097/CIN.0000000000000356)] [Medline: [28481754](#)]
50. Mahnke AN, Plasek JM, Hoffman DG, Partridge NS, Foth WS, Waudby CJ, et al. A rural community's involvement in the design and usability testing of a computer-based informed consent process for the Personalized Medicine Research Project. *Am J Med Genet A* 2014 Jan;164A(1):129-140 [[FREE Full text](#)] [doi: [10.1002/ajmg.a.36220](https://doi.org/10.1002/ajmg.a.36220)] [Medline: [24273095](#)]
51. Chalil Madathil K, Koikkara R, Obeid J, Greenstein JS, Sanderson IC, Fryar K, et al. An investigation of the efficacy of electronic consenting interfaces of research permissions management system in a hospital setting. *Int J Med Inform* 2013 Sep;82(9):854-863 [[FREE Full text](#)] [doi: [10.1016/j.ijmedinf.2013.04.008](https://doi.org/10.1016/j.ijmedinf.2013.04.008)] [Medline: [23757370](#)]

52. Morán-Sánchez I, Luna A, Pérez-Cárceles MD. Enhancing the informed consent process in psychiatric outpatients with a brief computer-based method. *Psychiatry Res* 2016 Nov 30;245:354-360. [doi: [10.1016/j.psychres.2016.08.045](https://doi.org/10.1016/j.psychres.2016.08.045)] [Medline: [27591409](https://pubmed.ncbi.nlm.nih.gov/27591409/)]
53. Chhin V, Roussos J, Michaelson T, Bana M, Bezjak A, Foxcroft S, et al. Leveraging Mobile Technology to Improve Efficiency of the Consent-to-Treatment Process. *JCO Clin Cancer Inform* 2017 Nov;1:1-8. [doi: [10.1200/CCI.17.00041](https://doi.org/10.1200/CCI.17.00041)] [Medline: [30657388](https://pubmed.ncbi.nlm.nih.gov/30657388/)]
54. Schneiderheinze H, Prokosch H, Apel H, Bellut L, Wullich B, Trollmann R, et al. Development and Usability Analysis of a Multimedia eConsent Solution. *Stud Health Technol Inform* 2019 Sep 03;267:297-303. [doi: [10.3233/SHTI190841](https://doi.org/10.3233/SHTI190841)] [Medline: [31483285](https://pubmed.ncbi.nlm.nih.gov/31483285/)]
55. Perrault EK, McCulloch SP. Concise Consent Forms Appreciated-Still Not Comprehended: Applying Revised Common Rule Guidelines in Online Studies. *J Empir Res Hum Res Ethics* 2019 Oct;14(4):299-306. [doi: [10.1177/1556264619853453](https://doi.org/10.1177/1556264619853453)] [Medline: [31169051](https://pubmed.ncbi.nlm.nih.gov/31169051/)]
56. McGowan CR, Houlihan CF, Kingori P, Glynn JR. The Acceptability of Online Consent in a Self-Test Serosurvey of Responders to the 2014-2016 West African Ebola Outbreak. *Public Health Ethics* 2018 Jul;11(2):201-212 [FREE Full text] [doi: [10.1093/phe/phx027](https://doi.org/10.1093/phe/phx027)] [Medline: [30135701](https://pubmed.ncbi.nlm.nih.gov/30135701/)]
57. Bunnell BE, Sprague G, Qanungo S, Nichols M, Magruder K, Lauzon S, et al. An Exploration of Useful Telemedicine-Based Resources for Clinical Research. *Telemed J E Health* 2020 Jan;26(1):51-65. [doi: [10.1089/tmj.2018.0221](https://doi.org/10.1089/tmj.2018.0221)] [Medline: [30785853](https://pubmed.ncbi.nlm.nih.gov/30785853/)]
58. Kim H, Bell E, Kim J, Sitapati A, Ramsdell J, Farcas C, et al. iCONCUR: informed consent for clinical data and bio-sample use for research. *J Am Med Inform Assoc* 2017 Mar 01;24(2):380-387 [FREE Full text] [doi: [10.1093/jamia/ocw115](https://doi.org/10.1093/jamia/ocw115)] [Medline: [27589942](https://pubmed.ncbi.nlm.nih.gov/27589942/)]
59. Kane EI, Gallo JJ. Perspectives of IRB chairs on the informed consent process. *AJOB Empir Bioeth* 2017;8(2):137-143 [FREE Full text] [doi: [10.1080/23294515.2016.1253628](https://doi.org/10.1080/23294515.2016.1253628)] [Medline: [28949842](https://pubmed.ncbi.nlm.nih.gov/28949842/)]
60. Haussen DC, Craft L, Doppelheuer S, Rodrigues GM, Al-Bayati AR, Ravindran K, et al. Legal authorized representative experience with smartphone-based electronic informed consent in an acute stroke trial. *J Neurointerv Surg* 2020 May;12(5):483-485. [doi: [10.1136/neurintsurg-2019-015283](https://doi.org/10.1136/neurintsurg-2019-015283)] [Medline: [31530653](https://pubmed.ncbi.nlm.nih.gov/31530653/)]
61. Simon CM, Klein DW, Schartz HA. Traditional and electronic informed consent for biobanking: a survey of U.S. biobanks. *Biopreserv Biobank* 2014 Dec;12(6):423-429. [doi: [10.1089/bio.2014.0045](https://doi.org/10.1089/bio.2014.0045)] [Medline: [25496155](https://pubmed.ncbi.nlm.nih.gov/25496155/)]
62. Sundby A, Boolsen MW, Burgdorf KS, Ullum H, Hansen TF, Middleton A, et al. The preferences of potential stakeholders in psychiatric genomic research regarding consent procedures and information delivery. *Eur Psychiatry* 2019 Jan;55:29-35. [doi: [10.1016/j.eurpsy.2018.09.005](https://doi.org/10.1016/j.eurpsy.2018.09.005)] [Medline: [30384109](https://pubmed.ncbi.nlm.nih.gov/30384109/)]
63. Bain KT, Holmes HM, Beers MH, Maio V, Handler SM, Pauker SG. Discontinuing medications: a novel approach for revising the prescribing stage of the medication-use process. *J Am Geriatr Soc* 2008 Oct;56(10):1946-1952 [FREE Full text] [doi: [10.1111/j.1532-5415.2008.01916.x](https://doi.org/10.1111/j.1532-5415.2008.01916.x)] [Medline: [18771457](https://pubmed.ncbi.nlm.nih.gov/18771457/)]
64. Gogtay NJ, Doshi BM, Kannan S, Thattai U. A study of warning letters issued to clinical investigators and institutional review boards by the United States Food and Drug Administration. *Indian J Med Ethics* 2011;8(4):211-214 [FREE Full text] [doi: [10.20529/IJME.2011.082](https://doi.org/10.20529/IJME.2011.082)] [Medline: [22106658](https://pubmed.ncbi.nlm.nih.gov/22106658/)]
65. Schenker Y, Wang F, Selig SJ, Ng R, Fernandez A. The impact of language barriers on documentation of informed consent at a hospital with on-site interpreter services. *J Gen Intern Med* 2007 Nov;22 Suppl 2:294-299 [FREE Full text] [doi: [10.1007/s11606-007-0359-1](https://doi.org/10.1007/s11606-007-0359-1)] [Medline: [17957414](https://pubmed.ncbi.nlm.nih.gov/17957414/)]

## Abbreviations

**CASP:** Critical Appraisal Skills Programme

**FDA:** US Food and Drug Administration

**MMAT:** Mixed Methods Appraisal Tool

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

*Edited by G Eysenbach; submitted 06.04.20; peer-reviewed by F Abujarad, C Nebecker, E Golembiewski; comments to author 15.06.20; revised version received 07.08.20; accepted 17.08.20; published 08.10.20.*

*Please cite as:*

De Sutter E, Zaçe D, Boccia S, Di Pietro ML, Geerts D, Borry P, Huys I

Implementation of Electronic Informed Consent in Biomedical Research and Stakeholders' Perspectives: Systematic Review  
*J Med Internet Res* 2020;22(10):e19129

URL: <https://www.jmir.org/2020/10/e19129>

doi: [10.2196/19129](https://doi.org/10.2196/19129)

PMID: [33030440](https://pubmed.ncbi.nlm.nih.gov/33030440/)

©Evelien De Sutter, Drieda Zaçe, Stefania Boccia, Maria Luisa Di Pietro, David Geerts, Pascal Borry, Isabelle Huys. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Review

# The Role of Technology and the Continuum of Care for Youth Suicidality: Systematic Review

Hannah Szlyk<sup>1\*</sup>, MSW, PhD; Jia Tan<sup>2\*</sup>, MSW

<sup>1</sup>School of Social Work, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

<sup>2</sup>The Brown School, Washington University in St Louis, Saint Louis, MO, United States

\* all authors contributed equally

**Corresponding Author:**

Hannah Szlyk, MSW, PhD

School of Social Work

Rutgers, The State University of New Jersey

390 George Street

New Brunswick, NJ, 08901

United States

Phone: 1 9178364225

Email: [hannah.szlyk@rutgers.edu](mailto:hannah.szlyk@rutgers.edu)

## Abstract

**Background:** Youth suicide is a global public health issue, and using technology is one strategy to increase participation in preventive interventions. However, there is minimal knowledge on how technology-enhanced interventions for youth correspond to the stages of care, from illness or risk recognition to treatment follow-up.

**Objective:** This systematic review aims to examine the efficacy of technology-enhanced youth suicide prevention and interventions across the continuum of care.

**Methods:** Four electronic databases were searched up to spring 2019 for youth suicide preventive interventions that used technology. The review was not restricted by study design and eligible studies could report outcomes on suicidality or related behaviors, such as formal treatment initiation. An adapted version of the Methodological Quality Ratings Scale was used to assess study quality.

**Results:** A total of 26 studies were identified. The findings support the emerging efficacy of technology-enhanced interventions, including a decline in suicidality and an increase in proactive behaviors. However, evidence suggests that there are gaps in the continuum of care and recent study samples do not represent the diverse identities of vulnerable youth.

**Conclusions:** The majority of identified studies were conducted in school settings and were universal interventions that aligned with the illness and risk recognition and help-seeking stages of the continuum of care. This field could be strengthened by having future studies target the stages of assessment and treatment initiation, include diverse youth demographics, and examine the varying roles of providers and technological components in emerging interventions.

(*J Med Internet Res* 2020;22(10):e18672) doi:[10.2196/18672](https://doi.org/10.2196/18672)

**KEYWORDS**

youth; suicide prevention; technology; continuum of care

## Introduction

**Background**

Youth suicide is a global public health crisis. In the United States, suicide is the second leading cause of death for children and youth aged 10-24 years [1]. Globally, suicide is the third leading cause of death for youths aged 15-19 years [2]. In addition to deaths by suicide, suicidality includes suicidal ideation and related behaviors, such as plans to attempt suicide

and actual suicide attempts [3]. Thus, research suggests that the risk of youth suicide may even be more pervasive as many youths experience suicidal ideation and nonfatal suicidal behaviors [4]. Universally, adolescence and early adulthood are vulnerable periods for when suicide risk is particularly elevated [5].

Therefore, it is important that youths have access to global systems of mental health care. Mental health services and resources may range from promoting illness recognition (in the

case of suicidality, this may include risk factor recognition) to providing targeted treatment and to offering follow-up services. This range, which spans the levels of intensity in care, is often called the *continuum of care* [6,7]. Suicidal individuals who are engaged in an integrated continuum of mental health care may experience decreases in suicidality [8]. However, youth engagement in the continuum of mental health care is often complicated as a consequence of developmental changes, the delayed detection of symptoms, and delayed access to treatment [9]. Thus, researchers need to ensure that available interventions are tailored to specifically meet youths' needs and correspond with the stages of the continuum of care [9].

Technology is one of the identified mediums to bridge gaps in the continuum of suicide preventive interventions [10]. Technology is especially relevant to engaging youths around the world. Research suggests that most young people in the United States and in developed countries have access to smartphones [11,12], whereas access increases among younger cohorts of emerging economies [12]. Therefore, the use of technology may address barriers to face-to-face care, such as access, reach, and stigma [10,13]. *Technology-enhanced interventions* use technology to solely deliver or serve as a component of an intervention and can include a mobile phone app, text messaging, telephone, videos, and web-based platforms [13]. Previous reviews on this topic have been restricted to gatekeeper interventions [14], including interventions across the lifespan or interventions designed to address broad mental health issues [15], focused on specific technologies [16], or that may be outdated as new interventions have since been developed [17]. No known review has explored how current technology-enhanced suicide interventions for youth correspond to the stages of the continuum of care. Thus, there is an incomplete understanding of the breadth and efficacy of preventive interventions that use technology and serve youth at risk of suicide.

## Objectives

To address these gaps in the literature, this systematic review aims to examine the efficacy of technology-enhanced youth suicide prevention and interventions across the continuum of care. The authors evaluated study outcomes in addition to suicidality, including help-seeking behaviors and coping skills, to better assess how the literature supports youth in leading lives worth living. The findings have implications for how suicidology may address identified gaps in the stages of the continuum of technology-enhanced suicide interventions and enhance care for vulnerable youths.

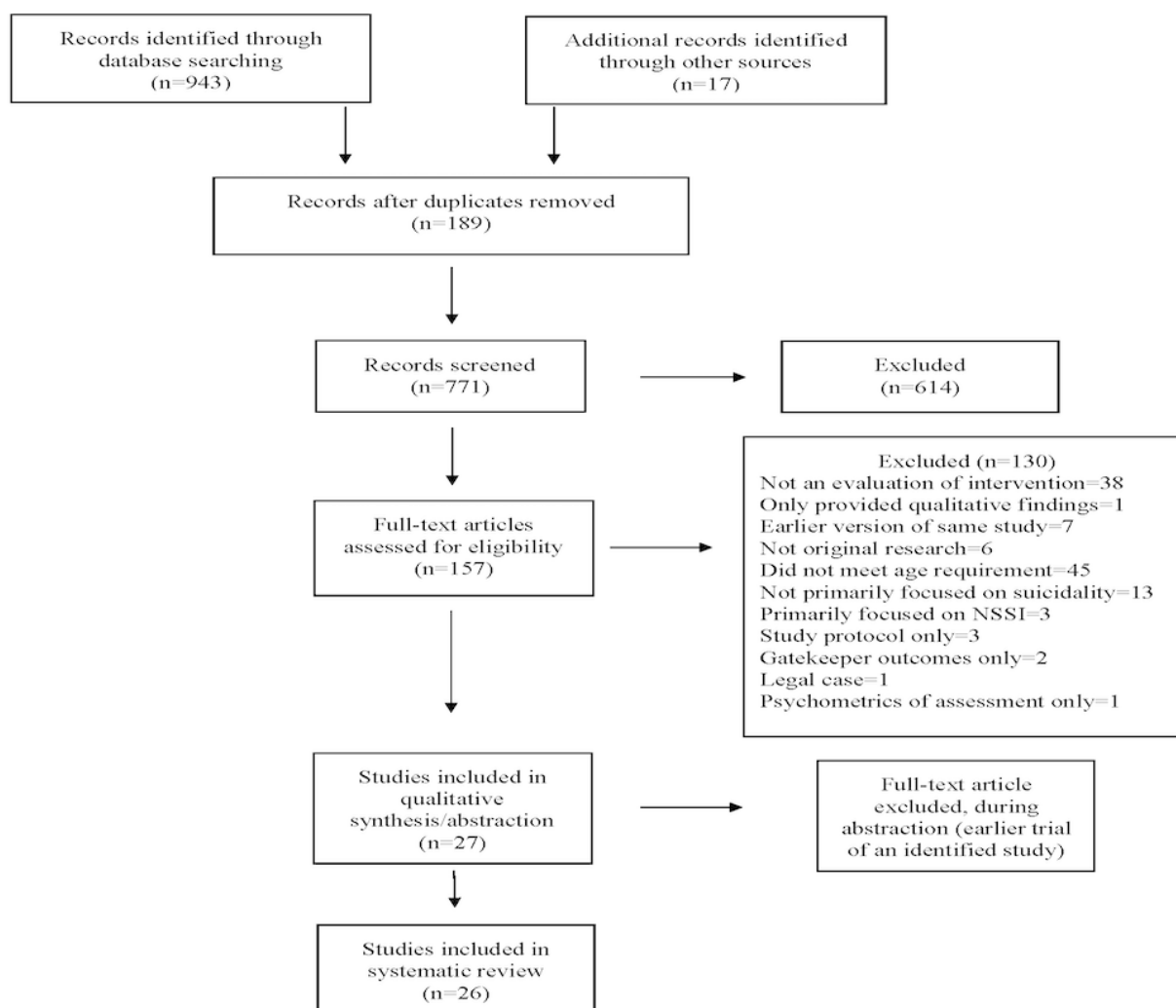
## Methods

### Search Strategy

The search was conducted in spring 2019, and the systematic review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Figure 1) [18].

Electronic databases were searched (PsycINFO, CINAHL [Cumulated Index to Nursing and Allied Health Literature], Ovid MEDLINE, ClinicalTrials.gov) using search terms specific to youth, suicidality, technology, and interventions (Multimedia Appendix 1). The study selection and data extraction were conducted by 2 assessors (the first and second authors). Eligible studies were required to focus on youths who were at potential risk, at high risk, or struggling with suicidality, be in English, and have the majority of participants between the ages of 12 and 24 years [19,20]. Selected studies adhered to the definition of technology-enhanced interventions (as previously defined) by Kreuze et al [13]. The investigators decided not to exclude studies that had primary outcome variables other than suicidality. This decision was based on a preliminary search of the literature, in which the investigators noticed that studies of technology-enhanced interventions for youth fell into several primary outcome domains (see the Data Extraction section for the outcome domains explored). The intention was to identify studies that may have been overlooked in previous reviews restricted to outcomes of suicidality and that applied to the stages of the continuum of care. Relevant systematic reviews were also cross-checked to adjust initial search terms and to potentially identify studies that had been missed in the final search.

Studies that focused on assisted suicide, nonsuicidal self-injury, postvention, or only gatekeeper outcomes were excluded from the final sample. In addition, the investigators excluded studies that only used technology to collect information about participant characteristics and behaviors, which only provided qualitative results, or only discussed the psychometrics of their assessment tool. No exclusion was placed on the trial design. When multiple publications of the same intervention were identified, the most recent or the most advanced trial was selected. Preliminary searches were organized using Endnote [21], and screening and data extraction were conducted using Covidence [22] and spreadsheets.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram for search strategy.

## Data Extraction

Data were extracted for study characteristics (Tables 1 and 2). As gender bias in peer review and in first author publications exists [23,24], the first author's gender identity was extracted and determined by searching department profiles. In addition, as suicidology is considered a multidisciplinary field, the investigators decided to document the first author's field of degree by searching web-based profiles and curriculum vitae. Studies' stages on the continuum of care were also extracted, and studies could cover more than 1 stage. The continuum of care framework used in this review was based on the literature from various health disciplines [25-27] and included the following stages: illness or suicide risk factor recognition, help seeking, assessment, treatment initiation, treatment module, treatment engagement, and follow-up.

Quality of study design and methodology were assessed using an adapted version of the Methodological Quality Rating Scale (MQRS; see Multimedia Appendix 1 for full scale) [28]. The MQRS was originally developed to review alcohol treatment outcomes and covered 12 domains, such as study design, documentation of quality control of treatments, and reports of participants lost to attrition [28,29]. Each domain is rated by

the study's strategies to decrease bias, with studies reporting more rigorous strategies receiving a higher quality score. The MQRS has been used to review other substance and mental health intervention studies [30,31] and has been adapted to evaluate the inclusion of theoretical frameworks and cultural linguistic adaptations in study development [32,33]. The investigators assessed interrater reliability by calculating the percent agreement for each MQRS domain. For example, there was an 81% agreement on both investigators' ratings by the domain of study design (see Multimedia Appendix 1 for the percent agreement for all MQRS domains).

The outcome attainment of each selected study was evaluated using a categorization system for statistical significance and outcome direction [33]. Outcomes were grouped into the following domains: suicidality (ie, suicidal ideation, planning suicide attempts, and attempts), co-occurring mental health issues (ie, anxiety, depression), youth's perceptions and knowledge about suicide, help-seeking behaviors, proactive coping behaviors, and formal treatment initiation, which is the official beginning of participation in outpatient or inpatient mental health treatment.

The investigators documented whether the outcomes related to these domains demonstrated statistically significant changes (at least  $P < .05$ ) and whether the change occurred in the desired direction (ie, a decrease in suicidal ideation) [33]. Differences in coding and ratings during the phases of study screening and data extraction were resolved through consensus between the 2 investigators.

## Results

### Study Characteristics

The final sample included 26 studies (Table 1) [34-59]. Although various professional disciplines were represented (ie, social work, medicine, public health), over half ( $n=16$ ) of the first authors held a doctoral degree in psychology. In total, 16 of the first authors were female; 13 studies were conducted in the United States. Among studies that provided participant ages,

the mean or median age ranged from 14.7 to 23 years. Female participants represented the majority of the sample participants. Of the studies that did report youth ethnicity, the majority of participants identified as White; no study reported the sexual or gendered minority identity of the participants. A total of 13 studies were conducted in an educational setting.

The selected studies included a variety of preventive interventions (Table 2); the SOS (Signs of Suicide) model was included in 2 studies: the original model for high schools [34] and the adapted model for middle schools [41]. The modalities of technology most often used were phones and web or web-based platforms. Half of the studies ( $n=13$ ) described indicated interventions for youth suicidality. A total of 12 studies were randomized controlled trials (RCTs). The outcomes of the adapted MQRS tool ranged from 0 (lowest quality) to 18 (highest quality), with an average score of 10.4 and range of 3 to 16.

**Table 1.** Study characteristics.

Authors (reference)	First author's field	Country	Setting	Sample size, n	Age (years), mean (SD) <sup>a</sup>	Majority gender of sample	Majority ethnicity of sample
<b>Universal</b>							
Aseltine et al [34]	Sociology	United States	High school	2100	Not available	Female	Hispanic, non-White
Bailey et al [35]	Behavioral health sciences	Australia	High school	129	16.7 (N/A) <sup>b</sup>	Male	White
Freedenthal [36]	Social work	United States	High school	146	15.8 (1.2)	Female	Hispanic, non-White
Haas et al [37]	Psychology	United States	University; web based	1162	Not available	Female	Not available
Han et al [38]	Public health	Australia; China	University; web based	257	Subsamples: • 18.6 (1.02) • 20.1 (2.08)	Female	Chinese
Pisani et al [39]	Psychology	United States	Rural high school	42	Not available	Female	White
Robinson et al [40]	Psychology	Australia	High school	69	16.4 (N/A)	Not available	Not available
Schilling et al [41]	Public health	United States	Middle school	386	Not available	Female	White
Wyman et al [42]	Psychology	United States	High school	2675	Not available	Female	White
<b>Selective</b>							
Dickter et al [43]	Psychology	United States	Hospital; web based	83	17.5 (2.04)	Female	White
Hetrick et al [44]	Psychology	Australia	Community mental health clinic	101	18.7 (2.8)	Female	Not available
Iorfino et al [45]	Psychology	Australia	Primary mental health clinic; web based	232	20.4 (2.59)	Female	Not available
King et al [46]	Psychology	United States	University; web based	76	22.9 (5.0)	Female	White
<b>Indicated</b>							
Bertolote et al [47]	Clinical sciences; psychiatry	Brazil, India, Sri Lanka, Iran, China	Emergency department	1867	23 <sup>c</sup> median age	Female	Indian
Czyz et al [48]	Psychology	United States	Hospital	36	15.4 (1.36)	Female	White
Hetrick et al [49]	Psychology	Australia	High school; web based	50	14.7 (1.4)	Female	Not available
King et al [50]	Psychology	United States	Hospital	448	15.6 (1.31)	Female	White
King et al [51]	Psychology	Australia	Hotline or counseling center	101	Not available	Female	Not available
Mehlum et al [52]	Medicine or psychology	Norway	Child and adolescent psychiatric outpatient	77	15.6 (1.5)	Female	Norwegian
Normand et al [53]	Psychiatry	France	Hospital	173	Subsamples: • 17.9 (1.9) • 18.4 (1.8)	Female	Not available



Authors (reference)	First author's field	Country	Setting	Sample size, n	Age (years), mean (SD) <sup>a</sup>	Majority gender of sample	Majority ethnicity of sample
O'Brien et al [54]	Social work	United States	Psychiatric out-patient	20	15.7 (1.6)	Female	White
Rosenbaum et al [55]	Psychiatry	United States	Emergency department	181	14.7 (2.0)	Female	Hispanic, non-White
Tan et al [56]	Psychology	China	Web based	725	21.2 (3.69)	Female	Not available
Yen et al [57]	Psychology	United States	Psychiatry inpatient unit	20	15.9 (1.5)	Female	White
Yen et al [58]	Psychology	United States	Psychiatry inpatient unit	50	15.7 (1.53)	Female	White
<b>All tiers</b>							
Silverstone et al [59]	Psychiatry	Canada	Middle and high school	6651	Not available	Female	Not available

<sup>a</sup>For mean age, full sample information has been provided; subsample information was reported when the mean and SD of the full sample could not be generated due to insufficient information.

<sup>b</sup>N/A: not applicable.

<sup>c</sup>Authors only provided median age across study sites.

**Table 2.** Study characteristics continued.

Authors (references)	Design	Intervention name	Description	Technology used	MQRS <sup>a</sup> score (0-18)
<b>Universal</b>					
Aseltine et al [34]	RCT <sup>b</sup>	SOS <sup>c</sup>	Aims to raise awareness of suicide risk and promote help seeking for youth and peers. Curricula delivered via video and discussion. Participation in program over 2 days	Video	14
Bailey et al [35]	Pre, posttest	safeTalk	A 1-time, 3-hour workshop delivered to students by a trainer and school staff. Uses presentations, video, discussion, questions, and role plays. Teaches students about suicide risk, perceptions about suicide, and help-seeking strategies	Video, phone	10
Freedenthal [36]	Quasi-experimental trial	Yellow Ribbon	A 60-min student leadership training for students and school staff conducted by program trainers. Training is focused on warnings signs of suicide among youth, myths about suicide, and the importance of seeking help for peers or for oneself. Content is delivered via a slide show and Ask4Help cards	Digital slide presentation	11
Haas et al [37]	Prospective cohort study	College screening	An interactive, web-based program for university students officially called the College Screening Project. The web-based screening identified at-risk students, supported them in getting help, and helped to determine the proportion of students who entered treatment	Web-based screen-er, email, web-based chat	9
Han et al [38]	RCT	ProHelp	A brief, 2-module web-based psychoeducational program that aims to teach students about risk factors for suicide, stigmatizing attitudes, and barriers to help seeking	Web-based platform	12
Pisani et al [39]	Field test	Text4Strength	An automated, interactive text messaging intervention developed for early adolescents in rural communities. It is an extension of the Sources of Strengths program. Youth received messages over 9 weeks that were related to topics of emotion regulation, social connections, and help seeking	Video, text messaging	11
Robinson et al [40]	Pilot study	Social Media Message	A social media message intervention that was developed by youth for at-risk peers. Designed to increase youth awareness about suicide, risk factors, and strategies to help peers and themselves. Participants evaluated 2 social media messages	Video, phone, tablets and computers, web based	3
Schilling et al [41]	RCT	SOS-Middle School	Similar to the SOS high school version. Features a 17-min DVD that includes 3 age-appropriate vignettes; a group discussion by middle school students about depression, suicide, bullying, self-injury, and getting help; and a student interview with a school-based counselor to model getting help. Delivered by trained school personnel	DVD or video	13

Authors (references)	Design	Intervention name	Description	Technology used	MQRS <sup>a</sup> score (0-18)
Wyman et al [42]	RCT	Sources of Strength	Aims to improve youth help-seeking behaviors and proactive coping to reduce the risk of suicide. The program has 3 standard phases: (1) school and community preparation, (2) peer leader training, and (3) schoolwide messaging through video and text messaging. Premise is that peer and staff training (varying from 1 to 6 hours) in curriculum encourages sustainability of the program	Video, text messaging	11
<b>Selective</b>					
Dickter et al [43]	Nonrandomized trial, 2 treatment groups	CATCH-IT: The Competent Adulthood Transition with Cognitive Behavioral and Interpersonal Training	The Competent Adulthood Transition with Cognitive Behavioral and Interpersonal Training consists of 14 self-guided, web-based modules that use techniques from CBT <sup>d</sup> and interpersonal psychotherapy. Aims to teach skills for increasing resiliency against depressive disorders and suicidality	Web-based platform	9
Hetrick et al [44]	Prospective cohort study	Monitoring Tool	A web-based tool for self-monitoring of depression and suicidal ideation that tracked changes in symptoms and alerted clinicians. Participating youth completed the tool between 2 and 8 times (duration varying between 4 and 83 days)	Web-based platform	10
Iorfino et al [45]	Prospective cohort study	Synergy Online System	An initial clinical assessment on the web before a face-to-face or web-based clinical appointment. The initial clinical assessment assesses a range of mental health outcomes (14 modules, approximately 45 min to complete). At the end of the suicidality module, the algorithms assess current and past suicidality, which alerts clinical staff if the current suicide risk is high	Video, web-based platform	6
King et al [46]	RCT	EBridge	The Electronic Bridge to Mental Health Services is a web-based screening and intervention program for college students at risk for suicide. The program provides students with feedback from the screening and information about resources and can link students with web-based counseling services. The program aims to increase help-seeking and eventual use of mental health services. Length of program depended on level of student interaction	Web-based platform	11
<b>Indicated</b>					
Bertolote et al [47]	RCT	BIC	A brief educational intervention with periodic follow-up contacts for suicide attempters conducted at global emergency departments and was part of the WHO <sup>e</sup> Multisite Intervention Study on Suicidal Behaviors (SUPRE-MISS). The BIC procedure includes a standard 1-hour individual information session at the time of discharge. Follow-up contacts by health professionals were 1 week; 2, 4, 7, and 11 weeks; and 4, 6, 12, and 18 months after discharge	Phone	17

Authors (references)	Design	Intervention name	Description	Technology used	MQRS <sup>a</sup> score (0-18)
Czyz et al [48]	RCT	MI-SafeCope	A motivational interview-enhanced safety planning intervention for teens hospitalized because of suicide risk. The intervention includes 3 components: an individual session, a family session, and a 2-week postdischarge booster call by phone (with the intervention counselor). Youth also provided assessments via text message at 1 and 3 months after discharge	Phone, text messaging	14
Hetrick et al [49]	RCT	Reframe-IT	An internet-based CBT program that aims to reduce suicide-related behaviors, depression, anxiety, and hopelessness and improve problem solving and cognitive and behavioral issues. The intervention consisted of 8 modules of CBT delivered on the web over 10 weeks	Web-based platform	12
King et al [50]	RCT	YST-II	The Youth-Nominated Support Team–Version II for suicidal adolescents provides psychoeducation and ongoing consultation for the parent-approved adult support persons that have been nominated by the adolescent. The support persons maintain regular supportive contact with the adolescents via phone for 3 months following hospitalization	Phone	14
King et al [51]	Pre-post tests	Kids Help Line	Trained help line counselors assessed changes in suicidality and mood for youth callers at the beginning of the session and at the conclusion of the phone session. Mean duration of calls was 40 min (range 10-120 min)	Phone	4
Mehlum et al [52]	RCT	DBT-A: Dialectical Behavior Therapy-Adolescent	Dialectical behavior therapy for adolescents lasts from 3 to 5 months, includes parents or other caregivers in weekly skills training groups, and has a skills module to support teens with emotion dysregulation and their families. This trial was delivered over 18 weeks and was delivered by trained mental health professionals. Coaching sessions were delivered over the phone	Phone	16
Normand et al [53]	Prospective cohort study	4-Phone-Calls	Hospital staff called the youth 1 week, 1 month, 6 months, and 12 months after discharge for a suicide attempt. The interviews during the phone calls included informal and formal assessment of current symptoms and the youth's safety	Phone	7
O'Brien et al [54]	Pilot study	Crisis Care	A smartphone app intervention developed specifically for suicidal adolescents and their parents to use after discharge from the hospital. The app provides access to coping skills and immediate access to help, if needed	Smartphone app	4
Rosenbaum et al [55]	RCT	FISP	The Family Intervention for Suicide Prevention is developed for youth admitted to emergency departments post-suicide attempt. Youth and their families participate in a CBT session aimed to increase motivation for follow-up treatment and safety postdischarge. Participants also received structured phone calls 48 hours and often 1, 2, and 4 weeks postdischarge to promote outpatient treatment attendance	Phone	11

Authors (references)	Design	Intervention name	Description	Technology used	MQRS <sup>a</sup> score (0-18)
Tan et al [56]	Pilot study	Microblog Intervention	An intervention developed for users of the Sina Weibo microblogging platform. Participants received direct messages designed to respond to high suicide risk postings. The intervention aimed to increase help seeking for at-risk users and peers	Web-based platform; Sina Weibo	5
Yen et al [57]	Pre-post test	STEP	Includes an in-person phase (4 sessions) and a remote delivery phase (text messaging and phone calls). The inpatient sessions focus on psychoeducation and coping skills. The remote delivery phase consists of weekly phone calls and daily text messages to provide skills practice reminders and to monitor mood	Phone, text messaging	12
Yen et al [58]	RCT	CLASP-A	The Coping Long Term with Active Suicide Program for Adolescents program is adapted for adolescents hospitalized for suicidal ideation or a suicide attempt. The program includes 3 individual sessions and 1 family session and a series of outpatient phone calls to adolescent and a designated parent or guardian over 6 months of follow-up post-discharge	Phone	15
<b>All tiers</b>					
Silverstone et al [59]	Pre-post and follow-up	EMPATHY	The multimodal program includes repeated data collection, identification of a high-risk group, a rapid intervention for the high-risk group (a supervised web-based CBT program), a universal CBT intervention, interactions with trained staff, and referrals to external medical and psychiatric services	Web-based platform	9

<sup>a</sup>MQRS: Methodological Quality Rating Scale.

<sup>b</sup>RCT: randomized controlled trial.

<sup>c</sup>SOS: Signs of Suicide.

<sup>d</sup>CBT: cognitive behavioral therapy.

<sup>e</sup>WHO: World Health Organization.

Regarding the continuum of care (Table 3), most studies were targeted to increase illness or risk factor recognition (n=11), to increase help seeking (n=10), and to guide youths through a treatment module (n=10; Table 2). The majority of studies (n=18) addressed >1 stage of the continuum of care. For example, 8 of the 9 universal interventions addressed illness or risk recognition and help seeking; 5 of the 11 indicated

interventions focused on the stages (at least) of participating in the treatment module and treatment engagement. As illustrated in Table 4, among the 6 common outcome domains measured, most studies (n=22) reported suicidality (ie, ideation, attempts) as an important study variable, followed by co-occurring mental health issues (n=12; ie, depression, distress, or anxiety).



**Table 3.** Studies and their stages on the continuum of care.

Authors (references)	Illness or risk recognition	Help seeking	Assessment	Treatment initiation	Treatment module	Treatment engagement	Follow-up
<b>Universal</b>							
Aseltine et al [34]	✓ <sup>a</sup>	✓	N/A <sup>b</sup>	N/A	N/A	N/A	N/A
Bailey et al [35]	✓	✓	N/A	N/A	N/A	N/A	N/A
Freedenthal [36]	✓	✓	N/A	N/A	N/A	N/A	N/A
Haas et al [37]	N/A	N/A	✓	✓	N/A	N/A	N/A
Han et al [38]	✓	✓	N/A	N/A	N/A	N/A	N/A
Pisani et al [39]	✓	✓	N/A	N/A	N/A	N/A	N/A
Robinson et al [40]	✓	✓	N/A	N/A	N/A	N/A	N/A
Schilling et al [41]	✓	✓	N/A	N/A	N/A	N/A	N/A
Wyman et al [42]	✓	✓	N/A	N/A	N/A	N/A	N/A
<b>Selective</b>							
Dickter et al [43]	N/A	N/A	N/A	N/A	✓	N/A	N/A
Hetrick et al [44]	N/A	N/A	✓	N/A	N/A	N/A	N/A
Iorfino et al [45]	N/A	N/A	✓	N/A	N/A	N/A	N/A
King et al [46]	✓	✓	N/A	✓	✓	✓	N/A
<b>Indicated</b>							
Bertolote et al [47]	N/A	N/A	N/A	N/A	✓	N/A	✓
Czyz et al [48]	N/A	N/A	N/A	N/A	✓	✓	✓
Hetrick et al [49]	N/A	N/A	N/A	N/A	✓	✓	N/A
King et al [50]	N/A	N/A	N/A	N/A	N/A	N/A	✓
King et al [51]	✓	N/A	N/A	N/A	N/A	N/A	N/A
Mehlum et al [52]	N/A	N/A	N/A	N/A	✓	✓	N/A
Normand et al [53]	N/A	N/A	N/A	N/A	N/A	N/A	✓
O'Brien et al [54]	N/A	N/A	N/A	N/A	✓	N/A	N/A
Rosenbaum et al [55]	N/A	N/A	N/A	N/A	N/A	✓	✓
Tan et al [56]	N/A	✓	N/A	N/A	N/A	N/A	N/A
Yen et al [57]	N/A	N/A	N/A	N/A	✓	✓	✓
Yen et al [58]	N/A	N/A	N/A	N/A	✓	✓	✓
<b>All tiers</b>							
Silverstone et al [59]	✓	N/A	✓	✓	✓	N/A	N/A
Total, n	11	10	4	3	10	7	7

<sup>a</sup>The study addresses that stage of the continuum.<sup>b</sup>N/A: not applicable.

**Table 4.** Measured and significant intervention outcomes.

Authors (references)	Suicidality	Co-occurring mental health issues	Perceptions and knowledge of suicide	Help seeking	Coping behavior	Treatment initiation	Significant outcomes
<b>Universal</b>							
Aseltine et al <sup>a</sup> [34]	✓ <sup>b</sup>	N/A <sup>c</sup>	✓	✓	N/A	N/A	Suicide attempts, decrease; adaptive attitudes about suicide, increase; knowledge about suicide, increase
Bailey et al [35]	✓	✓	✓	✓	N/A	N/A	Distress, decrease at T1 <sup>d</sup> and T2 <sup>d</sup> ; adaptive attitudes about suicide, increase; knowledge about suicide increase; help seeking for self at T2 and T3 <sup>d</sup> , increase
Freedenthal [36]	✓	N/A	N/A	✓	N/A	N/A	Help seeking by hotline, increase; help seeking from adult, decrease; help seeking from a peer, decrease
Haas et al [37]	N/A	✓	N/A	N/A	N/A	✓	Outpatient treatment initiation for students who received evaluation and dialogue with counselor, increase
Han et al <sup>a</sup> [38]	N/A	N/A	✓	✓	✓	N/A	Knowledge of suicide, increase at posttest; attitude about help seeking from a professional, increase
Pisani et al <sup>a,e</sup> [39]	N/A	✓	✓	N/A	✓	N/A	Not reported
Robinson et al [40]	✓	N/A	✓	✓	N/A	N/A	Not reported
Schilling et al <sup>a</sup> [41]	✓	N/A	✓	✓	N/A	N/A	Suicidal ideation, planning, or attempts decrease among intervention participants with pretest suicidal ideation; knowledge about suicide, increase
Wyman et al <sup>a</sup> [42]	✓	N/A	✓	✓	✓	N/A	Perceptions of seeking help from adults, increase; help seeking from nonmental health professional, increase; help seeking from a peer, increase
<b>Selective</b>							
Dickter et al [43]	✓	✓	N/A	N/A	N/A	N/A	Suicidal ideation, decrease
Hetrick et al [44]	✓	✓	N/A	N/A	N/A	N/A	Suicidal ideation, decrease; depression symptoms, significant decrease
Iorfino et al [45]	✓	N/A	N/A	N/A	N/A	N/A	Not reported

Authors (references)	Suicidality	Co-occurring mental health issues	Perceptions and knowledge of suicide	Help seeking	Coping behavior	Treatment initiation	Significant outcomes
King et al <sup>a</sup> [46]	✓	✓	✓	✓	N/A	✓	Stigma to seek help for mental health issues, decrease; help seeking from a mental health professional, increase; help seeking from family members, increase; help seeking from a peer, increase; outpatient treatment initiation, increase
<b>Indicated</b>							
Bertolote et al <sup>a</sup> [47]	✓	N/A	N/A	N/A	N/A	✓	Not reported
Czyz et al <sup>a</sup> [48]	✓	N/A	N/A	N/A	✓	N/A	Coping for suicidal behavior, increase; coping with safety plan, increase
Hetrick et al <sup>a</sup> [49]	✓	✓	N/A	N/A	✓	N/A	Not reported
King et al <sup>a</sup> [50]	✓	✓	N/A	N/A	N/A	N/A	Suicidal ideation, decrease at 6 weeks and 6 months of follow-up among multiple attempters
King et al [51]	✓	✓	N/A	N/A	N/A	N/A	Suicidal ideation, decrease; distress, decrease
Mehlum et al <sup>a</sup> [52]	✓	✓	N/A	N/A	N/A	✓	Suicidal ideation, decrease; depression symptoms, decrease
Normand et al [53]	✓	N/A	N/A	N/A	N/A	✓	Not reported
O'Brien et al [54]	N/A	N/A	✓	N/A	N/A	N/A	Not reported
Rosenbaum et al <sup>a</sup> [55]	✓	✓	N/A	N/A	N/A	✓	Outpatient treatment initiation, increase
Tan et al <sup>a,e</sup> [56]	✓	N/A	N/A	N/A	N/A	N/A	Not reported
Yen et al [57]	✓	N/A	N/A	N/A	N/A	✓	Suicidal ideation, decrease
Yen et al [58]	✓	N/A	N/A	N/A	N/A	✓	Not reported
<b>All tiers</b>							
Silverstone et al [59]	✓	✓	N/A	N/A	N/A	N/A	Suicidal ideation, planning, and attempts, decrease among actively suicidal participants; depression symptoms, decrease; anxiety, decrease

<sup>a</sup>Indicates the study was a randomized controlled trial.

<sup>b</sup>The study reports that outcome.

<sup>c</sup>N/A: not applicable.

<sup>d</sup>T1, T2, and T3 indicate the time trials of the larger study.

<sup>e</sup>The study's purpose was to promote help seeking, but no variables directly reported the help-seeking behaviors of the participants.

In total, 9 of the 12 RCT studies and 8 studies of other designs reported significant changes in study outcomes. A total of 22 studies measured suicidality outcomes and only 9 of those studies reported significant changes (4 of which were RCTs).

The next most common outcome domain measured was co-occurring mental health issues, and 5 of the 12 studies reported significant outcomes. In total, 6 of the 9 studies that measured perceptions and knowledge of suicide and 6 of the 8

studies that measured help-seeking behaviors noted significant findings. Finally, only 1 of the 5 studies that measured coping behaviors and 3 of the 8 studies that measured treatment initiation reported significant results.

## Discussion

### Principal Findings

The findings suggest that suicidologists around the world are working to utilize technology to prevent youth suicide. Results demonstrate that 17 interventions of varying study designs reported significant changes in at least one of the outcome domains. In addition, it is promising that the majority of selected studies were conducted in educational settings, which may increase opportunities for youth to learn about suicide risk, seek help, and participate in treatment beyond a formal, clinical setting. On the basis of the findings presented in [Tables 1-3](#), these school-based and university-based interventions were mostly universal interventions and aligned with the illness and risk recognition and help-seeking stages of the continuum of care. In addition, these interventions mainly used videos and web-based platforms and predominantly demonstrated efficacy in increasing help-seeking behaviors and youth knowledge and perceptions about suicide.

However, the results from this review also illustrate that efforts are needed to test technology-enhanced interventions across the continuum of care. Among this review's sample, few studies used technology to assess suicidality or to formally initiate mental health treatment. This gap in the continuum of care is crucial to address as these stages impact participation in treatment modules and, hopefully, prevent future deaths. During the initial search, the investigators noticed that many available electronic assessments were not specific to youth suicidality, had not been incorporated into an intervention trial, or were only in the early stages of development.

Regarding the other intervention tiers, only 4 studies, all using web-based platforms, were determined to be selective interventions and these studies spanned the stages of the continuum of care. Although 3 studies did note significant changes in multiple outcome domains, future research should focus on strategies to use technology to reach youths who are at higher risk of suicide across the continuum of care, as this appears to be an overlooked group in technology-enhanced interventions. It was also not surprising that the majority of indicated interventions addressed the treatment module, treatment engagement, and follow-up stages of care; these interventions also primarily used phones. Although several studies have reported significant improvements in youth suicidal behavior, future trials may extend treatment outcomes beyond suicidality to include coping behaviors (in this review, only 2 studies measured this) and other metrics that mark improvements in youth resiliency.

Although certain tiers of intervention and stages of the continuum of care were associated with specific types of technology, it may be premature to determine whether one modality is better associated with efficacy, acceptability, or feasibility. For example, it would be presumed that studies that

use phones would achieve more successful metrics than a study that used a less-established or newer technology, such as a mobile phone app or web-based platform. However, this review suggests that studies with significant findings do not use one specific type of technology and therefore other factors, such as study design, intervention curriculum, and youth sample, may have a greater impact on an intervention's success.

### Implications for Future Study Design

This review emphasizes that not all youth interventions that use technology are the same. Some interventions have more *human* and *face-to-face* involvement, whereas others are mainly automated or self-directed by the participant. *Behavioral intervention technology* (BIT) is a term used in other health disciplines to determine the level of human involvement and automation in an intervention [60]. In contrast, the suicidology literature most often evaluates the intervention as a whole and does not provide specific details about the technological components. Understanding the level of provider integration and the technological components of the BIT may also inform how interventions can be implemented in other settings and scaled to reach a broader youth consumer base [61]. For example, the review's findings confirm that universal interventions that target help seeking, illnesses, and risk recognition are already tested on a larger scale versus selective and indicated interventions. Therefore, incorporating the BIT terminology and models may help suicidologists determine the efficacy and acceptability of specific intervention components, determine which technologies are better suited for specific stages of the continuum of care, and to disseminate other tiers of interventions to mental health systems that service at-risk and suicidal youth.

Scores on the MQRS demonstrated variability in the quality of the studies and may be a consequence of not restricting the study design for this systematic review. For example, several non-RCT studies did not have multiple sessions or have multisite trials for their intervention. However, many studies, regardless of study design, have underdeveloped MQRS domains in common, such as documenting the study's theoretical foundation, conducting the study at more than 1 site, and reporting inclusion of a collateral data source. The limited inclusion of collateral data sources is problematic as it is considered best practice to supplement youth self-report with parent, guardian, or teacher observations in face-to-face interventions [62-65]. The issue of an incomplete client profile is most likely indicative of the challenges of data collection and digital interventions [66] and may be an issue that suicidologists implementing technology-enhanced interventions can collectively explore and tackle.

The scores on the MQRS also suggested that few studies adapted interventions to the cultural, social, and linguistic needs of their specific demographic. In addition, no study has reported the participants' sexual identity, although suicide risk is particularly heightened among sexual minority youth [67]. As global research demonstrates that youth suicidality varies across age cohorts and demographics [5], technology-enhanced interventions need to mirror and be tailored for this diversity.

## Implications for the Profession of Suicidology

Diversity among suicide scholars and professional perspectives may also influence the impact that technology-enhanced interventions have on youths' mental health [68-70]. It is promising that the first authors of 16 studies identified as female, considering the noted gender bias in peer review and grant funding [23,24,71]. This bias has been noted in suicidology as well, as the American Association of Suicidology has historically bestowed more men than women with its annual research awards [72]. In addition, doctoral training in psychology was most common among the first authors. Although this finding is not reflective of the potential professional diversity of the research team and of nonintervention studies, the subfield of technology-enhanced youth interventions may be mindful of how to mentor students and researchers from other disciplines to be principal investigators. Disciplines may include those who are involved in the frontline (such as nursing) or those who have extensive training in digital health literacy and computational methods (ie, health communication fields).

## Limitations

As the study sample was not restricted by research design, the investigators could not compare outcome effect sizes. Preliminary searches demonstrated that the pool of eligible studies that were also RCTs would be small and that the investigators did not want to overlook cutting-edge interventions that were in earlier stages of development. In addition, as this

review was restricted to specific search guidelines and because suicide research is ever evolving, relevant studies may not have been included. For example, many studies were ineligible as they did not collect or report participants' ages or validate that the participants were within the specific age range.

## Conclusions

This systematic review emphasizes the need for technology-enhanced interventions that extend beyond illness or risk recognition and help seeking, which are developed for diverse youth populations. Although technology shows promise in its utility to address suicidality and increase proactive behaviors, such as help seeking and coping skills, it is difficult to determine which types of technology are better associated with intervention efficacy, acceptability, and feasibility and better suited for specific stages of the continuum of care. The field of suicidology also faces challenges in capturing youth participants' demographics on digital platforms, supplementing youth self-reports with collateral information, developing interventions suitable for underserved demographics, and involving researchers from diverse backgrounds and disciplines. Adoption of BIT terminology and frameworks may improve the understanding of both the roles of providers and technological components in technology-enhanced suicide preventive interventions for youth and how these interventions can be successfully implemented across the continuum of care and within mental health care systems.

## Acknowledgments

Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number T32MH019960. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Literature search terms and adapted study quality scoring sheet.

[DOCX File, 26 KB - [jmir\\_v22i10e18672\\_app1.docx](#)]

## References

- 10 leading causes of death by age group, United States, 2017. Center for Disease Control and Prevention (CDC). 2018. URL: <https://www.cdc.gov/injury/wisqars/LeadingCauses.html> [accessed 2019-12-12]
- Suicide. World Health Organization (WHO). 2019. URL: <https://www.who.int/news-room/fact-sheets/detail/suicide> [accessed 2019-12-12]
- Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. *Am J Psychiatry* 2007 Jul;164(7):1035-1043 [FREE Full text] [doi: [10.1176/ajp.2007.164.7.1035](https://doi.org/10.1176/ajp.2007.164.7.1035)] [Medline: [17606655](#)]
- Nock MK, Green JG, Hwang I, McLaughlin KA, Sampson NA, Zaslavsky AM, et al. Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement. *JAMA Psychiatry* 2013 Mar;70(3):300-310 [FREE Full text] [doi: [10.1001/2013.jamapsychiatry.55](https://doi.org/10.1001/2013.jamapsychiatry.55)] [Medline: [23303463](#)]
- McLoughlin AB, Gould MS, Malone KM. Global trends in teenage suicide: 2003-2014. *QJM* 2015 Oct 31;108(10):765-780. [doi: [10.1093/qjmed/hcv026](https://doi.org/10.1093/qjmed/hcv026)] [Medline: [25638789](#)]
- Behar L. Changing Patterns of State Responsibility: A Case Study of North Carolina. *Journal of Clinical Child Psychology* 2010 Jun 07;14(3):188-195. [doi: [10.1207/s15374424jccp1403\\_4](https://doi.org/10.1207/s15374424jccp1403_4)]



7. Stroul B, Friedman R. Washington, DC: Georgetown University Child Development Center, CASSP Technical Assistance Center; 1986.
8. Gould MS, Lake AM, Galfalvy H, Kleinman M, Munfakh JL, Wright J, et al. Follow-up with Callers to the National Suicide Prevention Lifeline: Evaluation of Callers' Perceptions of Care. *Suicide Life Threat Behav* 2018 Feb;48(1):75-86. [doi: [10.1111/sltb.12339](https://doi.org/10.1111/sltb.12339)] [Medline: [28261860](https://pubmed.ncbi.nlm.nih.gov/28261860/)]
9. MacDonald K, Fainman-Adelman N, Anderson KK, Iyer SN. Pathways to mental health services for young people: a systematic review. *Soc Psychiatry Psychiatr Epidemiol* 2018 Oct;53(10):1005-1038 [FREE Full text] [doi: [10.1007/s00127-018-1578-y](https://doi.org/10.1007/s00127-018-1578-y)] [Medline: [30136192](https://pubmed.ncbi.nlm.nih.gov/30136192/)]
10. Berrouiguet S, Courtet P, Larsen M, Walter M, Vaiva G. Suicide prevention: Towards integrative, innovative and individualized brief contact interventions. *Eur Psychiatry* 2018 Jan;47:25-26. [doi: [10.1016/j.eurpsy.2017.09.006](https://doi.org/10.1016/j.eurpsy.2017.09.006)] [Medline: [29096129](https://pubmed.ncbi.nlm.nih.gov/29096129/)]
11. Anderson M, Jiang J. 2018. URL: <https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/> [accessed 2019-12-12]
12. Smartphone ownership is growing rapidly around the world, but not always equally. Taylor K, Silver L. 2019. URL: <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/> [accessed 2019-12-12]
13. Kreuze E, Jenkins C, Gregoski M, York J, Mueller M, Lamis DA, et al. Technology-enhanced suicide prevention interventions: A systematic review. *J Telemed Telecare* 2017 Jul;23(6):605-617. [doi: [10.1177/1357633X16657928](https://doi.org/10.1177/1357633X16657928)] [Medline: [27377792](https://pubmed.ncbi.nlm.nih.gov/27377792/)]
14. Kreuze E, Ruggiero KJ. Technology-Oriented Suicide Prevention Interventions for Adolescents and Adolescent Gatekeepers: A Qualitative Review. *Adolescent Res Rev* 2017 Jun 6;3(2):219-233. [doi: [10.1007/s40894-017-0060-5](https://doi.org/10.1007/s40894-017-0060-5)]
15. Christensen H, Batterham P, O'Dea B. E-health interventions for suicide prevention. *Int J Environ Res Public Health* 2014 Aug 12;11(8):8193-8212 [FREE Full text] [doi: [10.3390/ijerph110808193](https://doi.org/10.3390/ijerph110808193)] [Medline: [25119698](https://pubmed.ncbi.nlm.nih.gov/25119698/)]
16. Reyes-Portillo JA, Mufson L, Greenhill LL, Gould MS, Fisher PW, Tarlow N, et al. Web-based interventions for youth internalizing problems: a systematic review. *J Am Acad Child Adolesc Psychiatry* 2014 Dec;53(12):1254-1270.e5. [doi: [10.1016/j.jaac.2014.09.005](https://doi.org/10.1016/j.jaac.2014.09.005)] [Medline: [25457924](https://pubmed.ncbi.nlm.nih.gov/25457924/)]
17. Perry Y, Werner-Seidler A, Cascar AL, Christensen H. Web-Based and Mobile Suicide Prevention Interventions for Young People: A Systematic Review. *J Can Acad Child Adolesc Psychiatry* 2016;25(2):73-79 [FREE Full text] [Medline: [27274742](https://pubmed.ncbi.nlm.nih.gov/27274742/)]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
19. Young teens (ages 12-14 years of age). Center for Disease Control and Prevention(CDC). 2019. URL: <https://www.cdc.gov/ncbddd/childdevelopment/positiveparenting/adolescence.html> [accessed 2019-12-12]
20. World Health Organization (WHO). Adolescent health in the South-East Asia region. 2020. URL: <https://www.who.int/southeastasia/health-topics/adolescent-health> [accessed 2019-12-12]
21. Endnote. Citation and reference software. Philadelphia, PA: Clarivate Analytics; 2020.
22. Covidence. Systematic review software. Melbourne: Veritas Health Innovation; 2020.
23. Gender in the Global Research Landscape. Elsevier. 2017. URL: <https://www.elsevier.com/research-intelligence/campaigns/gender-17> [accessed 2019-12-12]
24. Ross E. Gender bias distorts peer review across fields. *Nature News*. 2017 Mar 23. URL: <https://www.nature.com/news/gender-bias-distorts-peer-review-across-fields-1.21685> [accessed 2020-05-01]
25. Gabert R, Ng M, Sogarwal R, Bryant M, Deepu RV, McNellan CR, et al. Identifying gaps in the continuum of care for hypertension and diabetes in two Indian communities. *BMC Health Serv Res* 2017 Dec 27;17(1):846 [FREE Full text] [doi: [10.1186/s12913-017-2796-9](https://doi.org/10.1186/s12913-017-2796-9)] [Medline: [29282052](https://pubmed.ncbi.nlm.nih.gov/29282052/)]
26. Graboyes EM, Kompelli AR, Neskey DM, Brennan E, Nguyen S, Sterba KR, et al. Association of Treatment Delays With Survival for Patients With Head and Neck Cancer: A Systematic Review. *JAMA Otolaryngol Head Neck Surg* 2019 Feb 01;145(2):166-177 [FREE Full text] [doi: [10.1001/jamaoto.2018.2716](https://doi.org/10.1001/jamaoto.2018.2716)] [Medline: [30383146](https://pubmed.ncbi.nlm.nih.gov/30383146/)]
27. Tofighi B, Sindhu SS, Chami C, Lewis CF, Dickson VV, Lee JD. Perspectives on the HIV continuum of care among adult opioid users in New York City: a qualitative study. *Harm Reduct J* 2019 Oct 12;16(1):58 [FREE Full text] [doi: [10.1186/s12954-019-0329-z](https://doi.org/10.1186/s12954-019-0329-z)] [Medline: [31606048](https://pubmed.ncbi.nlm.nih.gov/31606048/)]
28. Miller W, Brown J, Simpson T. What works? A methodological analysis of the alcohol treatment outcome literature. Needham Heights, MA: Allyn & Bacon; 1995:12-44.
29. Miller W, Wilbourne PL. Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. *Addiction* 2002 Mar;97(3):265-277. [doi: [10.1046/j.1360-0443.2002.00019.x](https://doi.org/10.1046/j.1360-0443.2002.00019.x)] [Medline: [11964100](https://pubmed.ncbi.nlm.nih.gov/11964100/)]
30. Whiteman KL, Naslund JA, DiNapoli EA, Bruce ML, Bartels SJ. Systematic Review of Integrated General Medical and Psychiatric Self-Management Interventions for Adults With Serious Mental Illness. *Psychiatr Serv* 2016 Nov 01;67(11):1213-1225 [FREE Full text] [doi: [10.1176/appi.ps.201500521](https://doi.org/10.1176/appi.ps.201500521)] [Medline: [27301767](https://pubmed.ncbi.nlm.nih.gov/27301767/)]
31. Vaughn MG, Howard MO. Integrated psychosocial and opioid-antagonist treatment for alcohol dependence: A systematic review of controlled evaluations. *Social Work Research* 2004 Mar 01;28(1):41-53. [doi: [10.1093/swr/28.1.41](https://doi.org/10.1093/swr/28.1.41)]

32. Cabassa LJ, Camacho D, Vélez-Grau CM, Stefancic A. Peer-based health interventions for people with serious mental illness: A systematic literature review. *J Psychiatr Res* 2017 Jan;84:80-89. [doi: [10.1016/j.jpsychires.2016.09.021](https://doi.org/10.1016/j.jpsychires.2016.09.021)] [Medline: [27701013](https://pubmed.ncbi.nlm.nih.gov/27701013/)]
33. Auslander W, Fisher C, Ollie M, Yu M. Teaching Master's and Doctoral Social Work Students to Systematically Evaluate Evidence-Based Interventions. *Journal of Teaching in Social Work* 2012 Sep;32(4):320-341. [doi: [10.1080/08841233.2012.707170](https://doi.org/10.1080/08841233.2012.707170)]
34. Aseltine RH, DeMartino R. An outcome evaluation of the SOS Suicide Prevention Program. *Am J Public Health* 2004 Mar;94(3):446-451. [doi: [10.2105/ajph.94.3.446](https://doi.org/10.2105/ajph.94.3.446)] [Medline: [14998812](https://pubmed.ncbi.nlm.nih.gov/14998812/)]
35. Bailey E, Kryszynska K, O'Dea B, Robinson J. Internet Forums for Suicide Bereavement. *Crisis* 2017 Nov;38(6):393-402. [doi: [10.1027/0227-5910/a000471](https://doi.org/10.1027/0227-5910/a000471)] [Medline: [28793818](https://pubmed.ncbi.nlm.nih.gov/28793818/)]
36. Freedenthal S. Adolescent help-seeking and the Yellow Ribbon Suicide Prevention Program: an evaluation. *Suicide Life Threat Behav* 2010 Dec;40(6):628-639. [doi: [10.1521/suli.2010.40.6.628](https://doi.org/10.1521/suli.2010.40.6.628)] [Medline: [21198332](https://pubmed.ncbi.nlm.nih.gov/21198332/)]
37. Haas A, Koestner B, Rosenberg J, Moore D, Garlow SJ, Sedway J, et al. An Interactive Web-Based Method of Outreach to College Students at Risk for Suicide. *Journal of American College Health* 2008 Jul;57(1):15-22. [doi: [10.3200/jach.57.1.15-22](https://doi.org/10.3200/jach.57.1.15-22)]
38. Han J, Batterham PJ, Caele AL, Wu Y, Xue J, van Spijker BAJ. Development and pilot evaluation of an online psychoeducational program for suicide prevention among university students: A randomised controlled trial. *Internet Interv* 2018 Jun;12:111-120 [FREE Full text] [doi: [10.1016/j.invent.2017.11.002](https://doi.org/10.1016/j.invent.2017.11.002)] [Medline: [30135775](https://pubmed.ncbi.nlm.nih.gov/30135775/)]
39. Pisani AR, Wyman PA, Gurditta K, Schmeelk-Cone K, Anderson CL, Judd E. Mobile Phone Intervention to Reduce Youth Suicide in Rural Communities: Field Test. *JMIR Ment Health* 2018 May 31;5(2):e10425 [FREE Full text] [doi: [10.2196/10425](https://doi.org/10.2196/10425)] [Medline: [29853439](https://pubmed.ncbi.nlm.nih.gov/29853439/)]
40. Robinson J, Bailey E, Hetrick S, Paix S, O'Donnell M, Cox G, et al. Developing Social Media-Based Suicide Prevention Messages in Partnership With Young People: Exploratory Study. *JMIR Ment Health* 2017 Oct 04;4(4):e40 [FREE Full text] [doi: [10.2196/mental.7847](https://doi.org/10.2196/mental.7847)] [Medline: [28978499](https://pubmed.ncbi.nlm.nih.gov/28978499/)]
41. Schilling EA, Lawless M, Buchanan L, Aseltine RH. "Signs of Suicide" shows promise as a middle school suicide prevention program. *Suicide Life Threat Behav* 2014 Dec;44(6):653-667. [doi: [10.1111/sltb.12097](https://doi.org/10.1111/sltb.12097)] [Medline: [24796660](https://pubmed.ncbi.nlm.nih.gov/24796660/)]
42. Wyman PA, Brown CH, LoMurray M, Schmeelk-Cone K, Petrova M, Yu Q, et al. An outcome evaluation of the Sources of Strength suicide prevention program delivered by adolescent peer leaders in high schools. *Am J Public Health* 2010 Sep;100(9):1653-1661. [doi: [10.2105/AJPH.2009.190025](https://doi.org/10.2105/AJPH.2009.190025)] [Medline: [20634440](https://pubmed.ncbi.nlm.nih.gov/20634440/)]
43. Dickter B, Bunge EL, Brown LM, Leykin Y, Soares EE, Van Voorhees B, et al. Impact of an online depression prevention intervention on suicide risk factors for adolescents and young adults. *Mhealth* 2019;5:11 [FREE Full text] [doi: [10.21037/mhealth.2019.04.01](https://doi.org/10.21037/mhealth.2019.04.01)] [Medline: [31231666](https://pubmed.ncbi.nlm.nih.gov/31231666/)]
44. Hetrick SE, Goodall J, Yuen HP, Davey CG, Parker AG, Robinson J, et al. Comprehensive Online Self-Monitoring to Support Clinicians Manage Risk of Suicide in Youth Depression. *Crisis* 2017 May;38(3):147-157. [doi: [10.1027/0227-5910/a000422](https://doi.org/10.1027/0227-5910/a000422)] [Medline: [27659516](https://pubmed.ncbi.nlm.nih.gov/27659516/)]
45. Iorfino F, Davenport TA, Ospina-Pinillos L, Hermens DF, Cross S, Burns J, et al. Using New and Emerging Technologies to Identify and Respond to Suicidality Among Help-Seeking Young People: A Cross-Sectional Study. *J Med Internet Res* 2017 Jul 12;19(7):e247 [FREE Full text] [doi: [10.2196/jmir.7897](https://doi.org/10.2196/jmir.7897)] [Medline: [28701290](https://pubmed.ncbi.nlm.nih.gov/28701290/)]
46. King CA, Eisenberg D, Zheng K, Czyz E, Kramer A, Horwitz A, et al. Online suicide risk screening and intervention with college students: a pilot randomized controlled trial. *J Consult Clin Psychol* 2015 Jun;83(3):630-636 [FREE Full text] [doi: [10.1037/a0038805](https://doi.org/10.1037/a0038805)] [Medline: [25688811](https://pubmed.ncbi.nlm.nih.gov/25688811/)]
47. Bertolote JM, Fleischmann A, De Leo D, Phillips MR, Botega NJ, Vijayakumar L, et al. Repetition of suicide attempts: data from emergency care settings in five culturally different low- and middle-income countries participating in the WHO SUPRE-MISS Study. *Crisis* 2010;31(4):194-201. [doi: [10.1027/0027-5910/a000052](https://doi.org/10.1027/0027-5910/a000052)] [Medline: [20801749](https://pubmed.ncbi.nlm.nih.gov/20801749/)]
48. Czyz EK, King CA, Biermann BJ. Motivational Interviewing-Enhanced Safety Planning for Adolescents at High Suicide Risk: A Pilot Randomized Controlled Trial. *J Clin Child Adolesc Psychol* 2019;48(2):250-262. [doi: [10.1080/15374416.2018.1496442](https://doi.org/10.1080/15374416.2018.1496442)] [Medline: [30142300](https://pubmed.ncbi.nlm.nih.gov/30142300/)]
49. Hetrick SE, Yuen HP, Bailey E, Cox GR, Templer K, Rice SM, et al. Internet-based cognitive behavioural therapy for young people with suicide-related behaviour (Reframe-IT): a randomised controlled trial. *Evid Based Ment Health* 2017 Aug;20(3):76-82. [doi: [10.1136/eb-2017-102719](https://doi.org/10.1136/eb-2017-102719)] [Medline: [28701336](https://pubmed.ncbi.nlm.nih.gov/28701336/)]
50. King CA, Klaus N, Kramer A, Venkataraman S, Quinlan P, Gillespie B. The Youth-Nominated Support Team-Version II for suicidal adolescents: A randomized controlled intervention trial. *Journal of Consulting and Clinical Psychology* 2009 Oct;77(5):880-893 [FREE Full text] [doi: [10.1037/a0016552](https://doi.org/10.1037/a0016552)] [Medline: [19803568](https://pubmed.ncbi.nlm.nih.gov/19803568/)]
51. King R, Nurcombe B, Bickman L, Hides L, Reid W. Telephone counselling for adolescent suicide prevention: changes in suicidality and mental state from beginning to end of a counselling session. *Suicide Life Threat Behav* 2003;33(4):400-411. [doi: [10.1521/suli.33.4.400.25235](https://doi.org/10.1521/suli.33.4.400.25235)] [Medline: [14695055](https://pubmed.ncbi.nlm.nih.gov/14695055/)]
52. Mehlum L, Tørmoen AJ, Ramberg M, Haga E, Diep LM, Laberg S, et al. Dialectical behavior therapy for adolescents with repeated suicidal and self-harming behavior: a randomized trial. *J Am Acad Child Adolesc Psychiatry* 2014 Oct;53(10):1082-1091. [doi: [10.1016/j.jaac.2014.07.003](https://doi.org/10.1016/j.jaac.2014.07.003)] [Medline: [25245352](https://pubmed.ncbi.nlm.nih.gov/25245352/)]

53. Normand D, Colin S, Gaboulaud V, Baubet T, Taieb O. How to stay in touch with adolescents and young adults after a suicide attempt? Implementation of a 4-phones-calls procedure over 1 year after discharge from hospital, in a Parisian suburb. *Encephale* 2018 Sep;44(4):301-307. [doi: [10.1016/j.encep.2017.10.011](https://doi.org/10.1016/j.encep.2017.10.011)] [Medline: [29395246](https://pubmed.ncbi.nlm.nih.gov/29395246/)]
54. McManama O'Brien KH, LeCloux M, Ross A, Gironde C, Wharff EA. A Pilot Study of the Acceptability and Usability of a Smartphone Application Intervention for Suicidal Adolescents and Their Parents. *Arch Suicide Res* 2017;21(2):254-264. [doi: [10.1080/13811118.2016.1182094](https://doi.org/10.1080/13811118.2016.1182094)] [Medline: [27136115](https://pubmed.ncbi.nlm.nih.gov/27136115/)]
55. Asarnow JR, Baraff LJ, Berk M, Grob CS, Devich-Navarro M, Suddath R, et al. An emergency department intervention for linking pediatric suicidal patients to follow-up mental health treatment. *Psychiatr Serv* 2011 Nov;62(11):1303-1309 [FREE Full text] [doi: [10.1176/ps.62.11.pss6211\\_1303](https://doi.org/10.1176/ps.62.11.pss6211_1303)] [Medline: [22211209](https://pubmed.ncbi.nlm.nih.gov/22211209/)]
56. Tan Z, Liu X, Liu X, Cheng Q, Zhu T. Designing Microblog Direct Messages to Engage Social Media Users With Suicide Ideation: Interview and Survey Study on Weibo. *J Med Internet Res* 2017 Dec 12;19(12):e381 [FREE Full text] [doi: [10.2196/jmir.8729](https://doi.org/10.2196/jmir.8729)] [Medline: [29233805](https://pubmed.ncbi.nlm.nih.gov/29233805/)]
57. Yen S, Ranney ML, Tezanos KM, Chuong A, Kahler CW, Solomon JB, et al. Skills to Enhance Positivity in Suicidal Adolescents: Results From an Open Development Trial. *Behav Modif* 2019 Mar;43(2):202-221 [FREE Full text] [doi: [10.1177/0145445517748559](https://doi.org/10.1177/0145445517748559)] [Medline: [29258328](https://pubmed.ncbi.nlm.nih.gov/29258328/)]
58. Yen S, Spirito A, Weinstock L, Tezanos K, Kolobaric A, Miller I. Coping long term with active suicide in adolescents: Results from a pilot randomized controlled trial. *Clin Child Psychol Psychiatry* 2019;24(4):847-859. [doi: [10.1177/1359104519843956](https://doi.org/10.1177/1359104519843956)] [Medline: [31064203](https://pubmed.ncbi.nlm.nih.gov/31064203/)]
59. Silverstone PH, Bercov M, Suen VYM, Allen A, Cribben I, Goodrick J, et al. Long-term Results from the Empowering a Multimodal Pathway Toward Healthy Youth Program, a Multimodal School-Based Approach, Show Marked Reductions in Suicidality, Depression, and Anxiety in 6,227 Students in Grades 6-12 (Aged 11-18). *Front Psychiatry* 2017;8:81 [FREE Full text] [doi: [10.3389/fpsy.2017.00081](https://doi.org/10.3389/fpsy.2017.00081)] [Medline: [28555115](https://pubmed.ncbi.nlm.nih.gov/28555115/)]
60. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety. *J Med Internet Res* 2018 Jun 06;20(6):e199 [FREE Full text] [doi: [10.2196/jmir.9275](https://doi.org/10.2196/jmir.9275)] [Medline: [29875089](https://pubmed.ncbi.nlm.nih.gov/29875089/)]
61. Hermes ED, Lyon AR, Schueller SM, Glass JE. Measuring the Implementation of Behavioral Intervention Technologies: Recharacterization of Established Outcomes. *J Med Internet Res* 2019 Jan 25;21(1):e11752 [FREE Full text] [doi: [10.2196/11752](https://doi.org/10.2196/11752)] [Medline: [30681966](https://pubmed.ncbi.nlm.nih.gov/30681966/)]
62. Ash P. Children Are Different: Liability Issues in Working With Suicidal and Dangerous Youths. *Focus (Am Psychiatr Publ)* 2019 Oct;17(4):355-359. [doi: [10.1176/appi.focus.20190018](https://doi.org/10.1176/appi.focus.20190018)] [Medline: [32047381](https://pubmed.ncbi.nlm.nih.gov/32047381/)]
63. Thompson EC, Fox KA, Lapomardo A, Hunt JI, Wolff JC. Youth Self Report Thought Problems and Sleep Difficulties Are Linked to Suicidal Ideation Among Psychiatrically Hospitalized Adolescents. *J Child Adolesc Psychopharmacol* 2020 Feb 13. [doi: [10.1089/cap.2019.0160](https://doi.org/10.1089/cap.2019.0160)] [Medline: [32053009](https://pubmed.ncbi.nlm.nih.gov/32053009/)]
64. Van Meter AR, Algorta GP, Youngstrom EA, Lechtman Y, Youngstrom JK, Feeny NC, et al. Assessing for suicidal behavior in youth using the Achenbach System of Empirically Based Assessment. *Eur Child Adolesc Psychiatry* 2018 Feb;27(2):159-169 [FREE Full text] [doi: [10.1007/s00787-017-1030-y](https://doi.org/10.1007/s00787-017-1030-y)] [Medline: [28748484](https://pubmed.ncbi.nlm.nih.gov/28748484/)]
65. Van Roy B, Groholt B, Heyerdahl S, Clench-Aas J. Understanding discrepancies in parent-child reporting of emotional and behavioural problems: Effects of relational and socio-demographic factors. *BMC Psychiatry* 2010 Jul 16;10(1). [doi: [10.1186/1471-244x-10-56](https://doi.org/10.1186/1471-244x-10-56)]
66. Berman NC, Sullivan A, Wilhelm S, Cohen IG. Effect of a legal prime on clinician's assessment of suicide risk. *Death Stud* 2016;40(1):61-67. [doi: [10.1080/07481187.2015.1068248](https://doi.org/10.1080/07481187.2015.1068248)] [Medline: [26207570](https://pubmed.ncbi.nlm.nih.gov/26207570/)]
67. Center for Disease Prevention and Control (CDC). 2018. URL: <https://www.cdc.gov/healthyyouth/data/yrbs/pdf/trendsreport.pdf> [accessed 2020-12-12]
68. Abrutyn S, Mueller AS. Toward a robust science of suicide: Epistemological, theoretical, and methodological considerations in advancing suicidology. *Death Stud* 2019 Sep 10;1-6. [doi: [10.1080/07481187.2019.1660081](https://doi.org/10.1080/07481187.2019.1660081)] [Medline: [31502928](https://pubmed.ncbi.nlm.nih.gov/31502928/)]
69. Maple M, Pearce T, Sanford RL, Cerel J. The Role of Social Work in Suicide Prevention, Intervention, and Postvention: A Scoping Review. *Australian Social Work* 2016 Aug 31;70(3):289-301. [doi: [10.1080/0312407x.2016.1213871](https://doi.org/10.1080/0312407x.2016.1213871)]
70. Standley CJ. Expanding our paradigms: Intersectional and socioecological approaches to suicide prevention. *Death Stud* 2020 Feb 12;1-9. [doi: [10.1080/07481187.2020.1725934](https://doi.org/10.1080/07481187.2020.1725934)] [Medline: [32048555](https://pubmed.ncbi.nlm.nih.gov/32048555/)]
71. Witteman HO, Hendricks M, Straus S, Tannenbaum C. Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *The Lancet* 2019 Feb;393(10171):531-540. [doi: [10.1016/s0140-6736\(18\)32611-4](https://doi.org/10.1016/s0140-6736(18)32611-4)]
72. Canetto S. Suicide as culturally scripted: Theory, evidence, and implications for prevention. In: Annual American Association Suicidology Conference. 2020 Presented at: American Association Suicidology; April 2020; Virtual conference.

## Abbreviations

**BIT:** behavioral intervention technology

**MQRS:** Methodological Quality Rating Scale

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 11.03.20; peer-reviewed by Y Perry, P Batterham, K McKay; comments to author 20.04.20; revised version received 26.05.20; accepted 25.06.20; published 09.10.20.*

*Please cite as:*

*Szlyk H, Tan J*

*The Role of Technology and the Continuum of Care for Youth Suicidality: Systematic Review*

*J Med Internet Res 2020;22(10):e18672*

URL: <http://www.jmir.org/2020/10/e18672/>

doi: [10.2196/18672](https://doi.org/10.2196/18672)

PMID: [33034568](https://pubmed.ncbi.nlm.nih.gov/33034568/)

©Hannah Szlyk, Jia Tan. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 09.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Review

# Fitbit-Based Interventions for Healthy Lifestyle Outcomes: Systematic Review and Meta-Analysis

Mickael Ringeval<sup>1</sup>, MSc; Gerit Wagner<sup>2</sup>, PhD; James Denford<sup>3</sup>, PhD; Guy Paré<sup>2</sup>, PhD; Spyros Kitsiou<sup>4</sup>, PhD

<sup>1</sup>École des Sciences de la Gestion, Université du Québec à Montréal, Montreal, QC, Canada

<sup>2</sup>Research Chair in Digital Health, HEC Montreal, Montreal, QC, Canada

<sup>3</sup>Department of Management, Faculty of Social Sciences and Humanities, Royal Military College of Canada, Kingston, ON, Canada

<sup>4</sup>Department of Biomedical and Health Information Sciences, College of Applied Health Sciences, University of Illinois at Chicago, Chicago, IL, United States

**Corresponding Author:**

Guy Paré, PhD

Research Chair in Digital Health

HEC Montreal

3000, Cote-Sainte-Catherine Road

Montreal, QC, H1Y3K9

Canada

Phone: 1 514 882 8672

Email: [guy.pare@hec.ca](mailto:guy.pare@hec.ca)

## Abstract

**Background:** Unhealthy behaviors, such as physical inactivity, sedentary lifestyle, and unhealthful eating, remain highly prevalent, posing formidable challenges in efforts to improve cardiovascular health. While traditional interventions to promote healthy lifestyles are both costly and effective, wearable trackers, especially Fitbit devices, can provide a low-cost alternative that may effectively help large numbers of individuals become more physically fit and thereby maintain a good health status.

**Objective:** The objectives of this meta-analysis are (1) to assess the effectiveness of interventions that incorporate a Fitbit device for healthy lifestyle outcomes (eg, steps, moderate-to-vigorous physical activity, and weight) and (2) to identify which additional intervention components or study characteristics are the most effective at improving healthy lifestyle outcomes.

**Methods:** A systematic review was conducted, searching the following databases from 2007 to 2019: MEDLINE, EMBASE, CINAHL, and CENTRAL (Cochrane). Studies were included if (1) they were randomized controlled trials, (2) the intervention involved the use of a Fitbit device, and (3) the reported outcomes were related to healthy lifestyles. The main outcome measures were related to physical activity, sedentary behavior, and weight. All the studies were assessed for risk of bias using Cochrane criteria. A random-effects meta-analysis was conducted to estimate the treatment effect of interventions that included a Fitbit device compared with a control group. We also conducted subgroup analysis and fuzzy-set qualitative comparative analysis (fsQCA) to further disentangle the effects of intervention components.

**Results:** Our final sample comprised 41 articles reporting the results of 37 studies. For Fitbit-based interventions, we found a statistically significant increase in daily step count (mean difference [MD] 950.54, 95% CI 475.89-1425.18;  $P<.001$ ) and moderate-to-vigorous physical activity (MD 6.16, 95% CI 2.80-9.51;  $P<.001$ ), a significant decrease in weight (MD -1.48, 95% CI -2.81 to -0.14;  $P=.03$ ), and a nonsignificant decrease in objectively assessed and self-reported sedentary behavior (MD -10.62, 95% CI -35.50 to 14.27;  $P=.40$  and standardized MD -0.11, 95% CI -0.48 to 0.26;  $P=.56$ , respectively). In general, the included studies were at low risk for bias, except for performance bias. Subgroup analysis and fsQCA demonstrated that, in addition to the effects of the Fitbit devices, setting activity goals was the most important intervention component.

**Conclusions:** The use of Fitbit devices in interventions has the potential to promote healthy lifestyles in terms of physical activity and weight. Fitbit devices may be useful to health professionals for patient monitoring and support.

**Trial Registration:** PROSPERO International Prospective Register of Systematic Reviews CRD42019145450; [https://www.crd.york.ac.uk/prospero/display\\_record.php?ID=CRD42019145450](https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019145450)

(*J Med Internet Res* 2020;22(10):e23954) doi:[10.2196/23954](https://doi.org/10.2196/23954)



**KEYWORDS**

Fitbit; wearables; healthy lifestyle; meta-analysis; literature review; fuzzy-set qualitative comparative analysis

**Introduction**

Unhealthy behaviors, such as physical inactivity, sedentary lifestyle, and unhealthful eating, remain highly prevalent and pose formidable challenges worldwide [1-4]. These public health problems are associated with mental health problems, cardiovascular diseases, and shorter life expectancies [5-8]. Despite this, only a minority of the population leads healthy lifestyles and meets the general recommendations of 10,000 steps and 20 minutes of moderate-to-vigorous physical activity (MVPA) per day [9-11]. This trend toward physical inactivity affects 23% of the world's population [12], especially adolescents (81%) and adults (23%), according to the World Health Organization [13]. It also contributes to obesity, with over 650 million people affected worldwide [14,15].

To address the lack of physical activity (PA) and resulting health issues, a substantial amount of research has been dedicated to tracker-based interventions [16-18], which may synergize with the growing use of wearable devices by consumers [19-23]. Among several brands of commercial wearables, Fitbit stands out as one of the most popular commercial wearable activity trackers, with more than 63 million devices sold worldwide in the last 10 years [20] and with an active community [21]. Compared with more traditional PA-related interventions, tracker-based interventions are less resource-intensive and time-consuming, enabling health care providers to cover broader groups of patients [22]. In many cases, these interventions combine components such as individual goal setting, the provision of incentives, social support, and social comparison [11,24-27]. In addition, clinical trials of tracker-based interventions differ in terms of the intervention's characteristics (eg, time of follow-up and theory-based nature of the intervention) and the populations addressed (eg, pre-existing conditions and age) [16,17,28].

Evidence on the effectiveness of tracker-based interventions is inconclusive [29]. Recent systematic literature reviews and meta-analyses have found that wearable-based interventions have small-to-medium-size effects on PA (ie, steps and MVPA) among adults [16,17,30-32], and there is no evidence of such effects among children and adolescents [18,33]. Furthermore, there is inconclusive evidence that wearable-based interventions are effective at achieving healthier lifestyles through decreases in sedentary behavior [16,34] or through weight loss [28,30,32,35]. The practice of pooling studies on different types of advanced wearable tracking devices with studies on uniaxial pedometers is increasingly being singled out as contributing to the inconclusive nature of the available evidence [36]. Another important explanation for the inconclusive evidence is that current reviews applied only a correlational approach, using meta-analysis tools to analyze the increasing complexity of wearable-based interventions, which typically involved multiple components [16,33]. The growing volume of clinical trials suggests that the effectiveness of tracker-based interventions may depend on complex configurations of interacting and equifinal features [16,31,33]. For instance, such configurations

could be the combined provision of tracking devices, social comparison, and gamification in an intervention administered to younger participants, or the combined provision of tracking devices and educational material in a theory-based intervention administered to elderly participants. Current meta-analytic techniques are not suitable for assessing the complex and equifinal effects of complex combinations of intervention components on specific outcomes because the core benefit of the meta-analysis is to reveal the importance of individual variables [37,38]. This review attempts to fill this important gap. It proposes a configurational approach that complements meta-analysis findings by assessing what combination of factors works best.

In short, the purpose of this review was to assess the effects of Fitbit-based interventions, compared with nonwearable control groups, on healthy lifestyle outcomes. A further purpose was to assess the most effective intervention components, beside the Fitbit device, and the study characteristics. We therefore conducted a meta-analysis on the effects of Fitbit-based interventions on a range of healthy lifestyle-related outcomes. We focused on Fitbit devices because they are among the most accurate commercially available wearables [39-43] and are, in some cases, comparable to research-grade monitors [44]. The restriction of this review to Fitbit devices is also due to the fact that this brand is by far the most frequently included in interventional studies found in MEDLINE and ClinicalTrials.gov [45]. This stream of research successfully incorporated Fitbit devices into lifestyle interventions to increase PA, reduce overweight or obesity, and manage chronic diseases such as cancer [46-49]. With this study, we also answer the call to assess the effect of wearables on a broader set of healthy lifestyle-related outcomes [28,31,33,50], including PA-related outcomes, which were the exclusive focus of most previous meta-analyses.

**Methods****Review Protocol**

We conducted and reported this review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [51,52]. The protocol for this review was registered in the International Prospective Register of Systematic Reviews (PROSPERO; CRD42019145450).

**Search Strategy**

The following databases were searched on July 13, 2019: MEDLINE, EMBASE, CINAHL, and Cochrane Controlled Register of Trials (CENTRAL). The search was designed to capture studies involving Fitbit activity tracking devices. No language restrictions were applied. The full search strategy is presented in [Multimedia Appendix 1](#). Electronic searches were supplemented with manual screening of the reference lists of the included articles. We also screened the articles retrieved in prior relevant systematic literature reviews.

## Study Selection

Studies were included if (1) a randomized controlled trial (RCT) design was used, (2) the intervention involved using a Fitbit device to improve PA and/or other health-related outcomes (eg, weight loss), and (3) the study reported outcomes related to healthy lifestyle measures (eg, steps, MVPA, weight, and BMI). Only peer-reviewed journal and conference papers were considered.

Articles were screened in a two-step process. First, all titles and abstracts were examined by one author (MR). Any citations that clearly did not meet the inclusion criteria were excluded. Second, all abstracts and full-text articles were examined independently by two authors (MR and GW). Any disagreements in the selection process were resolved through discussion with a third author (GP or SK).

## Data Extraction

Two authors (MR and GW) independently extracted data from each of the included studies. Discrepancies were resolved through discussion and meetings with a third author (SK). The following data were recorded: author; year; country in which the study was conducted; study design; participant characteristics; sample size; intervention description (eg, intervention duration, model of Fitbit used, intervention components, and theoretical basis); control or comparator group description; primary and secondary outcomes (including method of assessment); and main study results, including relevant subgroup analyses. Within- and between-group quantitative findings (eg, mean differences and significance) were summarized for each study.

## Risk of Bias Assessment

Two authors (MR and SK) assessed each study for risk of bias using the Cochrane Collaboration seven domain-based criteria as follows [53,54]: sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective outcome reporting (reporting bias), and other (other bias). Each criterion was scored as low, unclear, or high risk. Disagreements were resolved through discussion.

## Data Analysis

Because of the variability of the included studies, random-effects meta-analyses [55] were performed on the following most frequently reported outcomes using Review Manager (RevMan) [53]: daily step count, MVPA (min/day), sedentary behavior (min/day), and weight (kg). Data were converted to the same units in order to compare the findings. For instance, weekly step counts were divided by 7, whereas data presented as hours per day were divided by 60 to obtain minutes per day. Studies that included multiple intervention groups (eg, group A: Fitbit alone; group B: Fitbit + text messages) were entered once in the meta-analysis to avoid double counting the control group. We retained the group with the fewest interventional components (eg, Fitbit alone) that matched our initial review objectives, and excluded the other intervention group from the analysis [56]. In studies including a control group that received a delayed intervention, we took into consideration the reported

outcomes before the control group received the intervention. For instance, Li et al [57] reported PA changes resulting from a 2-month intervention during which the first half of the intervention was exclusively administered to the intervention group. In this case, we considered the reported outcomes at 1 month. Data presented as mean, standard error (SE), or 95% CI were converted to SD using the RevMan calculator. We analyzed objective and self-reported measures separately because self-reported outcomes have a higher risk of over-estimation [58,59]. Although some studies used different actigraph devices, PA measures (eg, steps and intensity of activity) were reported similarly. In this case, meta-analytic evaluations of the pooled mean difference (MD) in steps/day, min/day of MVPA, min/day of sedentary behavior, and weight (kg) between the intervention and comparison groups for the objective outcome measures were calculated using mean changes or postintervention data, depending on what the authors had reported. When any relevant data were missing, mean or mean changes and corresponding SD were requested from the corresponding author. Authors of studies that presented data in a graphical format were contacted to obtain the exact values. Forest plots for steps were drawn using GraphPad Prism software (GraphPad Software Inc), because the scale in Review Manager has a limit of 1000 points.

In the presence of high statistical heterogeneity in the outcomes reported in the meta-analysis ( $I^2 > 50\%$ ), we conducted subgroup analysis and fuzzy-set qualitative comparative analysis (fsQCA) to explore potential reasons for this heterogeneity and proposed several explanatory hypotheses in our protocol. We assumed that the treatment effect was influenced by (1) a theoretically grounded treatment, (2) the duration of the treatment, and/or (3) the subject's health condition. We considered these subgroup analyses because there is some evidence that theory-based interventions are more effective [60-62] and that the effects of wearable activity trackers may not be sustainable over time, favoring short interventions. Individuals with chronic conditions may respond to the treatment differently from healthy people because the treatment allows chronic patients to live better with their health conditions, while healthy individuals may consider it as a tool to prevent health problems. We also conducted post-hoc subgroup analyses between studies reporting postintervention values and those reporting mean changes from baseline values to explore whether there are any significant differences between the two reporting methods that may introduce bias in the principal meta-analyses. In addition, we identified a very small number of studies ( $n=2$ ) that reported significant differences between groups at baseline, and we conducted sensitivity analyses to assess whether these trials made any difference to the results of the principal meta-analyses. Finally, we assessed publication bias using funnel plot analysis for each outcome included in our meta-analysis. To permit publication bias assessment, funnel plot analysis can be conducted only on outcomes that include 10 or more studies [63].

The fsQCA method can identify complex (ie, nonlinear and nonadditive) causal patterns [37]. It is especially appropriate when dealing with complex interventions [38]. FsQCA considers the necessity and sufficiency of conditions for an outcome. In

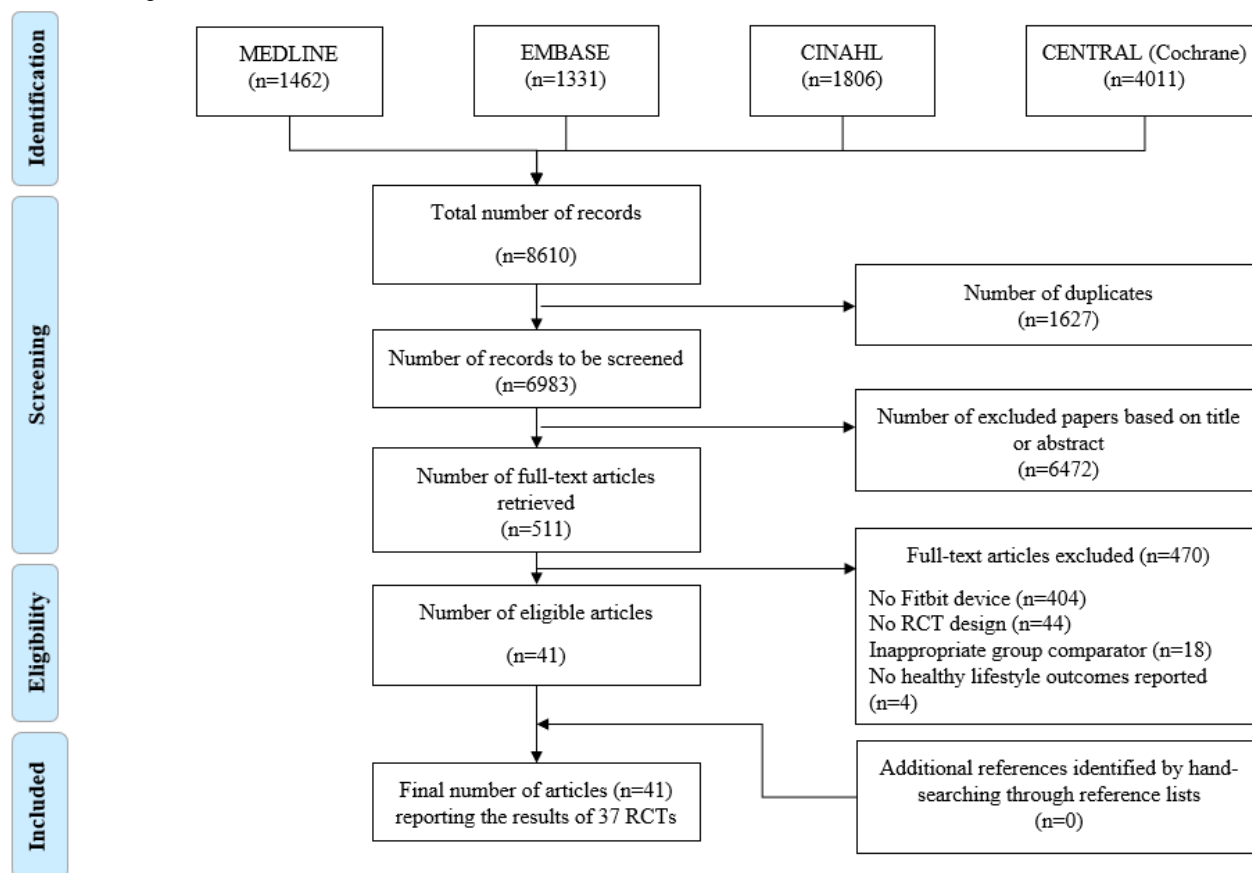
our case, we included the following two types of conditions: the main intervention components that were present in the included studies and the same study characteristics as in the subgroup analysis (ie, grounded in a theory, length of the intervention, and chronic disease in the subjects). We focused on a range of outcomes (ie, steps, MVPA, sedentary behaviors, and weight) that are important components of a healthy lifestyle [64-66]. FsQCA is an analytical method that allows us to assess which configurations of conditions or factors (ie, intervention components and study characteristics) lead to successful outcomes. [Multimedia Appendix 2](#) provides a detailed explanation of how fsQCA was applied.

## Results

### Study Selection

In total, 8610 articles were retrieved using the search strategy. A total of 1627 duplicates were removed, and 6983 records were screened by title and abstract, with 6472 records removed after the application of our selection criteria. The remaining 511 articles were retrieved and assessed for eligibility based on the full text. In total, 41 articles were included, reporting the results of 37 RCTs ([Figure 1](#)) [48,49,57,67-104]. All 41 articles are described in detail, based on study design and PICO (population, intervention, comparison, outcome) characteristics ([Multimedia Appendix 3](#)).

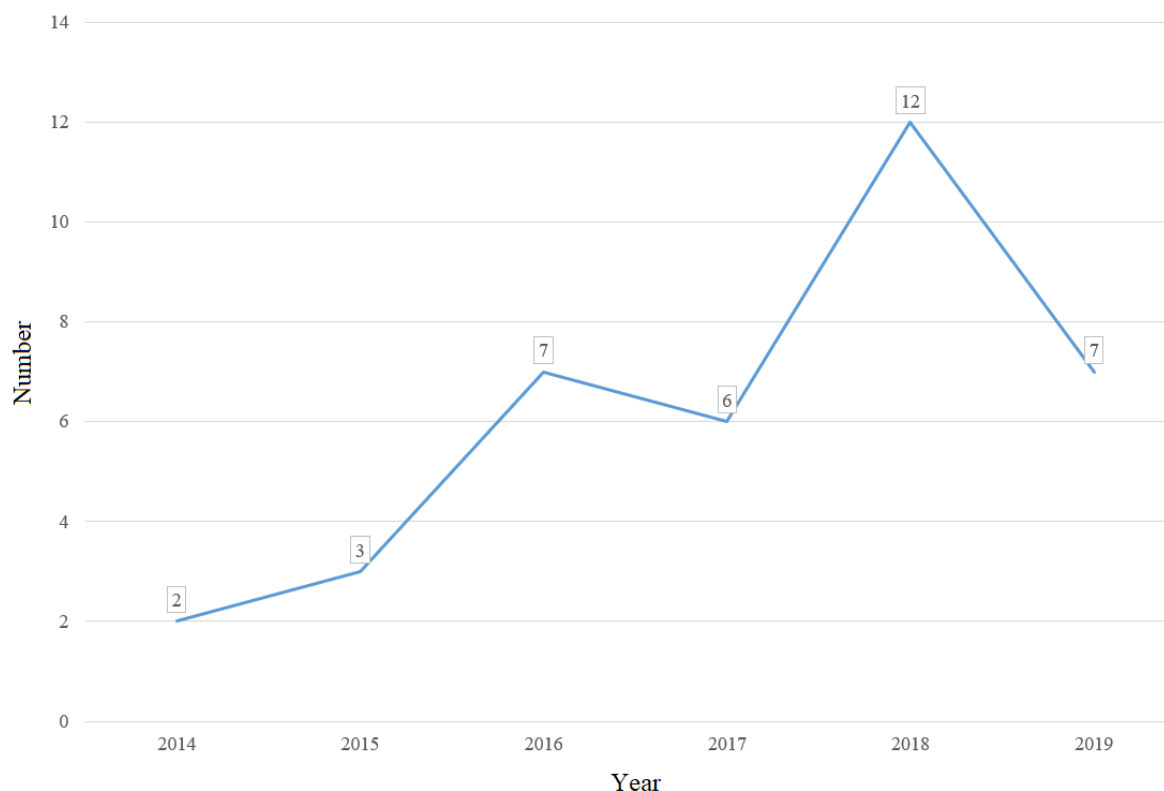
**Figure 1.** Flow diagram. RCT: randomized controlled trial.



### Study Characteristics

The studies were conducted in North America (27/37, 73%), Europe (4/37, 11%), Australia (4/37, 11%), and Asia (2/37, 5%). Approximately two thirds (24/37, 65%) of the studies were conducted in the United States.

The volume of Fitbit-based intervention studies has steadily increased since the first one was published in 2014. The number of articles published per year increased to 12 RCTs in 2018, with most of the studies (25/37, 68%) published in the last 3 years ([Figure 2](#)).

**Figure 2.** Number of Fitbit-based randomized controlled trials published each year.

Most of the studies (35/37, 95%) were parallel RCTs, while two studies used a factorial design and a cross-over design, respectively [79,101]. Among the parallel RCTs, 27 featured two arms [48,57,67-73,76,78,81,83-85,88-93,95,96,99,102-104], while eight featured multiple intervention arms [49,77,82,86,87,94,97,98]. For instance, Finkelstein et al [82] conducted an RCT with four groups; three groups receiving a Fitbit device and the other receiving interventional components differing in terms of the type of financial incentives.

The follow-up duration ranged from 1 week [79] to 1 year [82,90,95,101], with most studies lasting less than 5 months (20/37, 54%).

To summarize, the majority of the studies were conducted in Western countries, adopted a parallel RCT research design, and lasted less than 5 months.

### Population Characteristics

A total of 3779 participants were included, with a mean of 102 participants per study (median 68) and range from 16 [98] to 800 participants [82] per study.

Virtually all studies focused on individuals over the age of 18 years (36/37, 97%). Only one study included adolescents [93]. Thirteen studies included young adults (age 18-43.9 years) [73,77,79,81-83,86,90,91,94,97,99,102], 17 studies included middle-aged adults (age 44-64.9 years) [48,49,57,67-72,78,84,87-89,96,103,104], and six studies included older adults (age  $\geq 65$  years) [76,85,92,95,98,101].

Concerning the main characteristics of the targeted population, most studies (21/37, 57%) reported that their participants had a particular condition or were at risk. This included patients with cardiovascular risks [70,98] and patients having chronic diseases like chronic obstructive pulmonary disease [85] or cardiometabolic diseases [69]. In the remaining studies, participants were selected based on their personal or professional status (eg, employees or students, such as medical students) (8/37, 22%) or their health status (eg, postoperation and cancer survivor) (8/37, 22%). Only one study focused specifically on healthy subjects [81]. Table 1 summarizes the main characteristics of each targeted population.

**Table 1.** Characteristics of the participants.

Main characteristic and specific characteristics	Reference
<b>Having a condition or being at risk</b>	
Overweight/obese	[48,71,83,97,104]
Sedentary	[68,94,99,101]
Arthritis	[57,87,89]
Cardiovascular risks	[70,98]
Diabetes	[73,88]
Cardiometabolic diseases	[69]
Chronic low back pain	[67]
Chronic obstructive pulmonary disease	[85]
Prediabetes	[49]
With low ankle brachial index	[92]
<b>Personal/professional status</b>	
Students	[77,79,86,90,91,102]
Community-dwelling people	[95]
Employees	[82]
<b>Health status</b>	
Postoperation/posttreatment	[76,78,96,103]
Cancer survivor	[72,84,93]
Healthy	[81]

To summarize, the included studies had sample sizes of less than 100 adult subjects, who mainly had a chronic condition or were at risk of having one.

### Intervention Characteristics

The intervention components were highly heterogeneous (Multimedia Appendix 4). Four studies included at least one interventional arm involving only the use of a Fitbit device [77,79,81,91], while the majority (35/37, 95%) included at least one interventional arm involving a comprehensive program for improving PA and facilitating weight loss. For instance, the components of the intervention in the study by Amorim et al [67] included an information booklet on PA and sedentary behavior, a tailored PA plan, a face-to-face coaching session, 12 phone calls from a health coach, weekly personalized messages to encourage participants to achieve their goals, and a Fitbit device with its web-interfaced IMPACT mobile app. In addition to the wearable device, other intervention components included the use of an app or a website (sometimes different from the app provided by the device manufacturer), goal setting and prescription, messaging, education, counseling and feedback, social support, financial incentives, and the provision of another device (Multimedia Appendix 4). Further details on these intervention components are provided in Multimedia Appendix 5.

As expected, there was a wide variety of Fitbit devices used in the included studies. Most of them (17/37, 46%) used clip-on devices such as Fitbit Zip [76,79,82,83,87,88,90,92,96-98], Fitbit One [48,68,81,84,94], and Fitbit Ultra [85]. Seven studies did not specify which model was used [49,67,69,70,95,101,102].

The remaining studies used wrist-worn devices, such as a Fitbit Flex [57,71,73,77,86,89,91,93,103,104] and Fitbit Charge [72,78,99]. Use patterns with these devices were mentioned in 35% (13/37) of the studies. Beside reporting use or wear durations [49,88,103], the studies indicated Fitbit use as the frequency at which the subjects wore the device [48,67,93,95,96] and the number of subjects having Fitbit measurements [73,82,87,90,92]. This information was mainly assessed using device data or it was self-reported.

In summary, most of the interventions did not rely on theory and used Fitbit Zip or Flex tracking devices for interventional purposes along with several other components commonly related to goal setting and education.

### Control Group Characteristics

Most studies included some form of PA or other healthy lifestyle education component [49,67,68,70,72,73,82,84,86,87,90,99,103]. Other studies involved usual care [48,76,78,83,85,88,93,96-98], financial incentives [68,82], blinded wearables [94,101], or no intervention at all [77,79,81,91,92]. Participants in some control groups were put on a waiting list to receive the same intervention following a delay [57,69,71,89]. Two studies included a control group to which overlapping intervention components were allocated [102,104]. For example, in the RCT published by Vandelanotte et al [104], the comparison group received the same intervention components as the interventional arm, except for the Fitbit Flex.



## Study Outcomes

Taken together, the studies in our sample reported a wide range of outcomes that can be classified into the following several categories: PA-related outcomes (eg, steps, MVPA, and light PA), weight-related outcomes (eg, weight and BMI), sedentary behavior outcomes, dietary intake-related outcomes, oxygen uptake outcomes, sleep-related outcomes, quality of life, self-efficacy, and overall health ([Multimedia Appendix 6](#)). While steps were usually reported in steps/day, MVPA was reported in different ways, such as min/day, days/week, metabolic equivalent (MET)-min/week, MET-min/week in 10-minute bouts, etc. This was also the case for sedentary behavior outcomes. These behaviors were mostly reported in min/day [68,89,93,103,104], while few studies reported them as prolonged sedentary 30-minute bouts (%/day) [100] or sedentary activity (<5000 steps per day, %) [87].

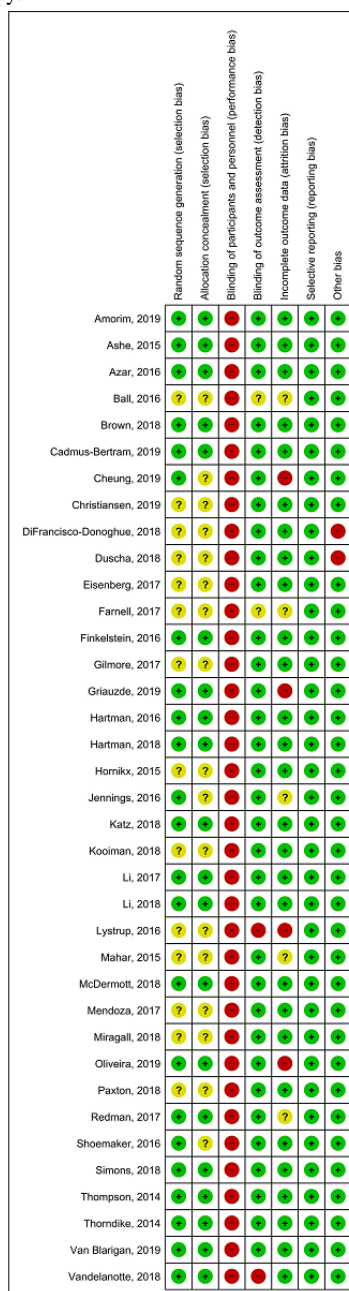
Furthermore, the studies featured several ways of measuring the reported outcomes, especially steps, MVPA, and sedentary behavior. For steps, eight studies (22%) used the intervention Fitbit device as a measurement tool [77-79,86,88,90,94,102]. This outcome was also frequently (8/37, 22%) assessed using a research-grade accelerometer (ie, actigraph) [67,68,72,76,82,96,99,103]. Other studies reported measures taken with a SenseWear Mini device [89], a Jawbone Up wearable [87], a Yamax pedometer [73], and a Dynaport MoveMonitor device [85]. One study did not report how the step counts were assessed [91]. MVPA outcomes were mostly measured using a research-grade accelerometer (13/37, 35%) [48,67,68,71,72,76,82,84,93,95,98,99,103], and, less often, they were self-reported (5/37, 14%) [67,79,86,88,104] or relied on the use of a SenseWear Mini device (2/37, 5%) [57,89]. One study used the intervention Fitbit device to measure MVPA [78]. Finally, research-grade accelerometers [68,82,93,98,103] and self-reported measures [82,104] were used to assess

sedentary behaviors. Two studies measured sedentary behaviors using a SenseWear Mini device [57,89] and one study used a Jawbone Up wearable [87].

To summarize, the available evidence was primarily based on studies that reported PA outcomes, such as steps and MVPA, mostly measured with an actigraph.

## Risk of Bias

Risk-of-bias judgements are presented in [Figure 3](#). Random sequence generation was assessed as being at low risk of bias (23/37, 62%) or unclear (14/37, 38%) in the included studies. Allocation concealment was assessed as being at low risk of bias (20/37, 54.05%) or unclear (17/37, 45.95%) in the included studies. The blinding of participants and personnel was assessed as being at high risk of bias for all the studies because the nature of the intervention and control conditions rendered blinding not feasible. Blinding of outcome assessment was evaluated only in terms of the primary outcome of interest, and was reported as low in cases where it was measured objectively. This was assessed as being at high risk of bias in two studies because of the use of subjective measures [104] and because the authors clearly mentioned that it was an unblinded clinical trial [90]. The management of incomplete outcome data was assessed as being at high risk of bias in four studies [49,73,90,95]. The reasons were high attrition (more than 25%) [95], high imbalance in loss of follow-up between groups [49], limitation of the analysis to subjects who had completed a running event [90], and long periods during the intervention when the activity monitor was not worn [73]. Selective reporting was assessed as being at low risk of bias for all the included studies. Finally, two studies were assessed as having a high risk for other bias because of conflicts of interest declared by the authors [78] and significant differences between the groups at baseline [77]. All the reasons in the risk of bias assessment can be found in [Multimedia Appendix 7](#).

**Figure 3.** Risk of bias summary for each included study.

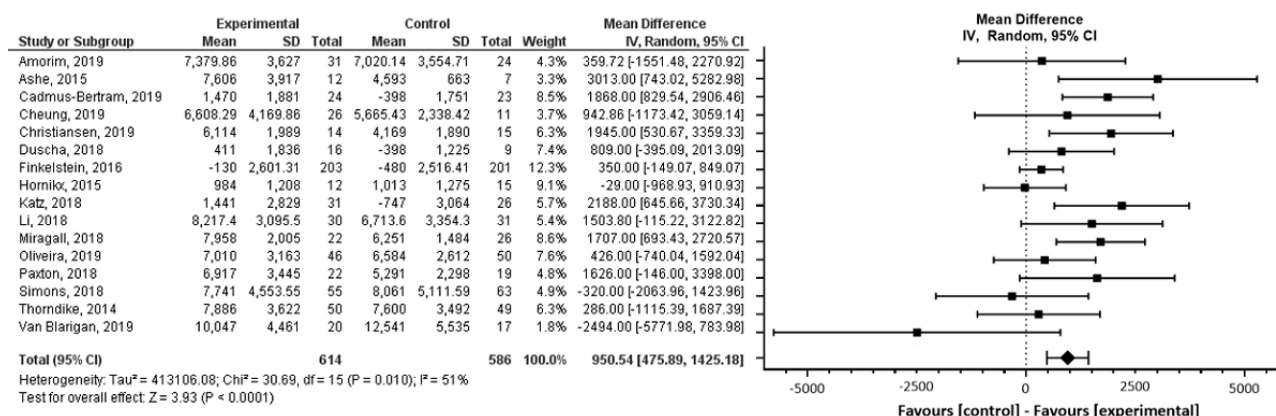
## Meta-Analysis Results

We proceed by presenting the main results of the meta-analysis. Detailed information on all subgroup meta-analyses can be found in [Multimedia Appendix 8](#).

### Steps

Of the 37 studies included in the review, 23 reported an outcome related to steps [67,68,72,73,76-79,82,85-92,94-96,99,102,103] and 16 reported this outcome in a way we could use in the

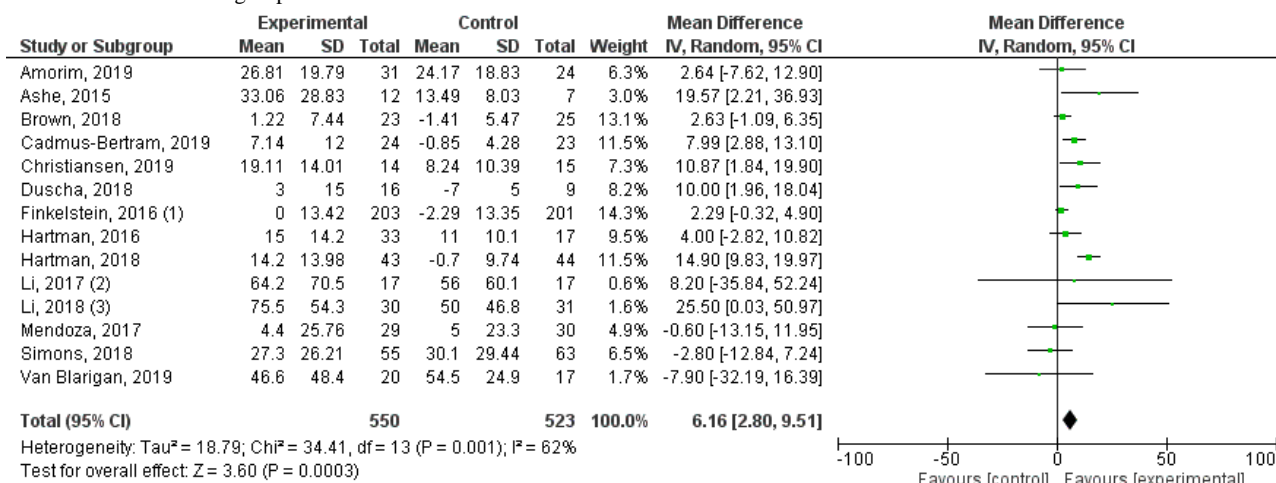
meta-analysis [67,68,72,73,76,78,82,85,87,89,94-96,99,102,103]. On average, Fitbit-based interventions were associated with a statistically significant increase in the number of daily steps when compared with the control groups (MD 950.54, 95% CI 475.89-1425.18;  $P < .001$ ; [Figure 4](#)) across most of the studies in the meta-analysis (13/16, 81%). Three studies showed a decrease in the number of steps [85,99,103]. There was high heterogeneity between study results in terms of the magnitude of the effects ( $I^2=51\%$ ).

**Figure 4.** Forest plot of mean difference in steps per day in studies comparing an intervention that included a Fitbit device with a control group that did not utilize such a device.

Subgroup analyses showed that the length of the intervention, the subjects' health condition, and theory-based interventions did not have significant impacts on the number of steps ( $P=.97$ ,  $P=.32$ , and  $P=.86$ , respectively). When we categorized studies by reporting method, we found no evidence of clinically or statistically significant ( $P=.86$ ) differences between studies that reported postintervention data and those that reported mean change from baseline data (Multimedia Appendix 8). A sensitivity analysis excluding the study by Van Blarigan et al [103], in which there was an imbalance in steps between the control and intervention groups at baseline, showed similar results ( $P=.87$ ). Funnel plot analysis showed no evidence of publication bias (Multimedia Appendix 9).

### MVPA

Of the 37 studies included in the review, 21 reported MVPA [48,57,67,68,71-73,76,78,79,82,84,86,88,89,93,95,98,99,103,104] and 14 studies reported this outcome in a way that we could use in the meta-analysis [48,57,67,68,71,72,76,78,82,84,89,93,99,103]. There was a statistically significant increase in minutes per day spent on MVPA in the Fitbit-based interventions compared with the comparison groups (MD 6.16, 95% CI 2.80-9.51;  $P<.001$ ; Figure 5). The study results featured high heterogeneity ( $I^2=62\%$ ).

**Figure 5.** Forest plot of mean difference in moderate-to-vigorous physical activity (MVPA; min/day) in studies comparing an intervention that included a Fitbit device with a control group that did not utilize such a device.

#### Footnotes

(1) reported as MVPA in bouts

(2) reported as MVPA in bouts > 10 minutes

(3) reported as MVPA in bouts > 10 minutes

The subgroup analyses showed that only theory-based interventions had a significant impact on MVPA ( $P<.001$ ), in contrast to the findings for length of follow-up and subjects' health condition ( $P=.28$  and  $P=.29$ , respectively). When we categorized studies by reporting method, we found no evidence of clinically or statistically significant ( $P=.93$ ) differences between studies that reported postintervention data and those that reported mean change from baseline data (Multimedia Appendix 8). A sensitivity analysis excluding the study by Van

Blarigan et al. [103], in which there was an imbalance in MVPA between the control and intervention groups at baseline, showed no significant ( $P=.92$ ) impact on the overall effect size. There was no evidence of publication bias (Multimedia Appendix 9).

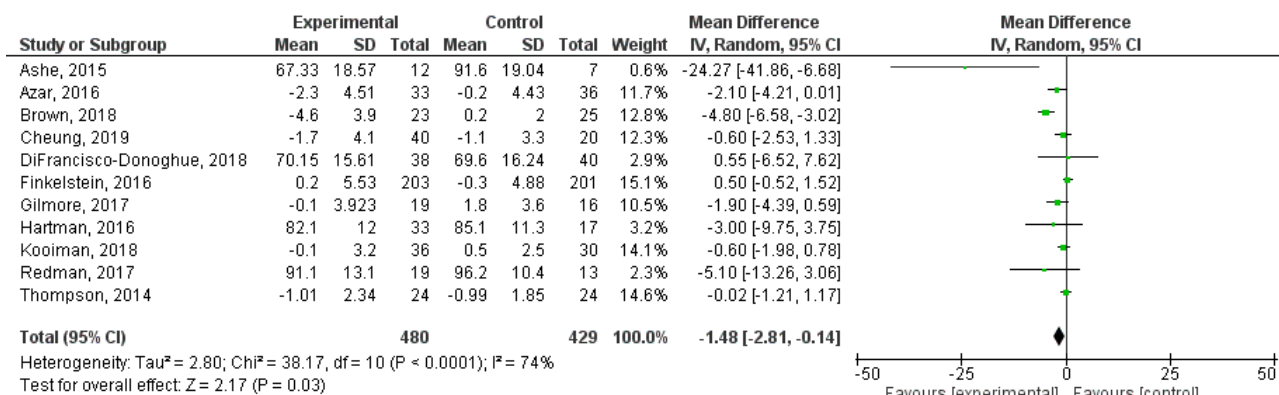
### Weight

Of the 37 studies included in the review, 15 reported an outcome related to weight [48,68-71,73,77,82-84,88,97,101,102,104] and 11 studies reported this outcome in a way that we could use

in the meta-analysis [48,68,69,71,73,77,82,83,88,97,101]. Weight was measured by the research team. A random-effects meta-analysis using MD performed on the 11 studies showed

a significant decrease in weight in the Fitbit-based interventions compared with the control groups (MD  $-1.48$ , 95% CI  $-2.81$  to  $-0.14$ ;  $P=.03$ ; Figure 6). Heterogeneity was high ( $I^2=74\%$ ).

**Figure 6.** Forest plot of mean difference in weight (kg) in studies comparing an intervention that included a Fitbit device with a control group that did not utilize such a device.



The subgroup analyses showed that only subjects' health condition had a significant impact on weight ( $P=.009$ ), in contrast to the findings for length of follow-up and theory-based interventions ( $P=.26$  and  $P=.31$ , respectively). When we categorized studies by reporting method, we found no evidence of clinically or statistically significant ( $P=.30$ ) differences between studies that reported postintervention data and those that reported mean change from baseline data (Multimedia Appendix 8). A sensitivity analysis excluding the study by Ashe et al. [68], in which there was a weight difference between the control and intervention groups at baseline, showed similar results ( $P=.86$ ). Publication bias was not detected for this outcome (Multimedia Appendix 9).

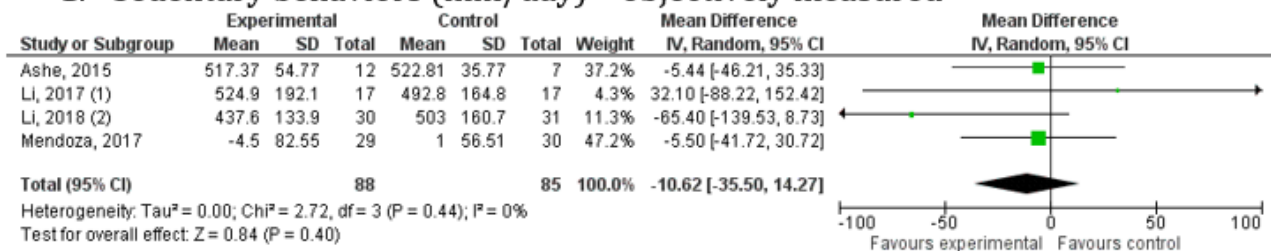
### Sedentary Behaviors

Of the 37 studies included in the review, 10 reported an outcome related to sedentary behaviors [57,68,82,87,89,93,98,103,104] and six reported this outcome in a way we could use in the

meta-analysis [57,68,82,89,93,104]. Sedentary behavior was measured objectively, except in two studies [82,104] that utilized a self-reported questionnaire to obtain daily sitting time or sedentary behavior. A random-effects meta-analysis was performed on four studies that objectively measured sedentary behavior using MD. The other two were assessed using standardized mean difference (SMD). For objective measures, there was a nonsignificant decrease in sedentary behavior following the intervention compared with the control comparator (MD  $-10.62$ , 95% CI  $-35.50$  to  $14.27$ ;  $P=.40$ ; Figure 7) across most of the studies in the meta-analysis (3/4, 75%), with a low level of heterogeneity ( $I^2=0\%$ ). For self-reported measures, there was a nonsignificant decrease in sedentary behavior following the intervention compared with the control comparator (SMD  $-0.11$ , 95% CI  $-0.48$  to  $0.26$ ;  $P=.56$ ; Figure 7), with a high level of heterogeneity ( $I^2=69\%$ ). Given the small sample size, no subgroup analysis could be conducted.

**Figure 7.** Forest plot of sedentary behaviors (min/day) in studies comparing an intervention that included a Fitbit device with a control group that did not utilize such a device.

### 1. Sedentary behaviors (min/day) – objectively measured

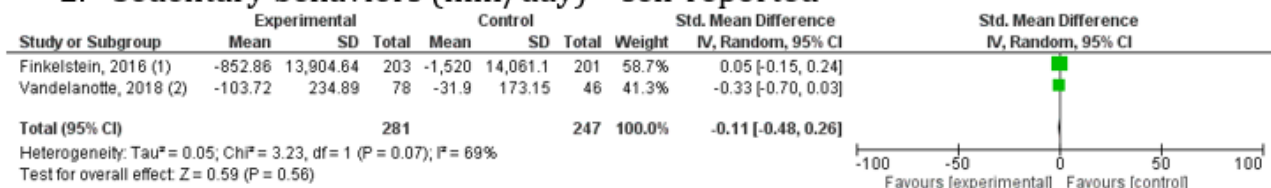


#### Footnotes

(1) reported in bouts > 20 minutes

(2) reported in bouts > 20 minutes

### 2. Sedentary behaviors (min/day) – self-reported



#### Footnotes

(1) assessed with the Global Physical Activity Questionnaire (GPAQ).

(2) assessed with the Workforce Sitting Questionnaire.

## FsQCA Results

We conducted all our analyses using a frequency cut-off of 2 per relevant configuration and a minimum raw consistency of 0.8, combined with a minimum proportional reduction in consistency (PRI consistency) of 0.6. This is consistent with the report by Rihoux and Ragin [37].

First, we conducted a separate analysis of each condition (intervention components and study characteristics) owing to the lack of cases for joint analyses with all conditions (Multimedia Appendix 2). Configuration analyses by

intervention components covered a relatively high number of observed cases (35.14%), whereas the study characteristics did not cover enough cases to analyze them further (no configuration had a high enough raw coverage to conduct an analysis) (Multimedia Appendix 10 and Multimedia Appendix 11). Based on these results and the meta-analysis results, we then combined the following conditions: goal setting, messaging, counseling, length of intervention (named “follow-up duration” below, see details in Multimedia Appendix 2), theory-based interventions, and subjects’ conditions. Figure 8 depicts the fsQCA results using the notation system from Ragin and Fiss [105].



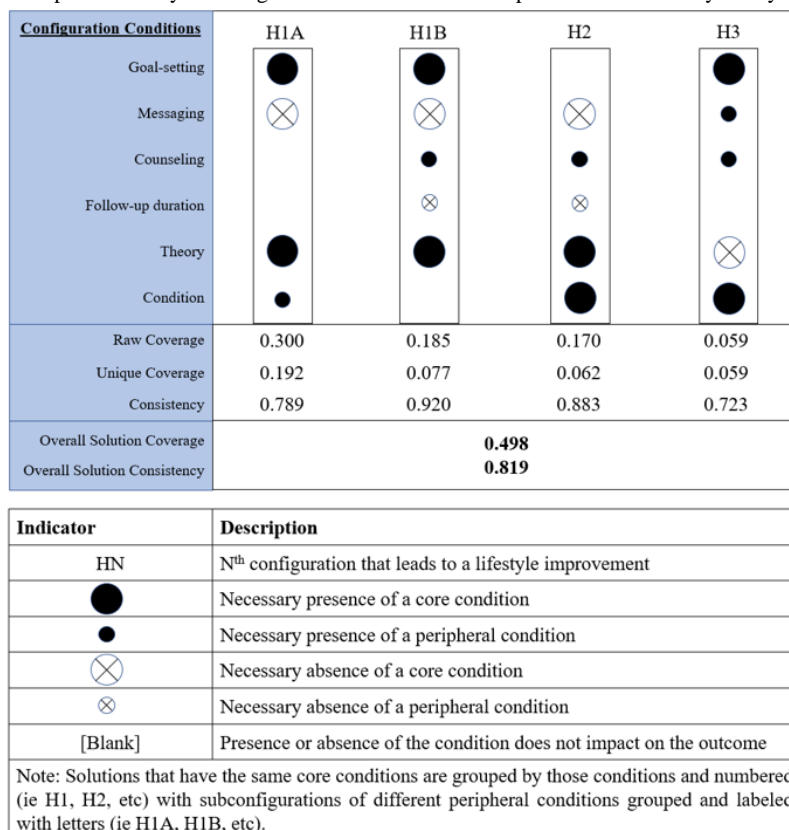
**Figure 8.** Fuzzy-set qualitative comparative analysis configurations associated with improvements in healthy lifestyle outcomes.

Figure 8 shows two measures that validated the solutions (consistency and coverage). HN indicates the Nth configuration that leads to a lifestyle improvement. Initially, overall solution consistency measures the degree to which all configurations together consistently result in an improvement in healthy lifestyle outcomes. In our case, overall consistency was 0.819, which is above the usually acceptable level of 0.80 [106]. Raw coverage shows empirical relevance and effectiveness of the solution for the outcome, although higher coverage does not necessarily mean theoretical importance [106]. Thus, there are

multiple paths to better outcomes for individuals with pre-existing conditions, but the most effective one (albeit for the short- or long-term interventions) centers on theory-based interventions with goal-setting but without messaging (H1A). In fact, in all but one configuration, goal setting was a condition for better outcomes, with or without either messaging or counselling. To differentiate between the numerous outcomes, Table 2 presents these results according to intervention length and subjects' health condition.

**Table 2.** Configurations leading to better lifestyle outcomes depending on the intervention length and subjects' health condition.

Intervention length and subjects' health condition	Follow-up duration (long follow-up duration) <sup>a,b</sup>	~Follow-up duration (short follow-up duration) <sup>b,c</sup>
Condition (subjects with a chronic condition) <sup>a</sup>	H1A H3	H1A H1B H2 H3
~Condition (healthy subjects) <sup>c</sup>	N/A <sup>d</sup>	H1B

<sup>a</sup>Configuration element that needs to be present for an improved lifestyle outcome.

<sup>b</sup>HN: Nth configuration that leads to a lifestyle improvement.

<sup>c</sup>Configuration element that needs to be absent for an improved lifestyle outcome.

<sup>d</sup>Configuration absent in our sample

Table 2 suggests that there are more paths to success for people with pre-existing conditions and none for people without pre-existing conditions who receive long-term interventions. Although it may appear reasonable to assume that the additive effects of more intervention components lead to better outcomes [16], this is not confirmed by our analyses. The configuration

with all the intervention components (H3) is not the most effective (it has the lowest raw coverage).

## Discussion

### Principal Findings

Our review summarizes the results of interventions that included a Fitbit wearable device to improve healthy lifestyle outcomes. Our meta-analysis results showed improvements in two PA outcomes, namely, steps and MVPA. Fitbit-based interventions also resulted in weight loss. However, sedentary behaviors did not improve, regardless of whether they were measured objectively or self-reported. These results are in line with prior reviews on wearables that showed no change in sedentary behaviors [16], and an improvement in PA outcomes [16-18,28,30-32,35,36,107] and weight loss [28,32,35,108]. However, the current evidence is mostly representative for adults and older subjects. The lack of studies focused on vulnerable populations, such as youth and adolescents, may be explained by some of the challenges faced when recruiting subjects from these populations and conducting RCTs (eg, securing consent from the legal guardian).

Considering the overall high heterogeneity in our meta-analysis results, we followed well-established guidelines to investigate them further [109]. We did this by applying the following two different methodologies: subgroup analysis and fsQCA. The former allows us to answer specific questions about a particular aspect of the study (eg, length and theory-based approach) and types of intervention components or patient characteristics (eg, age and condition), whereas the latter emphasizes the configuration of factors (eg, intervention components and study characteristics). Subgroup analyses showed significant ( $P<.001$ ) improvements in MVPA among nontheory-based interventions. This contradicts the results of McCullough et al [62], who found that theory-based interventions are more effective. We also observed that weight loss was more significant ( $P=.009$ ) among patients with chronic conditions. Further analyses using fsQCA uncovered additional interesting results. We found that both theory- and nontheory-based interventions contributed, but it would appear that this factor depends on specific conditions in order to lead to effective interventions. In all but one configuration, goal setting was a condition for better outcomes, with or without either messaging or counselling. This is true for lifestyle outcomes and weight outcomes (Multimedia Appendix 12, Multimedia Appendix 13, and Multimedia Appendix 14). Interestingly, neither the presence nor the absence of goal setting improved PA. Instead, the absence of messaging and/or absence of counselling were the most relevant conditions for improving PA. This was not expected, because previous studies have found these two intervention components effective on their own [110-112]. When combining messaging and counselling with other conditions, it appears that they are outweighed by other factors, such as education and subjects' health conditions. Future research is thus needed to investigate what mitigates the contributions made by messaging and counselling in interventions to improve PA. Of interest is the lack of better outcomes in long-term studies among participants without pre-existing conditions, and the more limited number of paths to success for longer term participants with health conditions (Table 2). This either suggests that there were fewer such studies or that it is difficult for subjects to maintain their

focus in order to achieve long-term results. The latter can easily be understood for weight outcomes since early weight loss is rapid and then tends to stall on a plateau for an extended period of time [113].

Moreover, we observed that better results were achieved with a combination of study and intervention components as compared to intervention components alone. In other words, there is complex causality at play, whereby individual and study characteristics are also important criteria to consider when evaluating the effectiveness of an intervention. This is coherent with the idea that technological interventions may not produce similar effects in different individuals [114]. This means that studies that do not adequately consider study characteristics and participant profiles may produce invalid conclusions regarding the effectiveness of an intervention.

Furthermore, we found that goal setting was the most promising intervention component, whereas messaging seemed to be mostly ineffective in complex interventions. These results can be illustrated with two studies from our sample. Amorim et al [67] found that setting goals increased outcomes related to steps and other activities, such as yoga and swimming. In contrast, Cheung et al [73] concluded that "the vast majority (of participants) found that the messages (on PA, nutrition, and general health and motherhood information and education) were helpful, although the reported effects on diet and PA were more modest." Finally, the length of an intervention does not appear to be relevant, because it was not significant in the subgroup analysis ( $P=.97$  for steps,  $P=.28$  for MVPA, and  $P=.26$  for weight), and the most dominant configuration (H1A) was not affected by this factor.

### Strengths and Limitations

Prior systematic literature reviews are limited by the quality and nature of the studies included. To avoid this, we included only studies featuring a Fitbit device as an interventional component. Despite this, and much like other reviews, the studies in our sample involved very heterogeneous interventions, rendering assessments of the effects of Fitbit interventions more difficult. However, a thorough systematic and transparent methodology was followed [115,116], and the use of meta-analysis tools and fsQCA allowed us to interpret the combined effects of Fitbit devices with the other interventional components and subject characteristics. Using these two methodologies enabled us to provide a fine-grained picture of the effectiveness of Fitbit-based interventions. Despite promising findings, applications of QCA in systematic reviews are still relatively new, especially in digital health research [38,117]. We hope that this review will help promote its application in future studies.

The results of our review must be interpreted in light of some limitations. First, even though we included a large range of outcomes, we could not assess the effectiveness of the interventions on each of them. Rather, we limited our analyses to PA outcomes, sedentary behaviors, and weight. While most of the articles in our sample examined well-studied outcomes (eg, steps and MVPA), other studies reported less common ones, such as cognition and dietary intake. Second, we could not assess the effectiveness of the Fitbit device itself on healthy

lifestyle outcomes. This was due to (1) the high complexity and variety of the interventions and (2) the number of studies that did not describe the Fitbit artifact. Indeed, the studies in our sample rarely described the wearable and ignored its specific features. As shown by Lyons et al [118] and Mercer et al [119], each device incorporates different behavioral change techniques that are linked to one or several features of the wearable, so providing a description of the features of the device and the associated app (if used) is essential for future research. Consideration of the features of these devices is also important because James et al [120] found that each set of features does not impact health outcomes equally. This study suggests that only the social interaction and data management features of activity trackers help improve well-being outcomes. Finally, we could not assess the effects of the different behavioral change techniques incorporated in the Fitbit devices as proposed in our protocol owing to high heterogeneity and the lack of information reported in the included studies. This gave us the opportunity to apply a new methodology (fsQCA) that is relevant to complex interventions in order to determine the most important conditions for Fitbit-based interventions.

## Conclusions and Future Research

Fitbit devices, included either as the primary component of an intervention or as part of a more comprehensive and complex intervention, have the potential to improve healthy lifestyle behaviors and, in particular, PA. The included studies encompassed mainly adult populations with pre-existing chronic conditions. Although the findings were not significant in all the RCTs, short-term interventions utilizing a Fitbit device generally resulted in improvements in terms of a healthy lifestyle. In addition to these activity trackers, we showed that goal setting is an effective complementary interventional component over the short and long term. Further research would be beneficial to determine the effect of a Fitbit device independent of other interventional components, as would investigations into the cost-effectiveness of Fitbit-based interventions. Given the potential associated with the use of PA trackers, further studies investigating their long-term use would be useful to guide potential clinical applications and future recommendations. Finally, future research could also focus on the effectiveness of such interventions in healthy subjects and consider subjective outcomes, such as psychological health and personal motivation.

---

## Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Search queries.

[[DOCX File, 20 KB - jmir\\_v22i10e23954\\_app1.docx](#)]

---

### Multimedia Appendix 2

Fuzzy-set qualitative comparative analysis methodology details.

[[DOCX File, 18 KB - jmir\\_v22i10e23954\\_app2.docx](#)]

---

### Multimedia Appendix 3

PICO table.

[[DOCX File, 48 KB - jmir\\_v22i10e23954\\_app3.docx](#)]

---

### Multimedia Appendix 4

Characteristics of the intervention components in the included studies.

[[XLSX File \(Microsoft Excel File\), 20 KB - jmir\\_v22i10e23954\\_app4.xlsx](#)]

---

### Multimedia Appendix 5

Intervention component details in the included studies.

[[DOCX File, 24 KB - jmir\\_v22i10e23954\\_app5.docx](#)]

---

### Multimedia Appendix 6

Reported outcomes in the included studies.

[[DOCX File, 295 KB - jmir\\_v22i10e23954\\_app6.docx](#)]

---

### Multimedia Appendix 7

Risk of bias assessment results.

[[PDF File \(Adobe PDF File\), 565 KB - jmir\\_v22i10e23954\\_app7.pdf](#)]

---

### Multimedia Appendix 8

Subgroup analysis results.

[DOCX File , 1598 KB - [jmir\\_v22i10e23954\\_app8.docx](#) ]

#### Multimedia Appendix 9

Funnel plots for publication bias.

[DOCX File , 25 KB - [jmir\\_v22i10e23954\\_app9.docx](#) ]

#### Multimedia Appendix 10

Fuzzy-set qualitative comparative analysis truth tables for main results.

[DOCX File , 22 KB - [jmir\\_v22i10e23954\\_app10.docx](#) ]

#### Multimedia Appendix 11

Fuzzy-set qualitative comparative analysis configuration charts for main results.

[DOCX File , 70 KB - [jmir\\_v22i10e23954\\_app11.docx](#) ]

#### Multimedia Appendix 12

Fuzzy-set qualitative comparative analysis truth tables for secondary analysis.

[DOCX File , 24 KB - [jmir\\_v22i10e23954\\_app12.docx](#) ]

#### Multimedia Appendix 13

Fuzzy-set qualitative comparative analysis configuration charts for secondary analysis.

[DOCX File , 205 KB - [jmir\\_v22i10e23954\\_app13.docx](#) ]

#### Multimedia Appendix 14

Fuzzy-set qualitative comparative analysis dataset.

[DOCX File , 17 KB - [jmir\\_v22i10e23954\\_app14.docx](#) ]

## References

1. Zhou B, Lu Y, Hajifathalian K, Bentham J, Di Cesare M, Danaei G, et al. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants. *The Lancet* 2016 Apr 09;387(10027):1513-1530. [doi: [10.1016/S0140-6736\(16\)00618-8](#)] [Medline: [27061677](#)]
2. Tuchman A. Diabetes and the public's health. *The Lancet* 2009 Oct 03;374(9696):1140-1141. [doi: [10.1016/s0140-6736\(09\)61730-x](#)] [Medline: [19810203](#)]
3. Kohl HW, Craig CL, Lambert EV, Inoue S, Alkandari JR, Leetongin G, et al. The pandemic of physical inactivity: global action for public health. *The Lancet* 2012 Jul 21;380(9838):294-305. [doi: [10.1016/S0140-6736\(12\)60898-8](#)] [Medline: [22818941](#)]
4. Heath GW, Parra DC, Sarmiento OL, Andersen LB, Owen N, Goenka S, et al. Evidence-based intervention in physical activity: lessons from around the world. *The Lancet* 2012 Jul 21;380(9838):272-281. [doi: [10.1016/S0140-6736\(12\)60816-2](#)] [Medline: [22818939](#)]
5. Global status report on noncommunicable diseases. World Health Organization. 2014. URL: <http://www.who.int/nmh/publications/ncd-status-report-2014/en/> [accessed 2020-10-02]
6. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet* 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: [10.1016/S0140-6736\(12\)61031-9](#)] [Medline: [22818936](#)]
7. Lear SA, Hu W, Rangarajan S, Gasevic D, Leong D, Iqbal R, et al. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. *The Lancet* 2017 Dec 16;390(10113):2643-2654. [doi: [10.1016/S0140-6736\(17\)31634-3](#)] [Medline: [28943267](#)]
8. Bhaskaran K, dos-Santos-Silva I, Leon D, Douglas I, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3·6 million adults in the UK. 2018 Dec 21;6(12):944-953 [FREE Full text] [doi: [10.1016/S2213-8587\(18\)30288-2](#)]
9. Updated Physical Activity Guidelines Now Available. American College of Sports Medicine. 2018. URL: <https://www.acsm.org/read-research/newsroom/news-releases/news-detail/2018/11/12/updated-physical-activity-guidelines-now-available> [accessed 2020-10-02]
10. American Heart Association Recommendations for Physical Activity in Adults and Kids. American Heart Association. 2018. URL: <https://www.heart.org/en/healthy-living/fitness/fitness-basics/aha-recs-for-physical-activity-in-adults> [accessed 2020-10-02]
11. Physical Activity Guidelines for Americans - 2nd edition. US Department of Health and Human Services. 2018. URL: [https://health.gov/paguidelines/second-edition/pdf/Physical\\_Activity\\_Guidelines\\_2nd\\_edition.pdf](https://health.gov/paguidelines/second-edition/pdf/Physical_Activity_Guidelines_2nd_edition.pdf) [accessed 2020-10-02]



12. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, et al. Progress in physical activity over the Olympic quadrennium. *The Lancet* 2016 Sep 24;388(10051):1325-1336. [doi: [10.1016/S0140-6736\(16\)30581-5](https://doi.org/10.1016/S0140-6736(16)30581-5)] [Medline: [27475270](https://pubmed.ncbi.nlm.nih.gov/27475270/)]
13. Global action plan on physical activity 2018–2030: more active people for a healthier world. World Health Organization. 2018. URL: <http://www.who.int/ncds/prevention/physical-activity/global-action-plan-2018-2030/en/> [accessed 2020-10-02]
14. Rodgers A, Woodward A, Swinburn B, Dietz WH. Prevalence trends tell us what did not precipitate the US obesity epidemic. *The Lancet Public Health* 2018 Apr;3(4):e162-e163 [FREE Full text] [doi: [10.1016/S2468-2667\(18\)30021-5](https://doi.org/10.1016/S2468-2667(18)30021-5)] [Medline: [29501260](https://pubmed.ncbi.nlm.nih.gov/29501260/)]
15. Ogden CL, Carroll MD, Fryar CD, Flegal KM. Prevalence of Obesity Among Adults and Youth: United States, 2011–2014. *NCHS Data Brief*. 2015 Nov. URL: [http://www.htuneup.com/diseases/d\\_overweight.pdf](http://www.htuneup.com/diseases/d_overweight.pdf) [accessed 2020-10-02]
16. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. *JMIR Mhealth Uhealth* 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: [10.2196/11819](https://doi.org/10.2196/11819)] [Medline: [30977740](https://pubmed.ncbi.nlm.nih.gov/30977740/)]
17. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis. *Am J Health Promot* 2019 Jun;33(5):778-791. [doi: [10.1177/0890117118816278](https://doi.org/10.1177/0890117118816278)] [Medline: [30586996](https://pubmed.ncbi.nlm.nih.gov/30586996/)]
18. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and Effectiveness of Using Wearable Activity Trackers in Youth: A Systematic Review. *JMIR Mhealth Uhealth* 2016 Nov 23;4(4):e129 [FREE Full text] [doi: [10.2196/mhealth.6540](https://doi.org/10.2196/mhealth.6540)] [Medline: [27881359](https://pubmed.ncbi.nlm.nih.gov/27881359/)]
19. Patients want a heavy dose of digital. Accenture. 2016. URL: <https://www.accenture.com/acnmedia/PDF-8/Accenture-Patients-Want-A-Heavy-Dose-of-Digital-Infographic-v2.pdf> [accessed 2020-10-02]
20. Fitbit Reports \$571M Q4'17 and \$1.616B FY'17 Revenue. Fitbit. 2018. URL: <https://investor.fitbit.com/press/press-releases/press-release-details/2018/Fitbit-Reports-571M-Q417-and-1616B-FY17-Revenue/default.aspx> [accessed 2020-10-02]
21. Pourquoi Fitbit? Fitbit. 2020. URL: <http://www.fitbit.com/fr/whyfitbit> [accessed 2020-10-02]
22. Gartner Says Worldwide Wearable Device Sales to Grow 26 Percent in 2019. Gartner. 2018. URL: <https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow-> [accessed 2020-10-02]
23. Paré G, Leaver C, Bourget C. Diffusion of the Digital Health Self-Tracking Movement in Canada: Results of a National Survey. *J Med Internet Res* 2018 May 02;20(5):e177 [FREE Full text] [doi: [10.2196/jmir.9388](https://doi.org/10.2196/jmir.9388)] [Medline: [29720359](https://pubmed.ncbi.nlm.nih.gov/29720359/)]
24. VanWormer JJ, Pronk NP, Kroeninger GJ. Clinical Counseling for Physical Activity: Translation of a Systematic Review Into Care Recommendations. *Diabetes Spectrum* 2009 Jan 01;22(1):48-55. [doi: [10.2337/diaspect.22.1.48](https://doi.org/10.2337/diaspect.22.1.48)]
25. Improving your health through physical activity. Government of Quebec. 2018. URL: <https://www.quebec.ca/en/health/advice-and-prevention/healthy-lifestyle-habits/physical-activity/improving-your-health-through-physical-activity/> [accessed 2020-10-02]
26. Physical activity and your health. Government of Canada. 2018. URL: <https://www.canada.ca/en/public-health/services/being-active/physical-activity-your-health.html> [accessed 2020-10-02]
27. Hillsdon M, Foster C, Thorogood M. Interventions for promoting physical activity. *Cochrane Database Syst Rev* 2005 Jan 25(1):CD003180 [FREE Full text] [doi: [10.1002/14651858.CD003180.pub2](https://doi.org/10.1002/14651858.CD003180.pub2)] [Medline: [15674903](https://pubmed.ncbi.nlm.nih.gov/15674903/)]
28. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: A systematic review. *Digit Health* 2019;5:2055207619839883. [doi: [10.1177/2055207619839883](https://doi.org/10.1177/2055207619839883)] [Medline: [30944728](https://pubmed.ncbi.nlm.nih.gov/30944728/)]
29. Piwek L, Ellis DA, Andrews S, Joinson A. The Rise of Consumer Health Wearables: Promises and Barriers. *PLoS Med* 2016 Feb;13(2):e1001953 [FREE Full text] [doi: [10.1371/journal.pmed.1001953](https://doi.org/10.1371/journal.pmed.1001953)] [Medline: [26836780](https://pubmed.ncbi.nlm.nih.gov/26836780/)]
30. de Vries HJ, Kooiman TJ, van Ittersum MW, van Brussel M, de Groot M. Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. *Obesity (Silver Spring)* 2016 Oct;24(10):2078-2091 [FREE Full text] [doi: [10.1002/oby.21619](https://doi.org/10.1002/oby.21619)] [Medline: [27670401](https://pubmed.ncbi.nlm.nih.gov/27670401/)]
31. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The Effect of Physical Activity Interventions Comprising Wearables and Smartphone Applications on Physical Activity: a Systematic Review and Meta-analysis. *Sports Med Open* 2018 Sep 03;4(1):42 [FREE Full text] [doi: [10.1186/s40798-018-0157-9](https://doi.org/10.1186/s40798-018-0157-9)] [Medline: [30178072](https://pubmed.ncbi.nlm.nih.gov/30178072/)]
32. Goode AP, Hall KS, Batch BC, Huffman KM, Hastings SN, Allen KD, et al. The Impact of Interventions that Integrate Accelerometers on Physical Activity and Weight Loss: A Systematic Review. *Ann Behav Med* 2017 Feb;51(1):79-93 [FREE Full text] [doi: [10.1007/s12160-016-9829-1](https://doi.org/10.1007/s12160-016-9829-1)] [Medline: [27565168](https://pubmed.ncbi.nlm.nih.gov/27565168/)]
33. Böhm B, Karwiese SD, Böhm H, Oberhoffer R. Effects of Mobile Health Including Wearable Activity Trackers to Increase Physical Activity Outcomes Among Healthy Children and Adolescents: Systematic Review. *JMIR Mhealth Uhealth* 2019 Apr 30;7(4):e8298 [FREE Full text] [doi: [10.2196/mhealth.8298](https://doi.org/10.2196/mhealth.8298)] [Medline: [31038460](https://pubmed.ncbi.nlm.nih.gov/31038460/)]
34. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. *Int J Behav Nutr Phys Act* 2017 Aug 11;14(1):105 [FREE Full text] [doi: [10.1186/s12966-017-0561-4](https://doi.org/10.1186/s12966-017-0561-4)] [Medline: [28800736](https://pubmed.ncbi.nlm.nih.gov/28800736/)]
35. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: A systematic review. *BMC Public Health* 2015 Jun 24;15:585 [FREE Full text] [doi: [10.1186/s12889-015-1947-3](https://doi.org/10.1186/s12889-015-1947-3)] [Medline: [26104189](https://pubmed.ncbi.nlm.nih.gov/26104189/)]



36. Cooper C, Gross A, Brinkman C, Pope R, Allen K, Hastings S, et al. The impact of wearable motion sensing technology on physical activity in older adults. *Exp Gerontol* 2018 Oct 02;112:9-19 [FREE Full text] [doi: [10.1016/j.exger.2018.08.002](https://doi.org/10.1016/j.exger.2018.08.002)] [Medline: [30103025](https://pubmed.ncbi.nlm.nih.gov/30103025/)]
37. Rihoux B, Ragin CC. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, California, USA: SAGE Publications; 2009.
38. Thomas J, O'Mara-Eves A, Brunton G. Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example. *Syst Rev* 2014 Jun 20;3:67 [FREE Full text] [doi: [10.1186/2046-4053-3-67](https://doi.org/10.1186/2046-4053-3-67)] [Medline: [24950727](https://pubmed.ncbi.nlm.nih.gov/24950727/)]
39. An H, Jones GC, Kang S, Welk GJ, Lee J. How valid are wearable physical activity trackers for measuring steps? *Eur J Sport Sci* 2017 Apr;17(3):360-368. [doi: [10.1080/17461391.2016.1255261](https://doi.org/10.1080/17461391.2016.1255261)] [Medline: [27912681](https://pubmed.ncbi.nlm.nih.gov/27912681/)]
40. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act* 2015 Dec 18;12(1):159 [FREE Full text] [doi: [10.1186/s12966-015-0314-1](https://doi.org/10.1186/s12966-015-0314-1)] [Medline: [26684758](https://pubmed.ncbi.nlm.nih.gov/26684758/)]
41. Nelson BW, Allen NB. Accuracy of Consumer Wearable Heart Rate Measurement During an Ecologically Valid 24-Hour Period: Intraindividual Validation Study. *JMIR Mhealth Uhealth* 2019 Mar 11;7(3):e10828 [FREE Full text] [doi: [10.2196/10828](https://doi.org/10.2196/10828)] [Medline: [30855232](https://pubmed.ncbi.nlm.nih.gov/30855232/)]
42. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. *Med Sci Sports Exerc* 2016 Mar 17;48(3):457-465 [FREE Full text] [doi: [10.1249/MSS.0000000000000778](https://doi.org/10.1249/MSS.0000000000000778)] [Medline: [26484953](https://pubmed.ncbi.nlm.nih.gov/26484953/)]
43. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. *JMIR Mhealth Uhealth* 2018 Apr 12;6(4):e94 [FREE Full text] [doi: [10.2196/mhealth.9754](https://doi.org/10.2196/mhealth.9754)] [Medline: [29650506](https://pubmed.ncbi.nlm.nih.gov/29650506/)]
44. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. *JMIR Mhealth Uhealth* 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: [10.2196/10527](https://doi.org/10.2196/10527)] [Medline: [30093371](https://pubmed.ncbi.nlm.nih.gov/30093371/)]
45. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. *J Med Internet Res* 2018 Mar 22;20(3):e110 [FREE Full text] [doi: [10.2196/jmir.9157](https://doi.org/10.2196/jmir.9157)] [Medline: [29567635](https://pubmed.ncbi.nlm.nih.gov/29567635/)]
46. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. *Am J Prev Med* 2015 Sep;49(3):414-418 [FREE Full text] [doi: [10.1016/j.amepre.2015.01.020](https://doi.org/10.1016/j.amepre.2015.01.020)] [Medline: [26071863](https://pubmed.ncbi.nlm.nih.gov/26071863/)]
47. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable Sensor/Device (Fitbit One) and SMS Text-Messaging Prompts to Increase Physical Activity in Overweight and Obese Adults: A Randomized Controlled Trial. *Telemed J E Health* 2015 Oct;21(10):782-792 [FREE Full text] [doi: [10.1089/tmj.2014.0176](https://doi.org/10.1089/tmj.2014.0176)] [Medline: [26431257](https://pubmed.ncbi.nlm.nih.gov/26431257/)]
48. Hartman SJ, Nelson SH, Cadmus-Bertram LA, Patterson RE, Parker BA, Pierce JP. Technology- and Phone-Based Weight Loss Intervention: Pilot RCT in Women at Elevated Breast Cancer Risk. *Am J Prev Med* 2016 Nov;51(5):714-721 [FREE Full text] [doi: [10.1016/j.amepre.2016.06.024](https://doi.org/10.1016/j.amepre.2016.06.024)] [Medline: [27593420](https://pubmed.ncbi.nlm.nih.gov/27593420/)]
49. Griaucze D, Kullgren JT, Liestenfeltz B, Ansari T, Johnson EH, Fedewa A, et al. A Mobile Phone-Based Program to Promote Healthy Behaviors Among Adults With Prediabetes Who Declined Participation in Free Diabetes Prevention Programs: Mixed-Methods Pilot Randomized Controlled Trial. *JMIR Mhealth Uhealth* 2019 Jan 09;7(1):e11267 [FREE Full text] [doi: [10.2196/11267](https://doi.org/10.2196/11267)] [Medline: [30626566](https://pubmed.ncbi.nlm.nih.gov/30626566/)]
50. McCallum C, Rooksby J, Gray CM. Evaluating the Impact of Physical Activity Apps and Wearables: Interdisciplinary Review. *JMIR Mhealth Uhealth* 2018 Mar 23;6(3):e58 [FREE Full text] [doi: [10.2196/mhealth.9054](https://doi.org/10.2196/mhealth.9054)] [Medline: [29572200](https://pubmed.ncbi.nlm.nih.gov/29572200/)]
51. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009 Jul 21;339:b2700 [FREE Full text] [doi: [10.1136/bmj.b2700](https://doi.org/10.1136/bmj.b2700)] [Medline: [19622552](https://pubmed.ncbi.nlm.nih.gov/19622552/)]
52. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
53. RevMan 5. Cochrane Community. URL: <https://community.cochrane.org/help/tools-and-software/revman-5/revman-5-download> [accessed 2020-10-02]
54. Higgins J, Savovic J, Page MJ, Elbers RG, Sterne JA. Chapter 8: Assessing risk of bias in a randomized trial. In: *Cochrane Handbook for Systematic Reviews of interventions*. London: Cochrane; 2019.
55. Hunter JE, Schmidt FL. Fixed Effects vs. Random Effects Meta - Analysis Models: Implications for Cumulative Research Knowledge. *International Journal of Selection and Assessment* 2008 Jun 28;8(4):275-292. [doi: [10.1111/1468-2389.00156](https://doi.org/10.1111/1468-2389.00156)]
56. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Chapter 25: Multiple Comparisons within a Study. In: *Introduction to Meta-Analysis*. Hoboken, New Jersey, USA: Wiley; Mar 11, 2008.

57. Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A Community-Based Physical Activity Counselling Program for People With Knee Osteoarthritis: Feasibility and Preliminary Efficacy of the Track-OA Study. *JMIR Mhealth Uhealth* 2017 Jun 26;5(6):e86 [[FREE Full text](#)] [doi: [10.2196/mhealth.7863](https://doi.org/10.2196/mhealth.7863)] [Medline: [28652228](#)]
58. Götte M, Seidel CC, Kesting SV, Rosenbaum D, Boos J. Objectively measured versus self-reported physical activity in children and adolescents with cancer. *PLoS One* 2017;12(2):e0172216 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0172216](https://doi.org/10.1371/journal.pone.0172216)] [Medline: [28207820](#)]
59. Limb ES, Ahmad S, Cook DG, Kerry SM, Ekelund U, Whincup PH, et al. Measuring change in trials of physical activity interventions: a comparison of self-report questionnaire and accelerometry within the PACE-UP trial. *Int J Behav Nutr Phys Act* 2019 Jan 22;16(1):10 [[FREE Full text](#)] [doi: [10.1186/s12966-018-0762-5](https://doi.org/10.1186/s12966-018-0762-5)] [Medline: [30670036](#)]
60. Beishuizen CR, Stephan BC, van Gool WA, Brayne C, Peters RJ, Andrieu S, et al. Web-Based Interventions Targeting Cardiovascular Risk Factors in Middle-Aged and Older People: A Systematic Review and Meta-Analysis. *J Med Internet Res* 2016 Mar 11;18(3):e55 [[FREE Full text](#)] [doi: [10.2196/jmir.5218](https://doi.org/10.2196/jmir.5218)] [Medline: [26968879](#)]
61. Donoghue K, Patton R, Phillips T, Deluca P, Drummond C. The effectiveness of electronic screening and brief intervention for reducing levels of alcohol consumption: a systematic review and meta-analysis. *J Med Internet Res* 2014 Jun 02;16(6):e142 [[FREE Full text](#)] [doi: [10.2196/jmir.3193](https://doi.org/10.2196/jmir.3193)] [Medline: [24892426](#)]
62. McCullough AR, Ryan C, Macindoe C, Yui N, Bradley JM, O'Neill B, et al. Behavior change theory, content and delivery of interventions to enhance adherence in chronic respiratory disease: A systematic review. *Respir Med* 2016 Jul;116:78-84 [[FREE Full text](#)] [doi: [10.1016/j.rmed.2016.05.021](https://doi.org/10.1016/j.rmed.2016.05.021)] [Medline: [27296825](#)]
63. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011 Jul 22;343:d4002. [doi: [10.1136/bmj.d4002](https://doi.org/10.1136/bmj.d4002)] [Medline: [21784880](#)]
64. Champion KE, Parmenter B, McGowan C, Spring B, Wafford QE, Gardner LA, et al. Effectiveness of school-based eHealth interventions to prevent multiple lifestyle risk behaviours among adolescents: a systematic review and meta-analysis. *The Lancet Digital Health* 2019 Sep;1(5):e206-e221. [doi: [10.1016/s2589-7500\(19\)30088-3](https://doi.org/10.1016/s2589-7500(19)30088-3)]
65. Hayman LL, Worel JN. Healthy lifestyle behaviors: the importance of individual and population approaches. *J Cardiovasc Nurs* 2014;29(6):477-478. [doi: [10.1097/JCN.0000000000000199](https://doi.org/10.1097/JCN.0000000000000199)] [Medline: [25290617](#)]
66. Kelly SA, Melnyk BM, Jacobson DL, O'Haver JA. Correlates among healthy lifestyle cognitive beliefs, healthy lifestyle choices, social support, and healthy behaviors in adolescents: implications for behavioral change strategies and future research. *J Pediatr Health Care* 2011;25(4):216-223. [doi: [10.1016/j.pedhc.2010.03.002](https://doi.org/10.1016/j.pedhc.2010.03.002)] [Medline: [21700136](#)]
67. Amorim AB, Pappas E, Simic M, Ferreira ML, Jennings M, Tiedemann A, et al. Integrating Mobile-health, health coaching, and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot randomised controlled trial. *BMC Musculoskelet Disord* 2019 Feb 11;20(1):71 [[FREE Full text](#)] [doi: [10.1186/s12891-019-2454-y](https://doi.org/10.1186/s12891-019-2454-y)] [Medline: [30744606](#)]
68. Ashe MC, Winters M, Hoppmann CA, Dawes MG, Gardiner PA, Giangregorio LM, et al. "Not just another walking program": Everyday Activity Supports You (EASY) model-a randomized pilot study for a parallel randomized controlled trial. *Pilot Feasibility Stud* 2015 Jan 12;1(1):4 [[FREE Full text](#)] [doi: [10.1186/2055-5784-1-4](https://doi.org/10.1186/2055-5784-1-4)] [Medline: [27175291](#)]
69. Azar KM, Koliwad S, Poon T, Xiao L, Lv N, Griggs R, et al. The Electronic CardioMetabolic Program (eCMP) for Patients With Cardiometabolic Risk: A Randomized Controlled Trial. *J Med Internet Res* 2016 May 27;18(5):e134 [[FREE Full text](#)] [doi: [10.2196/jmir.5143](https://doi.org/10.2196/jmir.5143)] [Medline: [27234480](#)]
70. Ball C, Carter K, Yeung C, Abdelmoneim S, Bauman J, Huang R, et al. Impact of Digital Health Methods for Weight Management on Atherosclerotic Cardiovascular Disease Risk in "at-risk" Women. *Canadian Journal of Cardiology* 2016 Apr;32(4):S9-S10. [doi: [10.1016/j.cjca.2016.02.026](https://doi.org/10.1016/j.cjca.2016.02.026)]
71. Brown JC, Yung RL, Gobbie-Hurder A, Shockro L, O'Connor K, Campbell N, et al. Randomized trial of a clinic-based weight loss intervention in cancer survivors. *J Cancer Surviv* 2018 Apr;12(2):186-195 [[FREE Full text](#)] [doi: [10.1007/s11764-017-0657-5](https://doi.org/10.1007/s11764-017-0657-5)] [Medline: [29101711](#)]
72. Cadmus-Bertram L, Tevaarwerk AJ, Sesto ME, Gangnon R, Van Remortel B, Date P. Building a physical activity intervention into clinical care for breast and colorectal cancer survivors in Wisconsin: a randomized controlled pilot trial. *J Cancer Surviv* 2019 Aug;13(4):593-602 [[FREE Full text](#)] [doi: [10.1007/s11764-019-00778-6](https://doi.org/10.1007/s11764-019-00778-6)] [Medline: [31264183](#)]
73. Cheung NW, Blumenthal C, Smith BJ, Hogan R, Thiagalingam A, Redfern J, et al. A Pilot Randomised Controlled Trial of a Text Messaging Intervention with Customisation Using Linked Data from Wireless Wearable Activity Monitors to Improve Risk Factors Following Gestational Diabetes. *Nutrients* 2019 Mar 11;11(3) [[FREE Full text](#)] [doi: [10.3390/nu11030590](https://doi.org/10.3390/nu11030590)] [Medline: [30862052](#)]
74. Christiansen M, Thoma L, Master H, Mathews D, Schmitt L, Ziegler M, et al. 1-year outcomes from a novel physical therapist-administered physical activity intervention after total knee replacement: a pilot study. *Arthritis Rheumatol* 2018;70:439-440 [[FREE Full text](#)]
75. Christiansen M, Thoma L, Master H, Mathews D, Schmitt L, White D. Preliminary findings of a novel physical therapist administered physical activity intervention after total knee replacement. *Osteoarthritis and Cartilage* 2018 Apr;26:S334. [doi: [10.1016/j.joca.2018.02.663](https://doi.org/10.1016/j.joca.2018.02.663)]

76. Christiansen MB, Thoma LM, Master H, Voinier D, Schmitt LA, Ziegler ML, et al. Feasibility and Preliminary Outcomes of a Physical Therapist-Administered Physical Activity Intervention After Total Knee Replacement. *Arthritis Care Res (Hoboken)* 2020 May;72(5):661-668. [doi: [10.1002/acr.23882](https://doi.org/10.1002/acr.23882)] [Medline: [30908867](https://pubmed.ncbi.nlm.nih.gov/30908867/)]
77. DiFrancisco-Donoghue J, Jung M, Stangle A, Werner WG, Zwibel H, Happel P, et al. Utilizing wearable technology to increase physical activity in future physicians: A randomized trial. *Prev Med Rep* 2018 Dec;12:122-127 [FREE Full text] [doi: [10.1016/j.pmedr.2018.09.004](https://doi.org/10.1016/j.pmedr.2018.09.004)] [Medline: [30234000](https://pubmed.ncbi.nlm.nih.gov/30234000/)]
78. Duscha BD, Piner LW, Patel MP, Craig KP, Brady M, McGarrah RW, et al. Effects of a 12-week mHealth program on peak VO and physical activity patterns after completing cardiac rehabilitation: A randomized controlled trial. *Am Heart J* 2018 May;199:105-114. [doi: [10.1016/j.ahj.2018.02.001](https://doi.org/10.1016/j.ahj.2018.02.001)] [Medline: [29754647](https://pubmed.ncbi.nlm.nih.gov/29754647/)]
79. Eisenberg MH, Phillips LA, Fowler L, Moore PJ. The Impact of E-diaries and Accelerometers on Young Adults' Perceived and Objectively Assessed Physical Activity. *Psychol Sport Exerc* 2017 May;30:55-63 [FREE Full text] [doi: [10.1016/j.psychsport.2017.01.008](https://doi.org/10.1016/j.psychsport.2017.01.008)] [Medline: [28966555](https://pubmed.ncbi.nlm.nih.gov/28966555/)]
80. Falck RS, Best JR, Li LC, Chan PC, Feehan LM, Liu-Ambrose T. Can we improve cognitive function among adults with osteoarthritis by increasing moderate-to-vigorous physical activity and reducing sedentary behaviour? Secondary analysis of the MONITOR-OA study. *BMC Musculoskelet Disord* 2018 Dec 21;19(1):447 [FREE Full text] [doi: [10.1186/s12891-018-2369-z](https://doi.org/10.1186/s12891-018-2369-z)] [Medline: [30577819](https://pubmed.ncbi.nlm.nih.gov/30577819/)]
81. Farnell G, Barkley J. The effect of a wearable physical activity monitor (Fitbit One) on physical activity behaviour in women: A pilot study. *jhse* 2017;12(4). [doi: [10.14198/jhse.2017.124.09](https://doi.org/10.14198/jhse.2017.124.09)]
82. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. 2016 Dec;4(12):983-995. [doi: [10.1016/S2213-8587\(16\)30284-4](https://doi.org/10.1016/S2213-8587(16)30284-4)]
83. Gilmore LA, Klempel MC, Martin CK, Myers CA, Burton JH, Sutton EF, et al. Personalized Mobile Health Intervention for Health and Weight Loss in Postpartum Women Receiving Women, Infants, and Children Benefit: A Randomized Controlled Pilot Study. *J Womens Health (Larchmt)* 2017 Jul;26(7):719-727 [FREE Full text] [doi: [10.1089/jwh.2016.5947](https://doi.org/10.1089/jwh.2016.5947)] [Medline: [28338403](https://pubmed.ncbi.nlm.nih.gov/28338403/)]
84. Hartman SJ, Nelson SH, Myers E, Natarajan L, Sears DD, Palmer BW, et al. Cancer 2018 Jan 01;124(1):192-202 [FREE Full text] [doi: [10.1002/cncr.30987](https://doi.org/10.1002/cncr.30987)] [Medline: [28926676](https://pubmed.ncbi.nlm.nih.gov/28926676/)]
85. Hornikx M, Demeyer H, Camillo CA, Janssens W, Troosters T. The effects of a physical activity counseling program after an exacerbation in patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot study. *BMC Pulm Med* 2015 Nov 04;15(1):136 [FREE Full text] [doi: [10.1186/s12890-015-0126-8](https://doi.org/10.1186/s12890-015-0126-8)] [Medline: [26530543](https://pubmed.ncbi.nlm.nih.gov/26530543/)]
86. H-Jennings F, Clément M, Brown M, Leong B, Shen L, Dong C. Promote Students' Healthy Behavior Through Sensor and Game: A Randomized Controlled Trial. *Med.Sci.Educ* 2016 May 3;26(3):349-355. [doi: [10.1007/s40670-016-0253-8](https://doi.org/10.1007/s40670-016-0253-8)]
87. Katz P, Margaretten M, Gregorich S, Trupin L. Physical Activity to Reduce Fatigue in Rheumatoid Arthritis: A Randomized Controlled Trial. *Arthritis Care Res (Hoboken)* 2018 Jan;70(1):1-10 [FREE Full text] [doi: [10.1002/acr.23230](https://doi.org/10.1002/acr.23230)] [Medline: [28378441](https://pubmed.ncbi.nlm.nih.gov/28378441/)]
88. Kooiman TJ, de Groot M, Hoogenberg K, Krijnen WP, van der Schans CP, Kooy A. Self-tracking of Physical Activity in People With Type 2 Diabetes: A Randomized Controlled Trial. *Comput Inform Nurs* 2018 Jul;36(7):340-349. [doi: [10.1097/CIN.0000000000000443](https://doi.org/10.1097/CIN.0000000000000443)] [Medline: [29742550](https://pubmed.ncbi.nlm.nih.gov/29742550/)]
89. Li LC, Sayre EC, Xie H, Falck RS, Best JR, Liu-Ambrose T, et al. Efficacy of a Community-Based Technology-Enabled Physical Activity Counseling Program for People With Knee Osteoarthritis: Proof-of-Concept Study. *J Med Internet Res* 2018 Apr 30;20(4):e159 [FREE Full text] [doi: [10.2196/jmir.8514](https://doi.org/10.2196/jmir.8514)] [Medline: [29712630](https://pubmed.ncbi.nlm.nih.gov/29712630/)]
90. Lystrup RM, West GF, Olsen C, Ward M, Stephens MB. Pedometry to Prevent Cardiorespiratory Fitness Decline-Is it Effective? *Mil Med* 2016 Oct;181(10):1235-1239. [doi: [10.7205/MILMED-D-15-00540](https://doi.org/10.7205/MILMED-D-15-00540)] [Medline: [27753558](https://pubmed.ncbi.nlm.nih.gov/27753558/)]
91. Mahar MT, Nanney LW, Das BM, Raedeke TD, Vick GA, Rowe DA. Effects of an Intervention using Movement Technology in a University Physical Activity Class. 2015;47:522. [doi: [10.1249/01.mss.0000477869.71815.de](https://doi.org/10.1249/01.mss.0000477869.71815.de)]
92. McDermott MM, Spring B, Berger JS, Treat-Jacobson D, Conte MS, Creager MA, et al. Effect of a Home-Based Exercise Intervention of Wearable Technology and Telephone Coaching on Walking Performance in Peripheral Artery Disease: The HONOR Randomized Clinical Trial. *JAMA* 2018 Apr 24;319(16):1665-1676 [FREE Full text] [doi: [10.1001/jama.2018.3275](https://doi.org/10.1001/jama.2018.3275)] [Medline: [29710165](https://pubmed.ncbi.nlm.nih.gov/29710165/)]
93. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: A pilot study. *Pediatr Blood Cancer* 2017 Dec 15;64(12):e26660. [doi: [10.1002/pbc.26660](https://doi.org/10.1002/pbc.26660)] [Medline: [28618158](https://pubmed.ncbi.nlm.nih.gov/28618158/)]
94. Miragall M, Domínguez-Rodríguez A, Navarro J, Cebolla A, Baños RM. Increasing physical activity through an Internet-based motivational intervention supported by pedometers in a sample of sedentary students: A randomised controlled trial. *Psychol Health* 2018 Apr;33(4):465-482. [doi: [10.1080/08870446.2017.1368511](https://doi.org/10.1080/08870446.2017.1368511)] [Medline: [28880576](https://pubmed.ncbi.nlm.nih.gov/28880576/)]
95. Oliveira JS, Sherrington C, Paul SS, Ramsay E, Chamberlain K, Kirkham C, et al. A combined physical activity and fall prevention intervention improved mobility-related goal attainment but not physical activity in older adults: a randomised trial. *J Physiother* 2019 Jan;65(1):16-22 [FREE Full text] [doi: [10.1016/j.jphys.2018.11.005](https://doi.org/10.1016/j.jphys.2018.11.005)] [Medline: [30581138](https://pubmed.ncbi.nlm.nih.gov/30581138/)]



96. Paxton RJ, Forster JE, Miller MJ, Gerron KL, Stevens-Lapsley JE, Christiansen CL. A Feasibility Study for Improved Physical Activity After Total Knee Arthroplasty. *J Aging Phys Act* 2018 Jan 01;26(1):7-13 [[FREE Full text](#)] [doi: [10.1123/japa.2016-0268](#)] [Medline: [28338406](#)]
97. Redman LM, Gilmore LA, Breaux J, Thomas DM, Elkind-Hirsch K, Stewart T, et al. Effectiveness of SmartMoms, a Novel eHealth Intervention for Management of Gestational Weight Gain: Randomized Controlled Pilot Trial. *JMIR Mhealth Uhealth* 2017 Sep 13;5(9):e133 [[FREE Full text](#)] [doi: [10.2196/mhealth.8228](#)] [Medline: [28903892](#)]
98. Shoemaker MJ, Oberholtzer NL, Jongekrijg LE, Bowen TE, Cartwright K, Hanson K, et al. Exercise- and Psychosocial-Based Interventions to Improve Daily Activity in Heart Failure: A Pilot Study. 2016 Dec 21;29(2):111-120. [doi: [10.1177/1084822316683660](#)]
99. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. Effect and Process Evaluation of a Smartphone App to Promote an Active Lifestyle in Lower Educated Working Young Adults: Cluster Randomized Controlled Trial. *JMIR Mhealth Uhealth* 2018 Aug 24;6(8):e10003 [[FREE Full text](#)] [doi: [10.2196/10003](#)] [Medline: [30143477](#)]
100. Sloan RA, Kim Y, Sahasranaman A, Müller-Riemenschneider F, Biddle SJ, Finkelstein EA. The influence of a consumer-wearable activity tracker on sedentary time and prolonged sedentary bouts: secondary analysis of a randomized controlled trial. *BMC Res Notes* 2018 Mar 22;11(1):189 [[FREE Full text](#)] [doi: [10.1186/s13104-018-3306-9](#)] [Medline: [29566746](#)]
101. Thompson WG, Kuhle CL, Koepp GA, McCrady-Spitzer SK, Levine JA. "Go4Life" exercise counseling, accelerometer feedback, and activity levels in older people. *Arch Gerontol Geriatr* 2014;58(3):314-319. [doi: [10.1016/j.archger.2014.01.004](#)] [Medline: [24485546](#)]
102. Thorndike AN, Mills S, Sonnenberg L, Palakshappa D, Gao T, Pau CT, et al. Activity monitor intervention to promote physical activity of physicians-in-training: randomized controlled trial. *PLoS One* 2014;9(6):e100251 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0100251](#)] [Medline: [24950218](#)]
103. Van Blarigan EL, Chan H, Van Loon K, Kenfield SA, Chan JM, Mitchell E, et al. Self-monitoring and reminder text messages to increase physical activity in colorectal cancer survivors (Smart Pace): a pilot randomized controlled trial. *BMC Cancer* 2019 Mar 11;19(1):218 [[FREE Full text](#)] [doi: [10.1186/s12885-019-5427-5](#)] [Medline: [30866859](#)]
104. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The Effectiveness of a Web-Based Computer-Tailored Physical Activity Intervention Using Fitbit Activity Trackers: Randomized Trial. *J Med Internet Res* 2018 Dec 18;20(12):e11321 [[FREE Full text](#)] [doi: [10.2196/11321](#)] [Medline: [30563808](#)]
105. Ragin C, Fiss P. Net effects versus configurations: An empirical demonstration. In: *Redesigning Social Inquiry – Fuzzy Sets and Beyond*. Chicago: University of Chicago Press; 2008:190-219.
106. Ragin C. Calibrating Fuzzy Sets. In: *Redesigning Social Inquiry – Fuzzy Sets and Beyond*. Chicago: University of Chicago Press; 2008.
107. Davergne T, Pallot A, Dechartres A, Fautrel B, Gossec L. Use of Wearable Activity Trackers to Improve Physical Activity Behavior in Patients With Rheumatic and Musculoskeletal Diseases: A Systematic Review and Meta-Analysis. *Arthritis Care Res (Hoboken)* 2019 Jun 23;71(6):758-767 [[FREE Full text](#)] [doi: [10.1002/acr.23752](#)] [Medline: [30221489](#)]
108. Cheatham SW, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. *J Sports Med Phys Fitness* 2018 Apr;58(4):534-548. [doi: [10.23736/S0022-4707.17.07437-0](#)] [Medline: [28488834](#)]
109. Deeks J, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: *Cochrane Handbook for Systematic Reviews of Interventions*. London: Cochrane; 2019.
110. Arroggi A, Schotte A, Bogaerts A, Boen F, Seghers J. Short- and long-term effectiveness of a three-month individualized need-supportive physical activity counseling intervention at the workplace. *BMC Public Health* 2017 Jan 09;17(1):52 [[FREE Full text](#)] [doi: [10.1186/s12889-016-3965-1](#)] [Medline: [28069016](#)]
111. Ludwig K, Arthur R, Sculthorpe N, Fountain H, Buchan DS. Text Messaging Interventions for Improvement in Physical Activity and Sedentary Behavior in Youth: Systematic Review. *JMIR Mhealth Uhealth* 2018 Sep 17;6(9):e10799 [[FREE Full text](#)] [doi: [10.2196/10799](#)] [Medline: [30224335](#)]
112. Proper KI, Hildebrandt VH, Van der Beek AJ, Twisk JW, Van Mechelen W. Effect of individual counseling on physical activity fitness and health. *American Journal of Preventive Medicine* 2003 Apr;24(3):218-226. [doi: [10.1016/s0749-3797\(02\)00645-1](#)]
113. Hall KD, Kahan S. Maintenance of Lost Weight and Long-Term Management of Obesity. *Med Clin North Am* 2018 Jan;102(1):183-197 [[FREE Full text](#)] [doi: [10.1016/j.mcna.2017.08.012](#)] [Medline: [29156185](#)]
114. Savoli A, Barki H, Pare G. Examining How Chronically Ill Patients' Reactions to and Effective Use of Information Technology Can Influence How Well They Self-Manage Their Illness. *MISQ* 2020 Jan 01;44(1):351-389. [doi: [10.25300/misq/2020/15103](#)]
115. Paré G, Tate M, Johnstone D, Kitsiou S. Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. *European Journal of Information Systems* 2017 Dec 19;25(6):493-508. [doi: [10.1057/s41303-016-0020-3](#)]

116. Templier M, Paré G. Transparency in literature reviews: an assessment of reporting practices across review types and genres in top IS journals. *European Journal of Information Systems* 2017 Nov 30;27(5):503-550. [doi: [10.1080/0960085x.2017.1398880](https://doi.org/10.1080/0960085x.2017.1398880)]
117. Sutcliffe K, Kneale D, Thomas J. How is Qualitative Comparative Analysis (QCA) used in systematic reviews to identify critical intervention components? *Cochrane*. 2019. URL: <https://abstracts.cochrane.org/sites/default/files/attachments/pdf/poster-7549.pdf> [accessed 2020-10-02]
118. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. *J Med Internet Res* 2014 Aug 15;16(8):e192 [FREE Full text] [doi: [10.2196/jmir.3469](https://doi.org/10.2196/jmir.3469)] [Medline: [25131661](https://pubmed.ncbi.nlm.nih.gov/25131661/)]
119. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior Change Techniques Present in Wearable Activity Trackers: A Critical Analysis. *JMIR Mhealth Uhealth* 2016 Apr 27;4(2):e40 [FREE Full text] [doi: [10.2196/mhealth.4461](https://doi.org/10.2196/mhealth.4461)] [Medline: [27122452](https://pubmed.ncbi.nlm.nih.gov/27122452/)]
120. James TL, Wallace L, Deane JK. Using Organismic Integration Theory to Explore the Associations Between Users' Exercise Motivations and Fitness Technology Feature Set Use. *MISQ* 2019 Jan 1;43(1):287-312. [doi: [10.25300/misq/2019/14128](https://doi.org/10.25300/misq/2019/14128)]

## Abbreviations

**FsQCA:** fuzzy-set qualitative comparative analysis

**MD:** mean difference

**MET:** metabolic equivalent

**MVPA:** moderate-to-vigorous physical activity

**PA:** physical activity

**RCT:** randomized controlled trial

**SMD:** standardized mean difference

*Edited by G Eysenbach; submitted 29.08.20; peer-reviewed by M Johansson, H Sasai; comments to author 04.09.20; revised version received 24.09.20; accepted 27.09.20; published 12.10.20.*

*Please cite as:*

Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S

Fitbit-Based Interventions for Healthy Lifestyle Outcomes: Systematic Review and Meta-Analysis

*J Med Internet Res* 2020;22(10):e23954

URL: <http://www.jmir.org/2020/10/e23954/>

doi: [10.2196/23954](https://doi.org/10.2196/23954)

PMID: [33044175](https://pubmed.ncbi.nlm.nih.gov/33044175/)

©Mickael Ringeval, Gerit Wagner, James Denford, Guy Paré, Spyros Kitsiou. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 12.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Review

# Tailored Web-Based Smoking Interventions and Reduced Attrition: Systematic Review and Meta-Analysis

Amika Shah<sup>1</sup>, MScCH; Michael Chaiton<sup>1,2</sup>, PhD; Dolly Baliunas<sup>1,3</sup>, PhD; Robert Schwartz<sup>1,2</sup>, PhD

<sup>1</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

<sup>2</sup>Ontario Tobacco Research Unit, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

<sup>3</sup>Office of Education, Centre for Addiction and Mental Health, Toronto, ON, Canada

**Corresponding Author:**

Michael Chaiton, PhD

Dalla Lana School of Public Health

University of Toronto

155 College St

Toronto, ON, M5S 2S1

Canada

Phone: 1 416 978 7096

Email: [michael.chaiton@camh.ca](mailto:michael.chaiton@camh.ca)

## Abstract

**Background:** The increasing number of internet users presents an opportunity to deliver health interventions to large populations. Despite their potential, many web-based interventions, including those for smoking cessation, face high rates of attrition. Further consideration of how intervention features impact attrition is needed.

**Objective:** The aim of this systematic review is to investigate whether tailored web-based smoking cessation interventions for smokers are associated with reduced rates of attrition compared with active or passive untailored web-based interventions. The outcomes of interest were dropout attrition at 1-, 3-, 6-, and 12-month follow-ups.

**Methods:** Literature searches were conducted in May 2018 and updated in May 2020 on MEDLINE (Medical Literature Analysis and Retrieval System Online), PsycINFO (Psychological Information), EMBASE (Excerpta Medica dataBASE), CINAHL (Cumulated Index to Nursing and Allied Health Literature), Scopus, and the Cochrane Tobacco Addiction Group Specialized Register with the following search terms: smoking cessation, tailored, or web- or internet-based. Included studies were published in English before or in May 2020 using a randomized controlled trial design. Studies were restricted to those with web-based delivery, a tailored intervention group, an untailored control group, and a reported outcome of smoking cessation. Studies were assessed for methodological quality using the Cochrane Risk of Bias tool. Two reviewers independently extracted the study characteristics and the number of participants lost to follow-up for each treatment group.

**Results:** A total of 13 studies were included in the systematic review, of which 11 (85%) were included in the meta-analysis. Tailoring had no statistically significant effect on dropout attrition at 1-month (risk ratio [RR]=1.02, 95% CI 0.95-1.09;  $P=.58$ ;  $I^2=78\%$ ), 3-month (RR=0.99, 95% CI 0.95-1.04;  $P=.80$ ;  $I^2=73\%$ ), 6-month (RR=1.00, 95% CI 0.95-1.05;  $P=.90$ ;  $I^2=43\%$ ), or 12-month (RR=0.97, 95% CI 0.92-1.02;  $P=.26$ ;  $I^2=28\%$ ) follow-ups. Subgroup analyses suggested that there was a statistically significant effect of tailoring between the active and passive subgroups at 1-month ( $P=.03$ ), 3-month ( $P<.001$ ), and 6-month ( $P=.02$ ) follow-ups but not at 12-month follow-up ( $P=.25$ ).

**Conclusions:** The results suggest that tailoring of web-based smoking cessation interventions may not be associated with reduced rates of dropout attrition at 1-, 3-, 6-, or 12-month follow-ups. Significant differences between studies that include untailored active and passive control groups suggest that the role of tailoring may be more prominent when studies include a passive control group. These findings may be because of variability in the presence of additional features, the definition of smokers used, and the duration of smoking abstinence measured. Future studies should incorporate active web-based controls, compare the impact of different tailoring strategies, and include populations outside of the Western countries.

(*J Med Internet Res* 2020;22(10):e16255) doi:[10.2196/16255](https://doi.org/10.2196/16255)

**KEYWORDS**

internet; world wide web; smoking cessation; web-based intervention

## Introduction

### Background

Smoking is the leading cause of preventable death [1]. Annually, over 7 million deaths worldwide are attributed to cigarette smoking [2], leading to a global loss of 150 million disability-adjusted life years [3]. The majority of smokers want to quit [4]; however, it can take 30 or more attempts to successfully quit [5]. Although smoking cessation is difficult, the price of not quitting is high—over 50% of long-term smokers die from smoking [6,7]. Fortunately, evidence-based smoking cessation interventions can double or triple the chances that a quit attempt will result in long-term cessation [8]. However, most effective clinical interventions have been found to be more costly and to have lower reach compared with public health interventions [9].

### Web-Based Interventions

Interventions delivered over the internet, termed *web-based interventions*, have the potential to increase the public health impact of smoking cessation interventions because of their widespread use and scalability [10]. Despite their potential, few web-based interventions succeed in delivering the full treatment intended because of high rates of attrition. In a seminal paper, Eysenbach [11] delineated 2 types of attrition in web-based interventions: nonusage and dropout. The former refers to users who do not use the web-based intervention, whereas the latter refers to users who do not complete the follow-up study procedures. Eysenbach [11] hypothesized that both types of attrition are related and may be explained by a common experience of the user *losing interest*. As such, research that contributes to the *science of attrition* is recommended to better understand the phenomenon of attrition in the context of trials of web-based interventions [11].

Like many web-based interventions [11], web-based smoking cessation interventions often report high rates of dropout attrition [12], posing threats to the validity of evidence surrounding these interventions. Although studies on smoking cessation often employ intention-to-treat analyses, where those lost to follow-up are assumed to be smokers, high rates of dropout attrition increase the risk of attrition bias, which may lead to the underestimation of the effectiveness of web-based smoking cessation interventions. Such findings have been reported by systematic reviews of web-based smoking cessation interventions that have identified attrition bias as a critical challenge in assessing the effectiveness of these interventions [12–14]. In addition, bias caused by attrition makes it difficult to identify features of web-based smoking cessation interventions that are most effective in promoting smoking cessation. Thus, further consideration of the features that influence dropout attrition is needed to fully understand the impact of web-based smoking cessation interventions, the features that characterize effective interventions, and their mechanisms of action.

### Tailoring

Tailoring is one feature that has received significant interest [13]. Tailored print smoking cessation materials have been

found to be more effective than untailored materials [15,16], suggesting that this feature may be important for their web-based equivalents. Likewise, systematic reviews of web-based smoking cessation interventions have highlighted the importance of tailoring in promoting smoking cessation [12–14]. For instance, Taylor [12] investigated the effectiveness of interactive, tailored, and combined (tailored and interactive) web-based interventions compared with nonactive and active control interventions. Combined interactive and tailored interventions were found to be moderately more effective than nonactive controls (eg, printed self-help guide) but not more effective than active controls (eg, counseling sessions) [12]. The authors noted that many of the studies had high rates of attrition, making it difficult to assess the effectiveness of tailored web-based smoking cessation interventions [12]. Thus, the ongoing challenge of high dropout attrition presents a need to explore the effect of intervention design features, such as tailoring, on attrition [15,17].

### Objectives

This study aims to investigate across randomized controlled trials (RCTs), whether tailored active web-based smoking cessation interventions reduce dropout attrition at 1-, 3-, 6-, and 12-month follow-ups compared with an untailored active or passive control. In the context of this review, we define the terms *tailoring*, *active*, and *passive* as follows.

### Tailoring

The term *tailoring* is often used interchangeably with *personalization* and *targeting*; however, important differences exist among these 3 terms [17]. Personalization refers to materials that have been customized using an individual's name, whereas targeting involves designing materials for a particular subgroup or population with one or more shared characteristics (eg, youth) [17]. In contrast, tailored materials refer to materials that are designed for the characteristics of a particular individual based on individual assessment (eg, quit date, level of motivation, and self-efficacy) [17]. These terms may be viewed as existing on a continuum ranging from generic to individually tailored [17], with each term varying in the level in which they consider and incorporate characteristics of the user into the intervention. Although we recognize the differences among these terms, for simplicity, we operationalize the term *tailored* as an umbrella term that encompasses personalization, targeting, and tailoring to the individual.

### Active

The term *active* is defined as interventions or control groups that require more than one engagement by the user. The multiple engagements must be part of the intervention and not simply part of the study procedures. For instance, a study with a single tailored email that requires multiple follow-up procedures would not be considered active in the context of this review. The authors chose to focus on active interventions, as attrition is more likely to be an issue for interventions and studies that require multiple engagements over time, rather than a single engagement.

### Passive

The term passive is defined as control groups that require a single engagement by the user. The authors chose to compare studies with either an active or passive control group to assess whether the effect of tailoring differs between studies that compare active tailored interventions to control groups that require similar (active) or different (passive) levels of engagement.

## Methods

### Protocol and Registration

This review is designed and reported in line with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) reporting guidelines (for checklist, see [Multimedia Appendix 1](#)) [18]. A protocol was developed to guide this review; however, the protocol was not registered. Details outlined in the protocol included the study rationale, research question, eligibility criteria, study selection process, outcomes of interest, and the processes followed in conducting the meta-analysis. No deviations were made from the protocol from start to the final review [18].

### Data Source and Search Strategy

Searches were conducted using the following electronic databases: MEDLINE (Medical Literature Analysis and Retrieval System Online), PsycINFO (Psychological Information), EMBASE (Excerpta Medica dataBASE), CINAHL (Cumulated Index to Nursing and Allied Health Literature), Scopus, and the Cochrane Tobacco Addiction Group Specialized Register. Search strategies for each database were created using synonyms of the 4 main search concepts: (1) *smoking cessation* (outcome measure), (2) *tailoring* (intervention feature), (3) *web- or internet-based* (technology), and (4) *randomized control trial* (study design). A sample search strategy for MEDLINE is provided in [Multimedia Appendix 2](#). An initial search was conducted in May 2018 for studies published in or before May 2018. The review was later updated in May 2020 to identify

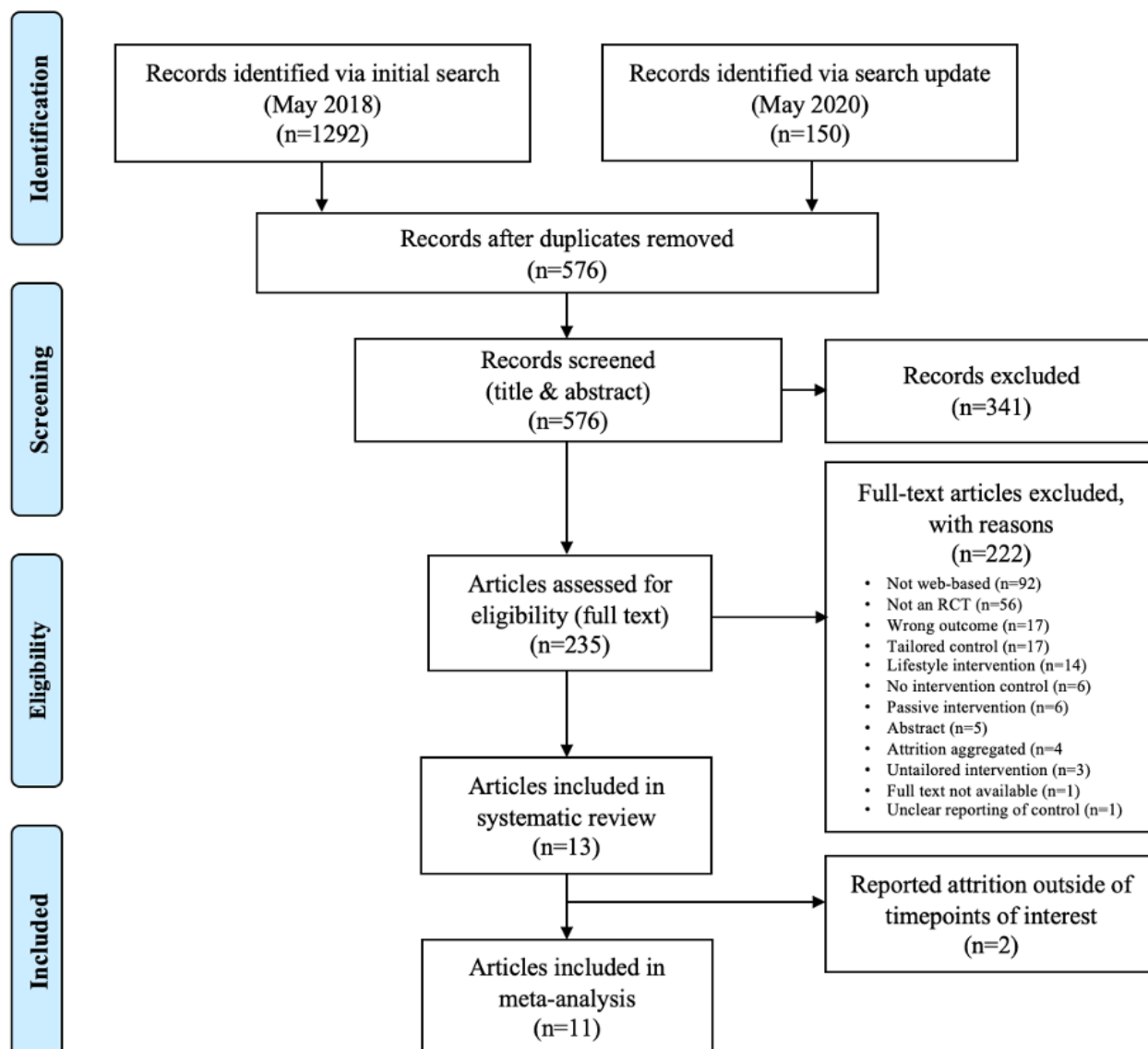
any studies published between May 2018 and May 2020. Search results were exported and managed on EndNote Web.

### Study Selection

Two authors (AS and MC) independently reviewed papers retrieved from the search strategy by title and abstract. Rayyan, a web-based application for systematic reviews, was used to facilitate the screening process between the 2 reviewers. The inclusion criteria were that studies (1) employed any type of randomized controlled trial (RCT) design (eg, crossover trials, parallel trials, and factorial trials), (2) included a tailored web-based smoking cessation intervention, (3) had an untailored control intervention (both web-based or nonweb-based were accepted), and (4) assessed smoking cessation through any method (ie, point prevalence estimate, self-report, and biochemical validation) at least 1 month after the start of the intervention. Papers could be published in or before May 2020. Only peer-reviewed articles published in English with a study population of smokers were included in the analysis. Dissertations, poster abstracts, and studies that described lifestyle interventions (interventions that targeted multiple health behaviors) were excluded ([Figure 1](#)).

Studies were selected based on the tailored intervention component. If the tailored component was not delivered online (ie, a computer-tailored letter that was printed and mailed to the participant), the study was excluded. Multicomponent interventions were included; however, if the tailored component was not delivered on the web (eg, tailored text messages accompanied by a web page), the study was deemed ineligible. These criteria were used to isolate the impact of tailoring and increase the comparability of the interventions included in the meta-analysis.

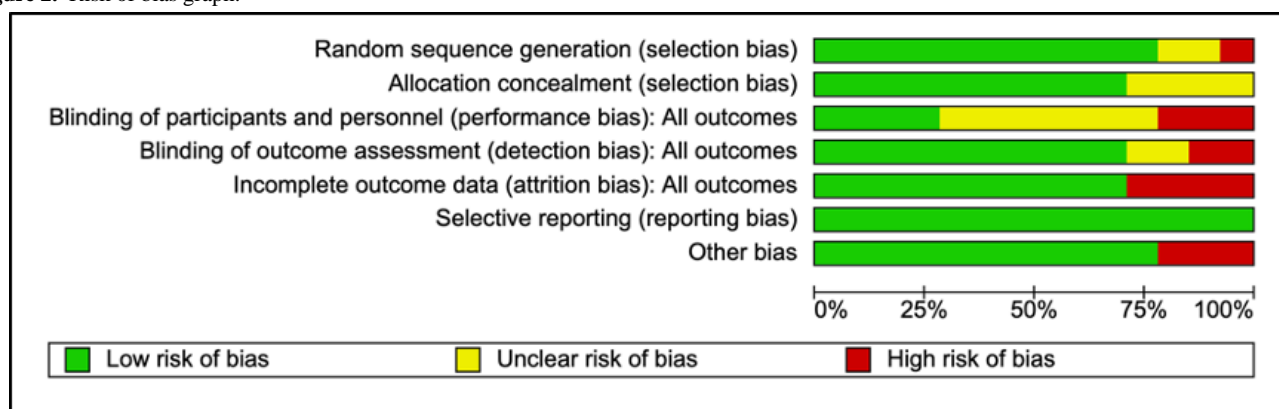
Studies that met the exclusion criteria when reviewed by title and abstract were excluded. The remaining studies were moved on to the next stage, where they were assessed using the full text. Full-text assessments of studies were later conducted for eligibility, with reasons for exclusion recorded. Disagreements in study eligibility were resolved through discussions between the two reviewers (AS and MC).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) diagram.

## Study Quality

Studies were assessed for methodological quality using the Cochrane Risk of Bias Tool [19], with reasons for risk level

recorded (Figure 2). Risk of bias assessment was conducted at the study level. Risk of bias assessments were considered during the data interpretation stage.

**Figure 2.** Risk of bias graph.



## Data Extraction and Analysis

The following details were extracted from the included studies: author, year of publication, study location, sample, population characteristics, intervention, theoretical framework used, tailored components, control group used, follow-up period, and key findings. In addition, the reported number of individuals lost to follow-up for the intervention and control arms were extracted from each study and compared against the original sample of individuals allocated to the intervention and control arms. If missing data were related to the outcome of interest (dropout attrition), the authors were contacted for additional information. Risk ratios (RRs) and 95% CIs were used to measure the difference in attrition between tailored interventions and untailored control groups. All significance tests were 2-tailed, with  $P < .05$  considered statistically significant. A meta-analysis was conducted with the software RevMan 5 (Cochrane Organization) using the Mantel–Haenszel fixed effects model to pool the RRs. A fixed effects model was chosen as the most conservative model with the expectation that there would be some heterogeneity in design. The  $I^2$  statistic was used to assess statistical heterogeneity at 3 upper limits (low,  $I^2 < 25\%$ ; moderate,  $I^2$  between 25% and 75%; high,  $I^2 > 75\%$ ) [20]. A sensitivity analysis was conducted to examine the effect of potential outliers on the meta-analysis results.

## Results

### Search Results

The results of the search strategy are summarized in the PRISMA flowchart shown in Figure 1. When the first search was conducted in May 2018, the search strategy yielded a total of 1292 studies, of which 801 (62.0%) were duplicates. Upon screening the remaining 491 studies by title and abstract, 274 (55.8%) studies were excluded. The remaining 217 studies were later screened by the full text, and 204 (94.0%) studies were excluded for not meeting the inclusion criteria. From this initial search, 13 studies were included in the systematic review [21–33]. The search update conducted in May 2020 identified a total of 150 studies. After removing duplicates, 56.7% (85/150) of the studies were screened by title and abstract against the inclusion criteria, and 79% (67/85) studies were subsequently excluded. The remaining 21% (18/85) of the studies were screened by their full texts. Within these 18 studies, 2 (11%) appeared to meet the inclusion criteria, although neither study could be included in the analysis. At the time of the search update, the full text of Kahler [34] was not available, and thus, the details required for the meta-analysis could not be extracted. Additionally, due to unclear details regarding the control intervention, we excluded the study by Altendorf et al [35].

## Study Characteristics

Details of the included studies, such as the author, year of publication, study location, sample, population characteristics, intervention, theoretical framework used, tailored components, control, and key findings, are outlined in Multimedia Appendix 3. The included studies were published from 2008 to 2018 in the following countries: the United States [21,24,26,27,32,33], Switzerland [22], Norway [23], Australia [25], France [28], Spain [29], Denmark [30], and the Netherlands [31]. Of the 13 included studies, 4 (31%) studies used an active control intervention, whereas 9 (69%) studies used a passive control intervention.

The 14 trials included a total of 12,661 participants: 6538 (51.64%) randomized to a tailored intervention and 6123 (48.63%) to an untailored control intervention. Participants were recruited from a variety of settings, including postsecondary institutions [24,29], hospitals or clinics [26,27,31,32], a Quitline [24], and from a website [21,22,28,33]. Follow-up periods ranged from 1 to 18 months. Five studies reported follow-up at 1 month [23,25,30,32,33], 6 studies at 3 months [21–23,26,28,33], 9 studies at 6 months [21,22,26–31,33], 6 studies at 12 months [21,22,26–31,33], and 1 study at 18 months [21]. One study reported outcomes at 8, 20, and 30 weeks [26] and another at 7 months [27]. Of the 13 studies included in the systematic review, 7 (54%) studies [23–26,28,29,33] reported significant differences in smoking cessation between the tailored intervention and untailored control at one or more follow-up periods.

## Study Quality

Most studies included in the analysis were found to have an unclear or high risk of performance bias (Figure 2). Moreover, 4 studies were found to be at high risk of attrition bias as characterized by studies with more than 50% of the sample randomized to a condition lost to attrition.

## Attrition

All the 13 included studies reported dropout attrition, whereas only 1 (8%) study [25] reported nonusage attrition. Dropout attrition across the 13 studies ranged from 5% to 67% for tailored interventions and from 3% to 64% for untailored interventions (Table 1). Among the 13 studies that met the inclusion criteria, 2 (15%) studies [24,25] were excluded from the meta-analysis, as they reported dropout attrition at timepoints that could not be compared with any other study included in the analysis. Although Borland [25] included follow-ups at 1 and 7 months, attrition was not reported for the 1-month follow-up, and thus, this study was excluded from the meta-analysis. The exclusion of these 2 studies resulted in a total of 11 studies that were included in the meta-analysis. The results of the meta-analysis with respect to the follow-up periods of interest (1, 3, 6, and 12 months) are described in the following section.



**Table 1.** Summary of attrition (n=13).

1st author, reference	Attrition type reported	Follow-up period	Intervention sample, n	Intervention loss to follow-up, n (%)	Control sample, n	Control loss to follow-up, n (%)
<b>An [24]</b>						
	Dropout	8 weeks	257	13 (5.1)	260	8 (3.1)
	Dropout	20 weeks	257	24 (9.3)	260	20 (7.7)
	Dropout	30 weeks	257	23 (8.9)	260	30 (11.5)
<b>Borland [25]</b>						
	Dropout and nonusage	1 month	809	Not reported	422	Not reported
	Dropout and nonusage	7 months	809	104 (12.9)	422	66 (15.6)
<b>Das [26]</b>						
	Dropout	3 months	105	16 (15.2)	111	20 (18.0)
	Dropout	6 months	105	13 (12.4)	111	15 (13.5)
	Dropout	12 months	105	13 (12.4)	111	14 (12.6)
<b>Graham [21]</b>						
	Dropout	3 months	651	151 (23.2)	679	142 (20.9)
	Dropout	6 months	651	168 (25.8)	679	154 (22.7)
	Dropout	12 months	651	180 (27.7)	679	187 (27.5)
	Dropout	18 months	651	201 (30.9)	679	213 (31.4)
<b>Harrington [27]</b>	Dropout	6 months	748	98 (13.1)	740	89 (12.0)
<b>Mavrot [22]</b>						
	Dropout	3 months	580	290 (50.0)	580	251 (43.3)
	Dropout	6 months	580	353 (60.9)	580	331 (57.1)
<b>Nguyen [28]</b>						
	Dropout	3 months	1242	639 (51.4)	1236	720 (58.3)
	Dropout	6 months	1242	667 (53.7)	1236	729 (59.0)
	Dropout	12 months	1242	732 (58.9)	1236	793 (64.2)
<b>Pardavila-Belio [29]</b>	Dropout	6 months	133	19 (14.3)	122	11 (9.0)
<b>Skov-Ettrup [30]</b>						
	Dropout	1 month	453	106 (23.4)	452	64 (14.2)
	Dropout	6 months	453	67 (14.8)	452	75 (16.6)
	Dropout	12 months	453	84 (18.5)	452	66 (14.6)
<b>Smit [31]</b>						
	Dropout	6 months	132	89 (67.4)	119	74 (62.2)
	Dropout	12 months	132	57 (43.2)	119	55 (46.2)
<b>Tsoh [32]</b>						
	Dropout	1 month	23	5 (21.7)	19	5 (26.3)
	Dropout	2 months	23	9 (39.1)	19	6 (31.6)
<b>Wangberg [23]</b>						
	Dropout	1 month	1029	613 (59.6)	1043	640 (64.1)
	Dropout	3 months	1029	648 (63.0)	1043	644 (61.7)
	Dropout	12 months	1029	303 (29.4)	1043	300 (28.8)
<b>Westmaas [33]</b>						
	Dropout	1 month	376	98 (26.1)	340	95 (27.9)

1st author, reference	Attrition type reported	Follow-up period	Intervention sample, n	Intervention loss to follow-up, n (%)	Control sample, n	Control loss to follow-up, n (%)
	Dropout	3 months	376	134 (35.6)	340	112 (32.9)
	Dropout	6 months	376	156 (41.5)	340	136 (40.0)

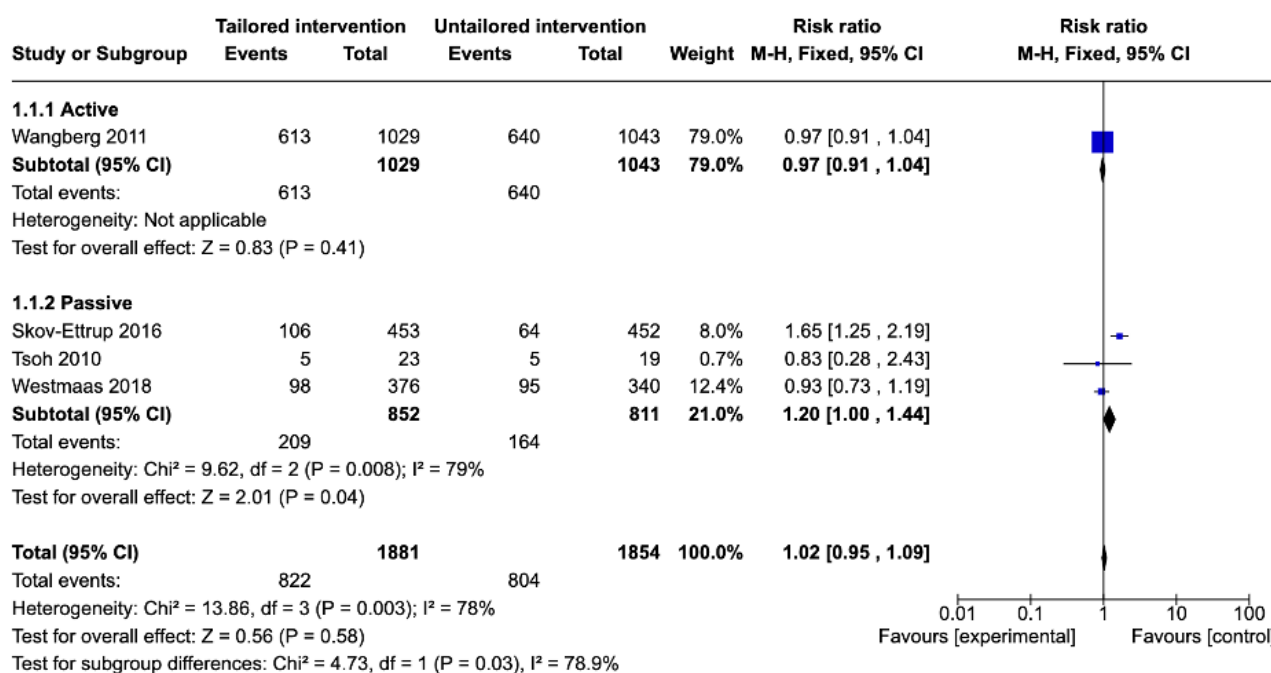
### One-Month Follow-Up

Of the 11 studies, 6 (55%) studies were included in the meta-analysis and 4 (36%) studies [23,30,32,33] reported attrition at 1-month follow-up. Only 1 (9%) study [23] compared a tailored active intervention with an untailored active control. There were no significant differences in the risk of attrition between tailored interventions and untailored controls at 1-month follow-up (RR=0.97, 95% CI 0.91-1.04;  $P=.41$ ). Across the 3 studies [30,32,33] that compared a tailored active intervention with an untailored passive control, the risk of attrition was higher among tailored interventions at 1-month follow-up (RR=1.20, 95% CI 1.00-1.44;  $P=.04$ ). This estimate

was statistically significant, although it had high heterogeneity ( $I^2=79\%$ ). The test for subgroup differences indicated significant differences between studies that included an active control and those that included a passive control at 1-month follow-up ( $\chi^2_1=4.7$ ;  $P=.03$ ).

When the 4 studies [23,30,32,33] were pooled, no significant differences in the risk of attrition between tailored interventions and untailored controls were found at 1-month follow-up (RR=1.02, 95% CI 0.95-1.09;  $P=.58$ ) with high heterogeneity ( $I^2=78\%$ ) in the estimate (Figure 3). These findings suggest that tailoring had no effect on dropout attrition at 1-month follow-up.

**Figure 3.** Comparison of attrition between tailored and untailored conditions at 1-month follow-up by control type.

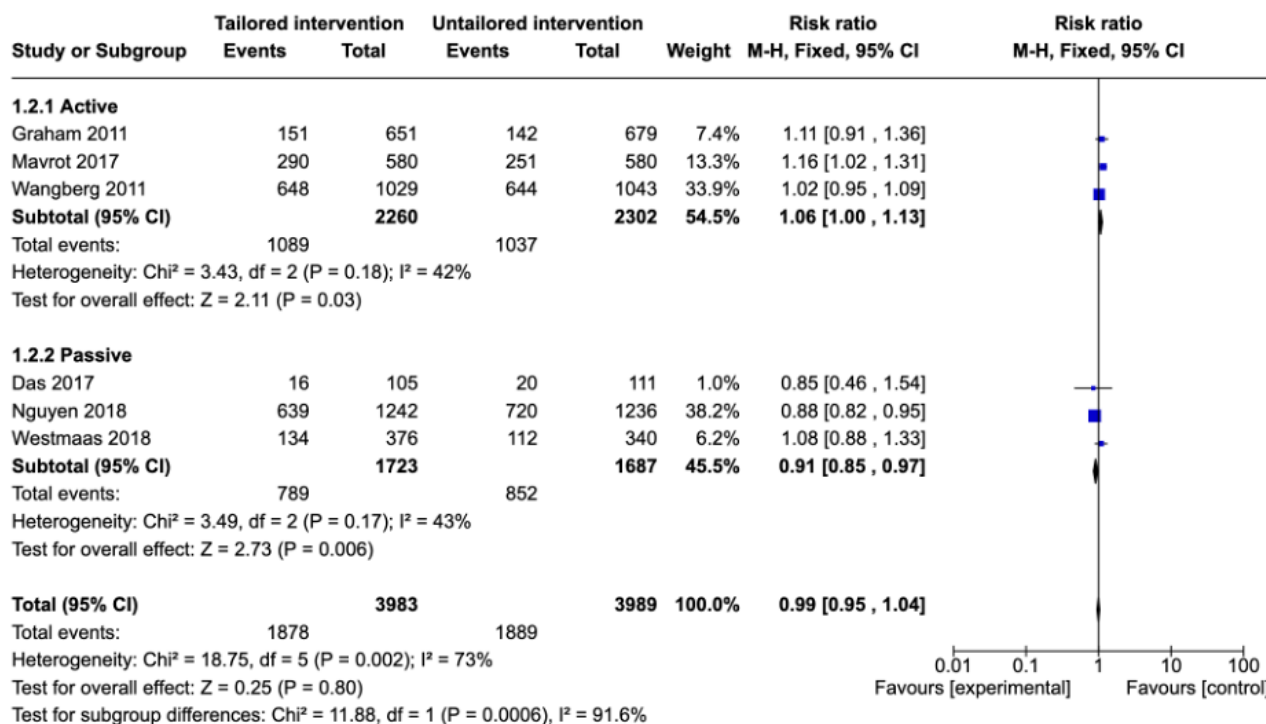


### Three-Month Follow-Up

Of the 11 studies included in the meta-analysis, 6 (55%) reported attrition at 3-month follow-up [21-23,26,28,33]. A total of 27% (3/11) of the studies [21-23] compared a tailored active intervention with an untailored active control. When these 3 studies were pooled, tailored interventions were at a significantly higher risk of attrition compared with untailored controls (RR=1.06, 95% CI 1.00-1.13;  $P=.03$ ) with moderate heterogeneity in the estimate ( $I^2=42\%$ ). The remaining 3 studies [26,28,33] compared a tailored active intervention with an untailored passive control. Among these studies, the risk of attrition was lower for the tailored intervention than for the

untailored intervention (RR=0.91, 95% CI 0.85-0.97;  $P=.006$ ). This estimate was statistically significant with moderate heterogeneity ( $I^2=43\%$ ). Significant differences between the active and passive subgroups were found at 3-month follow-up ( $\chi^2_1=11.9$ ;  $P<.001$ ).

When the 6 studies [23-25,28,30,35] that reported attrition at 3-month follow-up were pooled (Figure 4), no significant differences were found in the risk of attrition between the tailored interventions and their untailored controls (RR=0.99, 95% CI 0.95-1.04;  $P=.80$ ) with high heterogeneity in the estimate ( $I^2=73\%$ ). This finding suggests that tailoring had no effect on dropout attrition at 3-month follow-up.

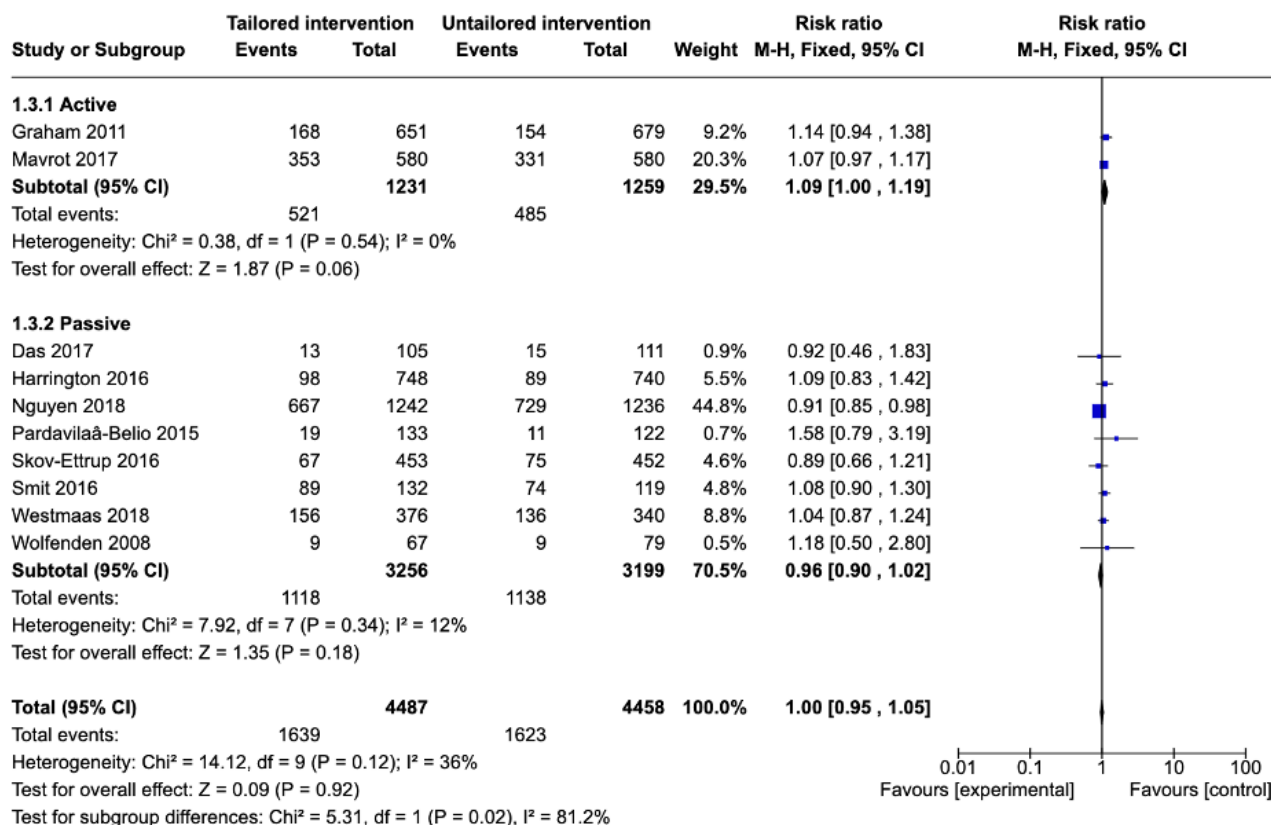
**Figure 4.** Comparison of attrition between tailored and untailored conditions at 3-month follow-up by control type.

### Six-Month Follow-Up

Of the 11 studies included in the meta-analysis, 9 (82%) reported attrition at 6-month follow-up [21,22,26-31,33]. Of these 9 studies, 2 compared a tailored active intervention with an untailored active control [21,22]. Differences in the risk of attrition between the tailored and untailored groups for these 2 studies were not significant ( $RR=1.09$ , 95% CI 1.00-1.19;  $P=.06$ ) with no heterogeneity in the estimate ( $I^2=0\%$ ). In addition, 7 of the 9 studies (78%) [26-31,33] compared a tailored active intervention with an untailored passive control. Among these 7 studies, there were no significant differences in the risk of attrition between the tailored intervention relative to the

untailored control ( $RR=0.96$ , 95% CI 0.90-1.02;  $P=.16$ ) with low heterogeneity in the estimate ( $I^2=22\%$ ). The test for subgroup differences suggested that significant differences existed between the active and passive subgroups at 6-month follow-up ( $\chi^2_1=5.4$ ;  $P=.02$ ).

When the 9 studies [21,22,26-31,33] that reported attrition at 6-month follow-up were pooled, no differences were found in dropout attrition between the tailored interventions and untailored controls ( $RR=1.00$ , 95% CI 0.95-1.05;  $P=.90$ ) with moderate heterogeneity ( $I^2=43\%$ , Figure 5). This suggests that tailoring had no effect on the risk of attrition at 6-month follow-up.

**Figure 5.** Comparison of attrition between tailored and untailored conditions at 6-month follow-up by control type.

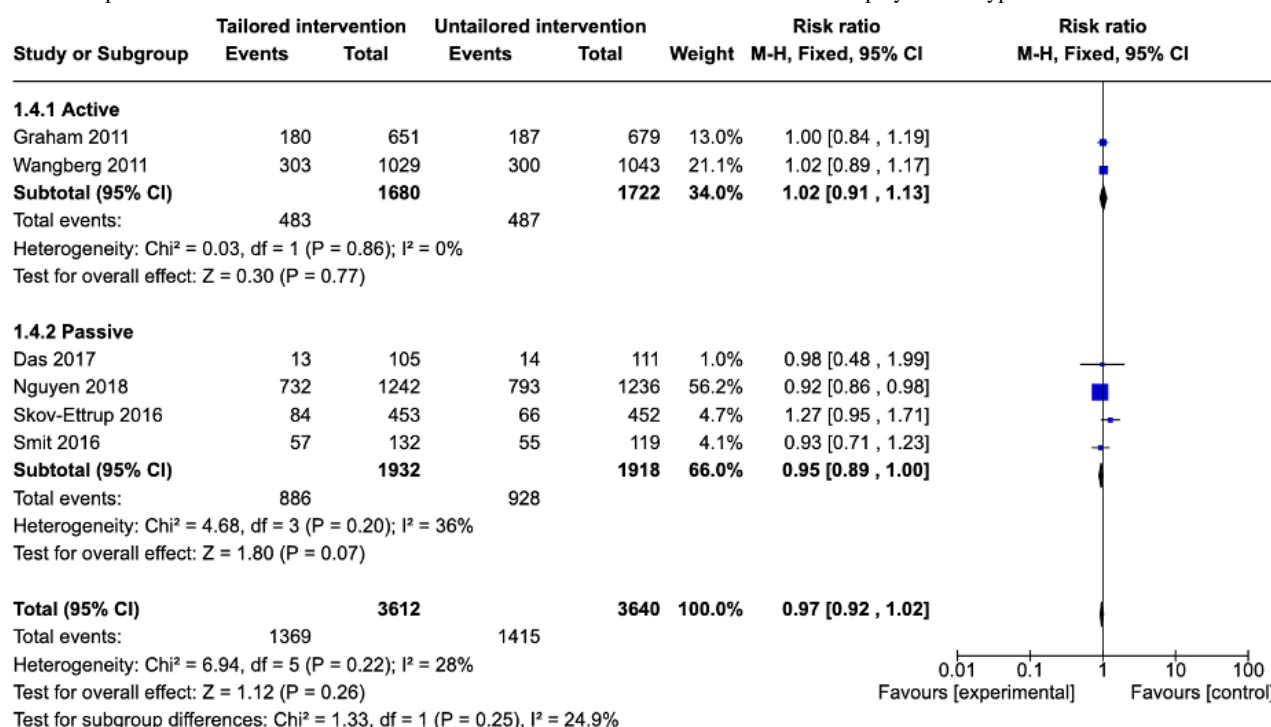
### Twelve-Month Follow-Up

Of the 11 studies included in the meta-analysis, 6 (55%) reported attrition at 12-month follow-up [21,23,26,28,30,31]. Of these 6 studies, 2 studies [21,23] compared a tailored active intervention with an untailored active control. The pooled effect of these 2 studies found that there were no significant differences in the risk of attrition between tailored interventions and untailored controls ( $RR=1.02$ , 95% CI 0.91-1.13;  $P=.77$ ) with no heterogeneity in the estimate ( $I^2=0\%$ ).

Four studies [26,28,30,31] reported attrition at 12-month follow-up that compared a tailored active intervention with an untailored passive control. The pooled effect of these 4 studies

found a lower risk of attrition among the tailored intervention ( $RR=0.95$ , 95% CI 0.89-1.00;  $P=.07$ ). This estimate was not statistically significant and had moderate heterogeneity ( $I^2=36\%$ ). The test for subgroup differences found no significant differences between the active and passive subgroups at 12-month follow-up ( $\chi^2_1=1.3$ ;  $P=.25$ ).

At 12-month follow-up, the pooled effect of the 6 studies [21,23,26,28,30,31] (Figure 6) demonstrated a lower risk of attrition for the tailored intervention compared with the untailored control ( $RR=0.97$ , 95% CI 0.92-1.02;  $P=.26$ ). This estimate was not statistically significant and had moderate heterogeneity ( $I^2=28\%$ ), suggesting that tailoring had no effect on dropout attrition at 12 months.

**Figure 6.** Comparison of attrition between tailored and untailored conditions at 12-month follow-up by control type.

## Sensitivity Analysis

Given the high heterogeneity of the results found for attrition at 1-, 3-, and 6-month follow-ups, a post hoc sensitivity analysis was conducted. Through visual inspection of the forest plots, the study by Westmaas [33] was suspected to be an outlier, and thus, this study was removed for the sensitivity analysis. Without the study by Westmaas [33] included in the meta-analysis, tailoring was found to have a higher risk of dropout attrition at 1-month follow-up ( $\text{RR}=1.03$ , 95% CI 0.96-1.11;  $P=.38$ ) with high heterogeneity ( $I^2=86\%$ ). Tailoring was found to have a lower risk of attrition both at 3-month follow-up ( $\text{RR}=0.99$ , 95% CI 0.94-1.03;  $P=.62$ ) with high heterogeneity ( $I^2=78\%$ ) and at 6-month follow-up ( $\text{RR}=0.99$ , 95% CI 0.94-1.05;  $P=.78$ ) with moderate heterogeneity ( $I^2=49\%$ , [Multimedia Appendix 4](#)). As none of these estimates were statistically significant, the sensitivity analysis suggests that tailoring did not have a statistically significant effect on attrition at 1-, 3-, and 6-month follow-ups.

## Discussion

### Principal Findings

This study sought to understand the impact of tailoring web-based interventions on attrition among studies using active and passive control groups. Although several reviews have investigated the efficacy of web-based smoking cessation interventions [13-15,36], to our knowledge, this is the first systematic review that compared the rate of attrition of tailored and untailored web-based smoking cessation interventions. Our review found no differences in the likelihood of attrition between tailored web-based interventions and untailored controls at 1, 3, 6, or 12-month follow-ups.

The results of this systematic review align with the findings of previous studies. Strecher [37] conducted a path analysis of tailoring in web-based smoking cessation interventions and found that the relationship between tailoring depth and smoking cessation was weakly mediated by longitudinal engagement. Moreover, in a trial of a web-based smoking cessation program, the number of web pages opened (a measure of engagement) did not predict 12-week cessation [38]. Thus, the results of this review, as well as previous studies, suggest that retention may not mediate the impact of tailoring on smoking cessation in the long term.

When considering the engagement level (active vs passive) of the control intervention, tailoring was found to have no effect early on at 1-month follow-up but increased attrition by the 3-month follow-up for studies with an active control. This increase in attrition, however, was diminished at 12-month follow-up. These results suggest that for studies with an active control group, tailoring had no effect on attrition in the long term at 12-month follow-up. The opposite was found to be true among studies that compared a tailored active intervention with an untailored passive control group, where tailoring was associated with increased attrition at 1-month follow-up and decreased attrition at 3-month follow-up. This effect diminished over time, with no effect of tailoring on attrition at 12-month follow-up. When stratified by the type of control intervention, significant differences in the effect of tailoring were found between the active and passive subgroups at 1-, 3-, and 6-month follow-ups.

A potential explanation for the increased attrition for studies with an active control is that participants in the active tailored intervention group may derive benefits from the intervention early on and, thus, seize their participation before the end of the study. For studies with a passive untailored control, the opposite



effect may have been observed, as there may be fewer other components to maintain interest in the intervention and in the study. Additional research is needed, however, to investigate this effect.

An important finding of this review was that only a few studies used an active control group ( $n=4$ ), which suggests a lack of RCTs that compare web-based interventions with active control groups. The lack of active control groups may stem from the application of traditional RCT designs, well suited for investigating drug efficacy, onto digital health interventions [39]. Although robust methods such as the RCT are important, web-based interventions often have multiple active ingredients, including content and technology features. As such, there is a need to consider what the active ingredients of the technology are hypothesized to be to isolate them between the intervention and the control.

Another notable result was that all studies included in the review took place in the Western countries—a finding that has been reported by a previous review [36]. Given that developing countries continue to be targeted by tobacco companies [40], face difficulties in adopting tobacco strategies [41], and are projected to account for 80% of smoking-related deaths in the next century [42], studies are needed to evaluate the effectiveness of web-based smoking cessation interventions in populations outside of the Western countries. Tailoring interventions for these populations may require consideration of different variables (ie, cultural factors) and have unique impacts on attrition and effectiveness.

### Implications for Future Work

The findings of this review highlight the importance of selecting an appropriate control condition when evaluating the effectiveness of web-based smoking cessation interventions and identifying the important design components of these interventions. Where possible, web-based interventions should be compared with interventions that require the same amount of engagement to isolate the hypothesized active ingredients of the intervention. Without appropriate control interventions, studies may risk misrepresenting the benefits and mechanisms of intervention features, such as tailoring.

In this study, tailoring was intentionally operationalized as a unitary construct [43] to investigate the impact of tailoring on attrition. Although beyond the scope of this review, future research is needed to explore more specific questions related to tailoring, such as the impact of various tailoring strategies on the effectiveness of web-based smoking cessation interventions. The findings of this review suggest, however, that an investigation of specific tailoring strategies may be difficult because of unclear reporting of tailoring among the included studies. Future studies on tailored web-based smoking cessation interventions should more clearly outline the tailored components using the tailoring reporting standards proposed by Harrington [44]. Reporting in accordance with these standards will facilitate more nuanced analyses of the differential impact of various tailoring strategies.

### Limitations

The findings of this review should be interpreted with caution because of several limitations. First, multiple studies included in the review were found to be of unclear or high risk of performance bias. Second, the high heterogeneity in the intervention type and components, as well as the study design, made it difficult to pool the data on attrition and generalize the results of the study. Indeed, some studies included in this review had interventions that were both tailored and interactive [21-24,27,32]. This is an important limitation of this study, as interactivity can include personalization but not in all cases. Although previous reviews have grouped tailoring and interactivity [14], we sought to isolate the impact of tailoring and, thus, did not include interactivity. Some studies also had features in addition to tailoring and interactivity, such as social support or coaching, which could have impacted attrition. Third, in this review, we focused on dropout attrition and thus our findings may not capture the impact of tailoring on nonusage attrition. As only 1 study included in this review reported nonusage attrition, we anticipate that such analyses may be difficult to perform without improvements in reporting. Fourth, this review did not account for user acceptance and experience of technologies, which may have affected attrition, particularly in the short term. Finally, given that technologies are rapidly being developed both for research and in the consumer market, this review may have missed web-based interventions that were publicly available, especially outside of the Western countries. The restriction on published RCT studies in the English language may have contributed to the lack of interventions delivered internationally.

### Conclusions

This study aimed to investigate the impact of tailoring on attrition in web-based smoking cessation interventions. A systematic review of the literature yielded 14 RCTs that compared tailored web-based interventions with untailored control interventions, with 4 (29%) of the studies using an active control intervention and 10 (71%) studies using a passive control intervention. A meta-analysis of attrition reported by 86% (12/14) of the included studies found no effects of tailoring on attrition at 1, 3, 6, and 12-month follow-ups. The findings of this review suggest that tailoring may not be associated with reduced rates of attrition in web-based smoking cessation studies in the long term, although the high heterogeneity of effects indicates that the findings should be interpreted with caution. Future studies that incorporate active web-based controls, compare the impact of different tailoring strategies, and include populations outside of the Western countries are needed. Moreover, although beyond the scope of this review, future reviews may consider the impact of other technology features on attrition. It is with a greater understanding of how intervention features impact attrition that technologies may be better designed to retain users who may benefit from web-based smoking cessation interventions.

## Acknowledgments

This review was conducted as part of the Research on Advancing Cessation Treatment project, which was funded by the Ontario Ministry of Health and Long-Term Care's Health System Research Fund (Canada).

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) checklist.

[[PDF File \(Adobe PDF File\), 111 KB - jmir\\_v22i10e16255\\_app1.pdf](#)]

### Multimedia Appendix 2

Sample search strategy (MEDLINE: Medical Literature Analysis and Retrieval System Online).

[[PDF File \(Adobe PDF File\), 43 KB - jmir\\_v22i10e16255\\_app2.pdf](#)]

### Multimedia Appendix 3

Characteristics of the included studies.

[[DOCX File, 32 KB - jmir\\_v22i10e16255\\_app3.docx](#)]

### Multimedia Appendix 4

Sensitivity analysis of tailoring at 1-, 3-, and 6-month follow-ups.

[[PDF File \(Adobe PDF File\), 630 KB - jmir\\_v22i10e16255\\_app4.pdf](#)]

## References

1. WHO Report on the Global Tobacco Epidemic, 2013: Enforcing Bans on Tobacco Advertising, Promotion and Sponsorship. World Health Organization. 2013. URL: [https://apps.who.int/iris/bitstream/handle/10665/85380/9789241505871\\_eng.pdf?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/85380/9789241505871_eng.pdf?sequence=1) [accessed 2019-08-05]
2. Tobacco. World Health Organization. URL: <http://www.who.int/news-room/fact-sheets/detail/tobacco> [accessed 2019-08-05]
3. GBD 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. *Lancet* 2017 May 13;389(10082):1885-1906 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(17\)30819-X](https://doi.org/10.1016/S0140-6736(17)30819-X)] [Medline: [28390697](#)]
4. Fong GT, Hammond D, Laux FL, Zanna MP, Cummings KM, Borland R, et al. The near-universal experience of regret among smokers in four countries: findings from the International Tobacco Control Policy Evaluation Survey. *Nicotine Tob Res* 2004 Dec;6 Suppl 3:S341-S351. [doi: [10.1080/14622200412331320743](https://doi.org/10.1080/14622200412331320743)] [Medline: [15799597](#)]
5. Chaiton M, Diemert L, Cohen JE, Bondy SJ, Selby P, Philipneri A, et al. Estimating the number of quit attempts it takes to quit smoking successfully in a longitudinal cohort of smokers. *BMJ Open* 2016 Jun 9;6(6):e011045 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2016-011045](https://doi.org/10.1136/bmjopen-2016-011045)] [Medline: [27288378](#)]
6. Doll R, Peto R, Boreham J, Sutherland I. Mortality from cancer in relation to smoking: 50 years observations on British doctors. *Br J Cancer* 2005 Feb 14;92(3):426-429 [[FREE Full text](#)] [doi: [10.1038/sj.bjc.6602359](https://doi.org/10.1038/sj.bjc.6602359)] [Medline: [15668706](#)]
7. Lam T. Absolute risk of tobacco deaths: one in two smokers will be killed by smoking: comment on "Smoking and all-cause mortality in older people". *Arch Intern Med* 2012 Jun 11;172(11):845-846. [doi: [10.1001/archinternmed.2012.1927](https://doi.org/10.1001/archinternmed.2012.1927)] [Medline: [22688993](#)]
8. Cummings KM, Hyland A. Impact of nicotine replacement therapy on smoking behavior. *Annu Rev Public Health* 2005;26:583-599. [doi: [10.1146/annurev.publhealth.26.021304.144501](https://doi.org/10.1146/annurev.publhealth.26.021304.144501)] [Medline: [15760302](#)]
9. Abrams DB, Orleans CT, Niaura RS, Goldstein MG, Prochaska JO, Velicer W. Integrating individual and public health perspectives for treatment of tobacco dependence under managed health care: a combined stepped-care and matching model. *Ann Behav Med* 1996;18(4):290-304. [doi: [10.1007/BF02895291](https://doi.org/10.1007/BF02895291)] [Medline: [18425675](#)]
10. Norman CD, McIntosh S, Selby P, Eysenbach G. Web-assisted tobacco interventions: empowering change in the global fight for the public's (e)Health. *J Med Internet Res* 2008 Nov 25;10(5):e48 [[FREE Full text](#)] [doi: [10.2196/jmir.1171](https://doi.org/10.2196/jmir.1171)] [Medline: [19033147](#)]
11. Eysenbach G. The law of attrition. *J Med Internet Res* 2005 Mar 31;7(1):e11 [[FREE Full text](#)] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](#)]
12. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2017 Sep 4;9:CD007078 [[FREE Full text](#)] [doi: [10.1002/14651858.CD007078.pub5](https://doi.org/10.1002/14651858.CD007078.pub5)] [Medline: [28869775](#)]
13. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. *Addiction* 2009 Nov;104(11):1792-1804. [doi: [10.1111/j.1360-0443.2009.02710.x](https://doi.org/10.1111/j.1360-0443.2009.02710.x)] [Medline: [19832783](#)]

14. Graham AL, Carpenter KM, Cha S, Cole S, Jacobs MA, Raskob M, et al. Systematic review and meta-analysis of internet interventions for smoking cessation among adults. *Subst Abuse Rehabil* 2016;7:55-69 [FREE Full text] [doi: [10.2147/SAR.S101660](https://doi.org/10.2147/SAR.S101660)] [Medline: [27274333](https://pubmed.ncbi.nlm.nih.gov/27274333/)]
15. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. *Ann Behav Med* 1999;21(4):276-283. [doi: [10.1007/BF02895958](https://doi.org/10.1007/BF02895958)] [Medline: [10721433](https://pubmed.ncbi.nlm.nih.gov/10721433/)]
16. Strecher VJ. Computer-tailored smoking cessation materials: a review and discussion. *Patient Educ Couns* 1999 Feb;36(2):107-117. [doi: [10.1016/s0738-3991\(98\)00128-1](https://doi.org/10.1016/s0738-3991(98)00128-1)] [Medline: [10223016](https://pubmed.ncbi.nlm.nih.gov/10223016/)]
17. Geraghty AW, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international Internet health interventions: a step towards global eHealth. *Health Promot Int* 2013 Sep;28(3):442-452 [FREE Full text] [doi: [10.1093/heapro/das029](https://doi.org/10.1093/heapro/das029)] [Medline: [22786673](https://pubmed.ncbi.nlm.nih.gov/22786673/)]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
19. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *Br Med J* 2011 Oct 18;343:d5928 [FREE Full text] [doi: [10.1136/bmj.d5928](https://doi.org/10.1136/bmj.d5928)] [Medline: [22008217](https://pubmed.ncbi.nlm.nih.gov/22008217/)]
20. Fletcher J. What is heterogeneity and is it important? *Br Med J* 2007 Jan 13;334(7584):94-96 [FREE Full text] [doi: [10.1136/bmj.39057.406644.68](https://doi.org/10.1136/bmj.39057.406644.68)] [Medline: [17218716](https://pubmed.ncbi.nlm.nih.gov/17218716/)]
21. Graham AL, Cobb NK, Papandonatos GD, Moreno JL, Kang H, Tinkelman DG, et al. A randomized trial of internet and telephone treatment for smoking cessation. *Arch Intern Med* 2011 Jan 10;171(1):46-53 [FREE Full text] [doi: [10.1001/archinternmed.2010.451](https://doi.org/10.1001/archinternmed.2010.451)] [Medline: [21220660](https://pubmed.ncbi.nlm.nih.gov/21220660/)]
22. Mavrot C, Stucki I, Sager F, Etter J. Efficacy of an internet-based, individually tailored smoking cessation program: a randomized-controlled trial. *J Telemed Telecare* 2017 Jun;23(5):521-528. [doi: [10.1177/1357633X16655476](https://doi.org/10.1177/1357633X16655476)] [Medline: [27365318](https://pubmed.ncbi.nlm.nih.gov/27365318/)]
23. Wangberg SC, Nilsen O, Antypas K, Gram IT. Effect of tailoring in an internet-based intervention for smoking cessation: randomized controlled trial. *J Med Internet Res* 2011 Dec 15;13(4):e121 [FREE Full text] [doi: [10.2196/jmir.1605](https://doi.org/10.2196/jmir.1605)] [Medline: [22169631](https://pubmed.ncbi.nlm.nih.gov/22169631/)]
24. An LC, Klatt C, Perry CL, Lein EB, Hennrikus DJ, Pallonen UE, et al. The RealU online cessation intervention for college smokers: a randomized controlled trial. *Prev Med* 2008 Aug;47(2):194-199. [doi: [10.1016/j.ypmed.2008.04.011](https://doi.org/10.1016/j.ypmed.2008.04.011)] [Medline: [18565577](https://pubmed.ncbi.nlm.nih.gov/18565577/)]
25. Borland R, Balmford J, Benda P. Population-level effects of automated smoking cessation help programs: a randomized controlled trial. *Addiction* 2013 Mar;108(3):618-628. [doi: [10.1111/j.1360-0443.2012.04091.x](https://doi.org/10.1111/j.1360-0443.2012.04091.x)] [Medline: [22994457](https://pubmed.ncbi.nlm.nih.gov/22994457/)]
26. Das S, Hickman NJ, Prochaska JJ. Treating smoking in adults with co-occurring acute psychiatric and addictive disorders. *J Addict Med* 2017;11(4):273-279 [FREE Full text] [doi: [10.1097/ADM.0000000000000320](https://doi.org/10.1097/ADM.0000000000000320)] [Medline: [28441272](https://pubmed.ncbi.nlm.nih.gov/28441272/)]
27. Kathleen FH, Young-Il K, Meifang C, Rekha R, Maria P, Rajani S, et al. Web-based intervention for transitioning smokers from inpatient to outpatient care: an RCT. *Am J Prev Med* 2016 Oct;51(4):620-629. [doi: [10.1016/j.amepre.2016.04.008](https://doi.org/10.1016/j.amepre.2016.04.008)] [Medline: [27647062](https://pubmed.ncbi.nlm.nih.gov/27647062/)]
28. Thanh VN, Guignard R, Lancrenon S, Bertrand C, Delva C, Berlin I, et al. Effectiveness of a fully automated internet-based smoking cessation program: a randomized controlled trial (STAMP). *Nicotine Tob Res* 2019 Jan 4;21(2):163-172. [doi: [10.1093/ntr/nty016](https://doi.org/10.1093/ntr/nty016)] [Medline: [29370407](https://pubmed.ncbi.nlm.nih.gov/29370407/)]
29. Pardavila-Belio MI, García-Vivar C, Pimenta AM, Canga-Armayor A, Pueyo-Garrigues S, Canga-Armayor N. Intervention study for smoking cessation in Spanish college students: pragmatic randomized controlled trial. *Addiction* 2015 Oct;110(10):1676-1683. [doi: [10.1111/add.13009](https://doi.org/10.1111/add.13009)] [Medline: [26053958](https://pubmed.ncbi.nlm.nih.gov/26053958/)]
30. Skov-Ettrup L, Dalum P, Bech M, Tolstrup JS. The effectiveness of telephone counselling and internet- and text-message-based support for smoking cessation: results from a randomized controlled trial. *Addiction* 2016 Jul;111(7):1257-1266. [doi: [10.1111/add.13302](https://doi.org/10.1111/add.13302)] [Medline: [26748541](https://pubmed.ncbi.nlm.nih.gov/26748541/)]
31. Smit ES, Hoving C, Cox VC, de Vries H. Influence of recruitment strategy on the reach and effect of a web-based multiple tailored smoking cessation intervention among Dutch adult smokers. *Health Educ Res* 2012 Apr;27(2):191-199. [doi: [10.1093/her/cyr099](https://doi.org/10.1093/her/cyr099)] [Medline: [22058288](https://pubmed.ncbi.nlm.nih.gov/22058288/)]
32. Tsoh JY, Kohn MA, Gerbert B. Promoting smoking cessation in pregnancy with video doctor plus provider cueing: a randomized trial. *Acta Obstet Gynecol Scand* 2010;89(4):515-523 [FREE Full text] [doi: [10.3109/00016341003678419](https://doi.org/10.3109/00016341003678419)] [Medline: [20196678](https://pubmed.ncbi.nlm.nih.gov/20196678/)]
33. Westmaas JL, Bontemps-Jones J, Hendricks PS, Kim J, Abrams LC. Randomised controlled trial of stand-alone tailored emails for smoking cessation. *Tob Control* 2018 Mar;27(2):136-146. [doi: [10.1136/tobaccocontrol-2016-053056](https://doi.org/10.1136/tobaccocontrol-2016-053056)] [Medline: [28522745](https://pubmed.ncbi.nlm.nih.gov/28522745/)]
34. Kahler CW, Cohn AM, Costantino C, Toll BA, Spillane NS, Graham AL. A digital smoking cessation program for heavy drinkers: pilot randomized controlled trial. *JMIR Form Res* 2020 Jun 8;4(6):e7570 [FREE Full text] [doi: [10.2196/formative.7570](https://doi.org/10.2196/formative.7570)] [Medline: [32348286](https://pubmed.ncbi.nlm.nih.gov/32348286/)]

35. Altendorf M, Hoving C, Van Weert JC, Smit ES. Effectiveness of message frame-tailoring in a web-based smoking cessation program: randomized controlled trial. *J Med Internet Res* 2020 Apr 3;22(4):e17251 [FREE Full text] [doi: [10.2196/17251](https://doi.org/10.2196/17251)] [Medline: [32242826](https://pubmed.ncbi.nlm.nih.gov/32242826/)]
36. Do HP, Tran BX, Le Pham Q, Nguyen LH, Tran TT, Latkin CA, et al. Which eHealth interventions are most effective for smoking cessation? A systematic review. *Patient Prefer Adherence* 2018;12:2065-2084 [FREE Full text] [doi: [10.2147/PPA.S169397](https://doi.org/10.2147/PPA.S169397)] [Medline: [30349201](https://pubmed.ncbi.nlm.nih.gov/30349201/)]
37. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. *J Med Internet Res* 2008 Nov 4;10(5):e36 [FREE Full text] [doi: [10.2196/jmir.1002](https://doi.org/10.2196/jmir.1002)] [Medline: [18984557](https://pubmed.ncbi.nlm.nih.gov/18984557/)]
38. Strecher VJ, Shiffman S, West R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. *Addiction* 2005 May;100(5):682-688. [doi: [10.1111/j.1360-0443.2005.01093.x](https://doi.org/10.1111/j.1360-0443.2005.01093.x)] [Medline: [15847626](https://pubmed.ncbi.nlm.nih.gov/15847626/)]
39. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. *PLoS Med* 2009 Aug;6(8):e1000126 [FREE Full text] [doi: [10.1371/journal.pmed.1000126](https://doi.org/10.1371/journal.pmed.1000126)] [Medline: [19688038](https://pubmed.ncbi.nlm.nih.gov/19688038/)]
40. Sebríe E, Glantz SA. The tobacco industry in developing countries. *Br Med J* 2006 Feb 11;332(7537):313-314 [FREE Full text] [doi: [10.1136/bmj.332.7537.313](https://doi.org/10.1136/bmj.332.7537.313)] [Medline: [16470028](https://pubmed.ncbi.nlm.nih.gov/16470028/)]
41. World Health Organization. Economics of Tobacco Toolkit. Assessment of the Economic Costs of Smoking Geneva, Switzerland: WHO 2011. World Health Organization.: World Health; 2020. URL: [https://apps.who.int/iris/bitstream/handle/10665/44596/9789241501576\\_eng.pdf](https://apps.who.int/iris/bitstream/handle/10665/44596/9789241501576_eng.pdf)
42. WHO Report on the Global Tobacco Epidemic 2008: The MPOWER Package. World Health Organization. 2008. URL: [https://apps.who.int/iris/bitstream/handle/10665/43818/9789241596282\\_eng.pdf;jsessionid=DC45BB004C41FD035C9C0E3B85AB10C7?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/43818/9789241596282_eng.pdf;jsessionid=DC45BB004C41FD035C9C0E3B85AB10C7?sequence=1) [accessed 2019-08-05]
43. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. *Health Educ Res* 2008 Jun;23(3):454-466 [FREE Full text] [doi: [10.1093/her/cyn004](https://doi.org/10.1093/her/cyn004)] [Medline: [18349033](https://pubmed.ncbi.nlm.nih.gov/18349033/)]
44. Harrington NG, Noar SM. Reporting standards for studies of tailored interventions. *Health Educ Res* 2012 Apr;27(2):331-342. [doi: [10.1093/her/cyr108](https://doi.org/10.1093/her/cyr108)] [Medline: [22156230](https://pubmed.ncbi.nlm.nih.gov/22156230/)]

## Abbreviations

**MEDLINE:** Medical Literature Analysis and Retrieval System Online

**PRISMA:** Preferred Reporting Items for Systematic reviews and Meta-Analyses

**RCT:** randomized controlled trial

**RR:** risk ratio

*Edited by G Eysenbach; submitted 14.09.19; peer-reviewed by R Bajpai, A Lazard, M Kwon, S Cha; comments to author 27.11.19; revised version received 24.04.20; accepted 03.06.20; published 19.10.20.*

*Please cite as:*

Shah A, Chaiton M, Baliunas D, Schwartz R

Tailored Web-Based Smoking Interventions and Reduced Attrition: Systematic Review and Meta-Analysis

*J Med Internet Res* 2020;22(10):e16255

URL: <https://www.jmir.org/2020/10/e16255>

doi: [10.2196/16255](https://doi.org/10.2196/16255)

PMID: [33074158](https://pubmed.ncbi.nlm.nih.gov/33074158/)

©Amika Shah, Michael Chaiton, Dolly Baliunas, Robert Schwartz. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 19.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Review

# Determining if Telehealth Can Reduce Health System Costs: Scoping Review

Centaine L Snoswell<sup>1,2</sup>, PhD, MPH, BPharm; Monica L Taylor<sup>1,2</sup>, MPH, BSc (Hons); Tracy A Comans<sup>2</sup>, PhD, BEcon, BPT; Anthony C Smith<sup>1,2,3</sup>, PhD, MEd, BNurs; Leonard C Gray<sup>2</sup>, PhD, MMed, MBBS; Liam J Caffery<sup>1,2</sup>, PhD, DipAppSci, BInfoTech

<sup>1</sup>Centre for Online Health, The University of Queensland, Brisbane, Australia

<sup>2</sup>Centre for Health Services Research, The University of Queensland, Brisbane, Australia

<sup>3</sup>Centre for Innovative Medical Technology, University of Southern Denmark, Odense, Denmark

**Corresponding Author:**

Liam J Caffery, PhD, DipAppSci, BInfoTech

Centre for Online Health

The University of Queensland

Building 33, Princess Alexandra Hospital

Brisbane,

Australia

Phone: 61 7 3176 7704

Email: [l.caffery@uq.edu.au](mailto:l.caffery@uq.edu.au)

## Abstract

**Background:** Telehealth represents an opportunity for Australia to harness the power of technology to redesign the way health care is delivered. The potential benefits of telehealth include increased accessibility to care, productivity gains for health providers and patients through reduced travel, potential for cost savings, and an opportunity to develop culturally appropriate services that are more sensitive to the needs of special populations. The uptake of telehealth has been hindered at times by clinician reluctance and policies that preclude metropolitan populations from accessing telehealth services.

**Objective:** This study aims to investigate if telehealth reduces health system costs compared with traditional service models and to identify the scenarios in which cost savings can be realized.

**Methods:** A scoping review was undertaken to meet the study aims. Initially, literature searches were conducted using broad terms for telehealth and economics to identify economic evaluation literature in telehealth. The investigators then conducted an expert focus group to identify domains where telehealth could reduce health system costs, followed by targeted literature searches for corresponding evidence.

**Results:** The cost analyses reviewed provided evidence that telehealth reduced costs when health system-funded travel was prevented and when telehealth mitigated the need for expensive procedural or specialist follow-up by providing competent care in a more efficient way. The expert focus group identified 4 areas of potential savings from telehealth: productivity gains, reductions in secondary care, alternate funding models, and telementoring. Telehealth demonstrated great potential for productivity gains arising from health system redesign; however, under the Australian activity-based funding, it is unlikely that these gains will result in cost savings. Secondary care use mitigation is an area of promise for telehealth; however, many studies have not demonstrated overall cost savings due to the cost of administering and monitoring telehealth systems. Alternate funding models from telehealth systems have the potential to save the health system money in situations where the consumers pay out of pocket to receive services. Telementoring has had minimal economic evaluation; however, in the long term it is likely to result in inadvertent cost savings through the upskilling of generalist and allied health clinicians.

**Conclusions:** Health services considering implementing telehealth should be motivated by benefits other than cost reduction. The available evidence has indicated that although telehealth provides overwhelmingly positive patient benefits and increases productivity for many services, current evidence suggests that it does not routinely reduce the cost of care delivery for the health system.

(*J Med Internet Res* 2020;22(10):e17298) doi:[10.2196/17298](https://doi.org/10.2196/17298)



**KEYWORDS**

cost-benefit analysis; telemedicine; review

**Introduction**

The sustainability of health systems is a major concern for governments worldwide. The financial viability of the health system is of particular concern, in light of both increasing costs and the increasing ratio of health expenditure to gross domestic product (GDP). For example, in Australia, in the decade to 2017, health care expenditure nearly doubled and the ratio of health expenditure to GDP increased from 8.75% to 10.28% [1]. Similarly, in the United States during the same time frame, health expenditure grew by 50% and the ratio of health expenditure to GDP increased from 15.9% to 17.9% [2]. This has catalyzed an imperative to reduce the cost of providing health care, as these increases are not sustainable long term.

Telehealth is the delivery of clinical health services using information and communication technologies to bridge the geographic separation of the clinician and consumer. Telehealth could potentially impact costs due to shorter interactions, reduced travel, economies of scale, increased revenues, or moving elements of care from clinicians to technology (eg, monitoring device) or to the patient themselves. The potential to reduce the cost of health care is one of the predominant reasons for the interest in implementing telehealth, followed closely by a desire to improve access to health care. Telehealth is often used to substitute a proportion of in-person encounters, and this substitution raises the question of a relative cost reduction.

Demonstrated cost-effective interventions may not be implemented due to budget constraints [3]. Hence, many health care organizations may be limited to implementing only telehealth interventions where cost reduction can be realized within a budgetary cycle and there is no cost increase associated with implementation. However, these programs are few, and

those that start successfully are not always sustainable in the long run, scalable, or transferable to other settings. Although proponents of telehealth often have projected health care savings, there is a dearth of evidence to support this view. Important questions relating to cost and sustainability remain unanswered [4].

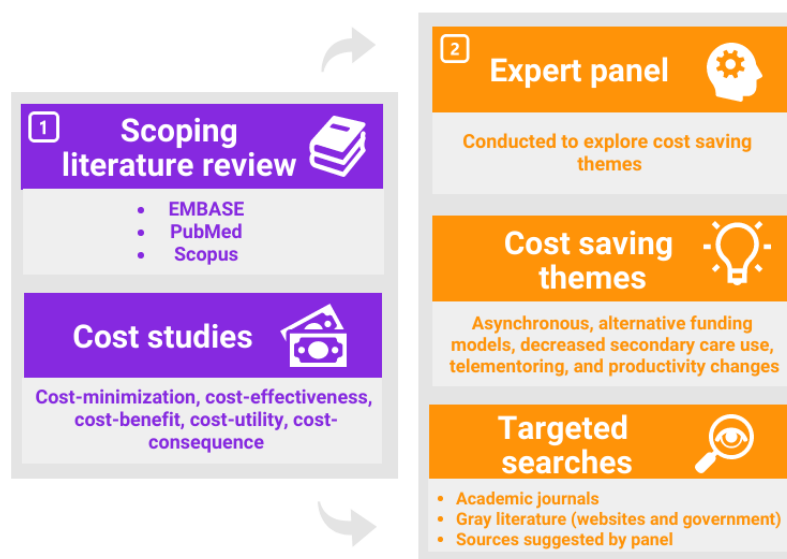
Reviews of telehealth cost-effectiveness are often limited to one clinical specialty, service modality, or country [5-8]. Limited research has looked collectively at available evidence, and to our knowledge, none have synthesized it from the perspective of the health system. Despite the differences between international health systems, identifying and collating information regarding the cost-saving potential of telehealth is valuable.

The aims of this scoping review were twofold. First, to investigate if telehealth reduces health system costs compared with traditional service models using international evidence, and second, to identify the scenarios in which cost savings can be realized.

**Methods****Study Design**

The Arksey and O'Malley scoping review method was used to achieve the aims of this study [9]. This methodology involved an initial literature review, followed by an expert focus group, and finally targeted literature searches based on the focus group discussion (Figure 1) [9]. A scoping review method was chosen due to the volume of literature [10]. Reporting of the methods and results was performed in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist [11].

**Figure 1.** Scoping review methods.



## Initial Literature Review (Economic Evidence)

To identify literature reporting the results of economic evaluations in telehealth, initial literature searches were conducted using terms relating to telehealth, telemedicine, and economics. Searches were conducted using the PubMed, EMBASE, and Scopus databases. Search results were restricted to publications available in English, and no restriction was placed on the country of origin or health system model described. The scope of the review was restricted to short- to

mid-term (6 months to 3 years) term and cost implications reported from the perspective of the health system payer (ie, health service or hospital). After the completion of database searches, duplicates were removed, titles and abstracts were screened by 2 authors (MT and CS), and a full-text review of articles was conducted by 3 authors (MT, LC, and CS). Included studies were categorized as cost-minimization analysis (CMA), cost-effectiveness analysis (CEA), and cost-utility analysis (CUA) studies [6,12,13]. Further descriptions of these categories are provided in Table 1.

**Table 1.** Description of cost analysis types.

Method	Description
CMA <sup>a</sup> [14]	CMA requires either proof or a stated assumption that the two comparators are equally effective, and therefore, the analysis only examines the difference in cost between the comparators. When comparing CMA, it is important to examine the items included for costing for each comparator as well as the final reported result.
CEA <sup>b</sup> [12]	CEA quantifies both the costs and a measurable effect (eg, blood pressure in mm Hg or days to diagnosis) from the comparators and presents them as a cost per increment of effectiveness. Due to the variety in measured effects, CEA are not easily comparable unless they use the same measure for effectiveness.
CUA <sup>c</sup>	CUA uses measures of cost and health-related quality of life (often expressed as a utility value) to compare interventions with usual care. Although more comparable, it is important to examine not only the cost estimations but also the method of eliciting health-related quality of life within each study.

<sup>a</sup>CMA: cost-minimization analysis.

<sup>b</sup>CEA: cost-effectiveness analysis.

<sup>c</sup>CUA: cost-utility analysis.

Data were extracted from the included studies by two investigators, and any disagreements in data extraction were discussed until a consensus was reached. The results were synthesized, and descriptions of systems in which telehealth reduces costs to the health system were identified and reported. All articles were quality assessed using the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) [15]. CHEERS is a 24-point quality assessment tool for published economic evaluations. It includes domains to assess the reporting of relevant economic principles such as currency, discounting, time horizon, effectiveness measures, choice of outcome, assumptions, and model choice. The CMA studies were assessed according to 20 of the 24 points in the checklist; choice of health outcome, measurement of effectiveness, measurement and valuation of preference-based outcomes, and incremental costs and outcomes were not applicable. All costs were converted to 2019 US \$ to allow for ease of comparison. The synthesis of the findings was narrative.

## Expert Panel Focus Group

Ethics approval to undertake the focus group was received from The University of Queensland's Human Research Ethics Committee (Approval # 2018002428).

The framework proposed by Arksey and O'Malley [9] identifies that reviews can be enhanced by including a consultation process. To this end, investigators conducted an expert focus group to identify the domains where telehealth could reduce health system costs.

National experts in telehealth and economics were sent an invitation to attend the focus group on "How can telehealth reduce health system costs?". Experts were identified and recruited through existing relationships between the research

team and the Australasian Telehealth Society who assisted with recommending individuals with both health economics and telehealth expertise. A total of 16 experts were invited to participate and 9 agreed to represent 7 different organizations. Participation was via focus group (n=7) or via email and telephone with the investigators (n=2).

Attendees were given the option to attend in person or via videoconference. Before attendance, invitees were emailed a discussion paper. Invitees were asked if they agreed or disagreed with the points in the discussion paper and if any additional factors should be considered. Some invitees who were unable to attend the meeting provided feedback via email or phone. Open-ended questions on the potential for cost savings arising from telehealth were used to facilitate discussion in the focus group. The discussion was recorded and notes were recorded. Topics suggested by the expert panel were synthesized and categorized under distinct topic headings (domains). These domains were used to direct further literature searches. By including the focus group consultative exercise where participants were given potential domains, it was possible to elicit more ideas on the subject matter, resulting in richer data [9].

## Targeted Literature Searches

Subsequent to the focus groups, further literature searches were conducted to locate evidence on domains identified by the expert panel. Literature searches were conducted using both broad search terms and domain-specific terms to identify supporting evidence. Searches were conducted in PubMed, EMBASE, Scopus, and gray literature. Additionally, members of the expert panel contributed relevant evidence items (designated as hand searches). No restrictions on publication date were set during

any search process. Searches were conducted from May 2018 to January 2019. Articles were omitted when the telehealth modality referenced telephone only. Search terms included, but were not limited to, telehealth, telemedicine, store-and-forward and other telehealth nomenclature, and domain-specific search terms such as secondary care or productivity. No restrictions were placed on the country of origin or the health system described in the analyses.

The level of evidence was described using the National Health and Medical Research Council guidelines [16], where, for example, I is a systematic review of randomized controlled trials (RCTs), II is an RCT, III is a comparative study with or without controls, and IV is a case series. The results of the subsequent

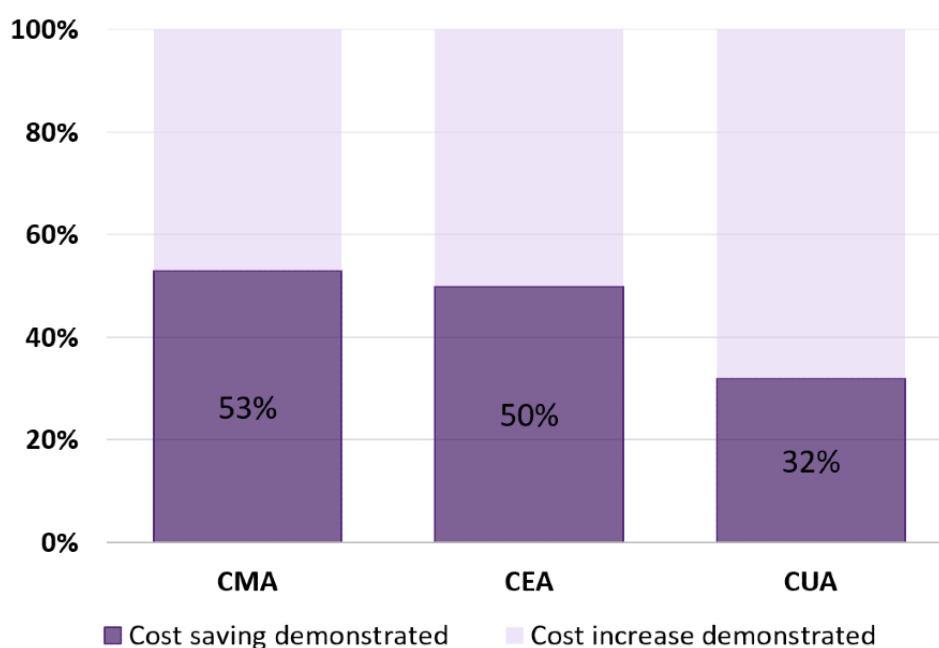
articles identified through the targeted searches were summarized.

## Results

### CMA

Searches and screening identified 17 cost-minimization studies that reported their results from the perspective of the health system. Of these 17, 9 studies (53%) reported telehealth to be cost saving compared with conventional care (Figure 2), 6 studies (35%) reported telehealth to be cost saving after a workload threshold was achieved, and 2 studies (12%) reported telehealth to be more expensive than conventional care. The overall quality of reporting was sound, with an average score of 15 out of 20 (Table 2).

**Figure 2.** Proportion of studies identified that saved costs. CEA: cost-effectiveness analysis; CMA: cost-minimization analysis; CUA: cost-utility analysis.



**Table 2.** Cost-minimization analysis demonstrating lower costs for telehealth from the perspective of the health system.

Reference	Telehealth modality and clinical focus	CHEERS <sup>a</sup> score (out of 20)	Findings in US \$ 2019	Initial investment in US \$ 2019	Reason for lower cost in the telehealth group
Kovács et al (2017) [17], Hungary	Store-and-forward system for screening for retinopathy of prematurity	18	Cost per examination for telehealth was less than in-person examination.	\$199,959.03 including equipment and implementation costs	Saved patient transport costs; saved working hours
Buyse et al (2008) [18], Belgium	Remote monitoring for high-risk pregnancy replacing extended hospital admissions	14	Cost reduction for remote monitoring of \$233,958 per year.	\$15,409.17	Saved admitted days. If an average of 8 patients were suitable for and accepted remote monitoring each month, an average of 14.7 admitted days could be replaced by remote monitoring
Xu et al (2008) [19], Australia	Videoconference for ear, nose and throat consultations	19	Costs of \$108 per consultation for telehealth versus \$155 for in-person consultation when caseload >100 consultations per year; saving realized despite a patient-end pediatrician cost to telehealth.	\$31,509.38	Saved patient and family travel
Armstrong et al (2007) [20], United States	Store-and-forward system for dermatology screening, diagnosis, and triage	15	Tele dermatology practice had an hourly operating cost of \$361 versus \$456 for conventional care	Not reported	When the patient-end is in a rural area (cheaper to rent space in those clinics)
Smith et al (2007) [21], Australia	Pediatric videoconference service for consultation reducing travel requirements	17	At caseload >774 cases/5 years telehealth is cost savings compared with in-person; \$598,203 saved over 5 years.	Not reported	Saved patient transport costs
Pare et al (2006) [22], Canada	Remote monitoring for patients with chronic obstructive pulmonary disorder by nurses, in place of regular home visits	13	Telehealth realizes \$361 in savings per patient or \$8566 total service cost savings compared with traditional in-home care program over 6-months (~\$13,713 per annum).	\$24,609.38	Reduced home visits by nurses (saved travel) and reduced salary for home-visit nurses (increased productivity) and reduced hospitalizations (secondary care usage)
Labiris et al (2005) [23], Greece	Multispecialty videoconference consultations service (mainly orthopedics and dermatology) in place of in-person consultations	13	Cost per consultation for telehealth \$327 versus \$333 for conventional care.	\$34,356.78	Saved transportation costs
Norum et al (2005) [24], Norway	Hybrid system for radiotherapy involving store-and-forward simulation planning and remote oncologist supervision via videoconference	15	At workload >9-12 patients, telehealth is less expensive when patient transport by air is required.	\$112,115.99	Avoided emergency transfers
Scuffham et al (2002) [25], United Kingdom (Scotland)	Generalist dentist videoconference with specialist dentists from a metropolitan center reducing the need for travel by patients or specialists	19	Teledentistry (\$233) is more expensive compared with outreach (\$156) but less expensive when compared in-person care (\$662) per patient treated.	Not reported	Saved patient travel costs/subsidy

Reference	Telehealth modality and clinical focus	CHEERS <sup>a</sup> score (out of 20)	Findings in US \$ 2019	Initial investment in US \$ 2019	Reason for lower cost in the telehealth group
Bjørvig et al (2002) [26], Norway	Store-and-forward system for diabetic retinopathy screening	17	At caseloads of >110, telehealth is cost saving. At workloads <110, telehealth is more expensive than conventional care; at very low workloads (n=20), telehealth is around 20 times more expensive than conventional care per consultation; at high workloads (n=200), telehealth costs around 67% of conventional care per consultation.	Not reported	Saved patient travel
Harno et al (2001) [27], Finland	Review and triage of orthopedic cases via videoconference	12	Telehealth was \$3954 (total service cost) less expensive than the traditional referral model.	Not reported	Triage by VC <sup>b</sup> decreased the number of in-person hospital visits
Bergmo et al (2000) [28], Norway	Store-and-forward system for dermatology screening, diagnosis, and triage	16	At caseload >195 patients per year, telehealth (\$96,042.79) costs less than hybrid out-reach/patient travel service as a whole (\$179,634.98), patient travel (\$333,568.03) or locally employed dermatologists (\$81,355.24); actual workload was 375 patients.	\$81355.24	Saved patient and/or clinician travel
Harno et al (2000) [29], Finland	Triage of specialist cases via email and/or videoconference	13	Telehealth is less expensive with saving of \$10,874 over 8 months for the service.	Not reported	Triage by email and/or VC decreased the number of hospital visits
McCue et al (1998) [30], United States	Review and triage of specialist cases (HIV, cardiology, and oral surgery) by videoconference	13	Net saving of \$22 per consultation using telemedicine.	Not reported	Main saving is from saved transport
McCue et al (1997) [31], United States	Review and triage of specialist cases (HIV, cardiology, and oral surgery) by videoconference	11	Telehealth was cost saving realizing total service cost saving of \$24,352 over the 7-month study period (~\$21,700 per annum) or cost per visit for telehealth (\$430) versus conventional care (\$835).	\$251,995.49	Transport savings and medical cost savings

<sup>a</sup>CHEERS: Consolidated Health Economic Evaluation Reporting Standards.

<sup>b</sup>VC: video consultation.

The most common situation where telehealth reduced health system costs reported in these studies was when it offsets patient or clinician travel funded or subsidized by the health system [17,19,21,25,26,28,30,31]. Hence, savings are most likely to be realized in the public health system, as it is unusual that other service models cover patient transport costs with the exception of the Department of Veterans Affairs. It is more likely that savings will be realized when patient travel is substituted with telehealth versus when clinician travel is substituted due to the volume, that is, saved transport cost for one clinician versus saved transport costs for many patients. One of the reviewed studies found telehealth to be less expensive than subsidized patient travel but more expensive than outreach clinics (where clinicians travel to outlying areas to provide consults) [25]. Prevention of emergency transfers was another way in which telehealth could contribute to reduced travel costs [24].

Other scenarios where potential savings can be realized include remote monitoring of high-risk pregnancies, reducing the need for in-hospital monitoring [18]. Remote monitoring in lieu of in-home visits also realized savings due to saved staff travel costs, the reduced salary of home-visit nurses, and the reduced number of hospitalizations that resulted from continual monitoring [22]. Triage, such as when a nurse screens a patient to determine if tertiary care is necessary, was also shown to save costs by reducing the number of hospital visits [30,31]. The two services that found telehealth to be more expensive than conventional care were due to the additional salary of a patient-end clinician (eg, general practitioner) attending a specialist consultation via videoconference [32,33].

In a number of CMA studies, telehealth was found to be cost saving as compared with standard care models only after a certain caseload was exceeded, known as a threshold or

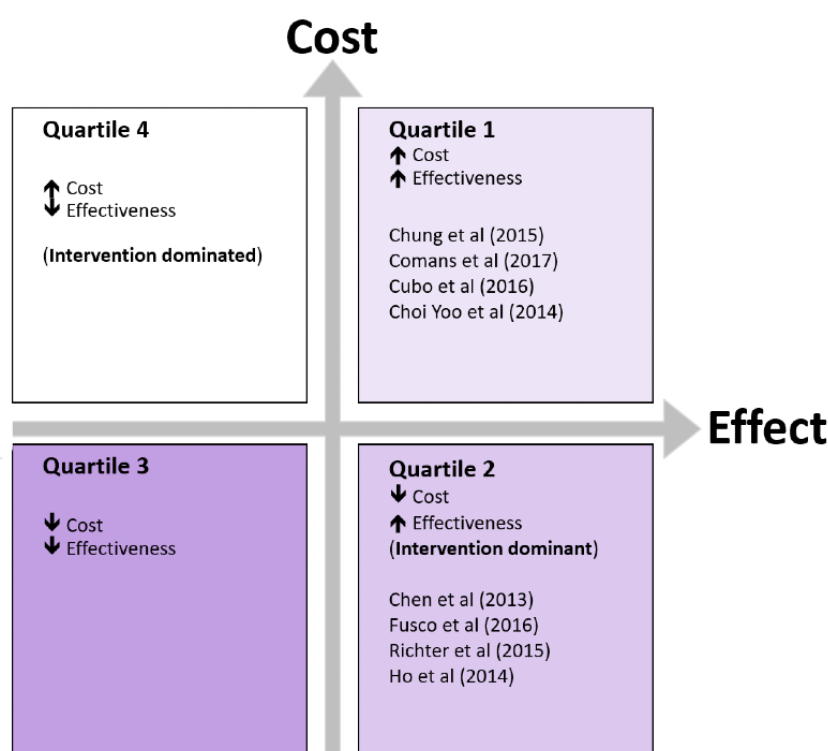


break-even point [19,21,24-26,28]. The break-even point was when the initial investment (typically in equipment setup and staff training) was offset by realized savings later on. Many studies calculated a break-even point; however, only one study reported the payback period [17]. The payback period ranged from near immediate to 9 years after implementation [34]. The payback period could be reduced if the activity was higher and the service had the capacity for increased activity [23,31]. Cost savings for telehealth have been slow to materialize for many early adopters due to the time lag between capital investments and the broader adoption of telehealth [35]. The marginal costs of a teleconsultation were found to be less than the marginal costs of conventional consultations, indicating that telehealth is likely to benefit from economies of scale [21].

### CEA

A total of 8 cost-effectiveness studies from the past 5 years that reported costs from the perspective of the health system were identified [36-43]. Of the 8 studies, 4 (50%) were in quartile 2 of the cost-effectiveness plane (Figures 2 and 3), meaning they demonstrated cost savings and an increased or equivalent clinical effectiveness for telehealth compared with conventional in-person care. These are also known as dominant strategies and are recommended for implementation. Two of these studies reported results from a remote monitoring intervention for heart failure [36,42]. The remaining studies (4/8, 50%) were in quartile 1 of the cost-effectiveness plane, meaning increased clinical effectiveness for increased costs, requiring a value judgment to be made.

**Figure 3.** Cost-effectiveness studies mapped on cost-effectiveness plane.



Moreover, 3 of the 4 cost-effectiveness studies in quartile 1 scored highly on quality as assessed using the CHEERS checklist [15], indicating comprehensive reporting of study results.

The telehealth services that were shown to reduce direct health system costs and be equally or more effective than their comparators were in smoking cessation [43], cardiovascular remote monitoring [36,42], and physiotherapy telerehabilitation after orthopedic surgery [41] (Table 3). Telehealth interventions

for cardiovascular remote monitoring [36,42] and postdischarge monitoring for neonates [44] have been shown to reduce hospital admissions, readmissions, and emergency department (ED) presentations, which have the potential to reduce overall costs to the health system [36,42,44]. In these studies, hospital events were used as a measure of the effectiveness of the telehealth intervention. The premise being that preventing hospital events would reduce costs for the health system if it was translated into a dollar value.

**Table 3.** Summary of cost-effectiveness studies that demonstrated lower health system costs in the telehealth model.

Reference	Telehealth modality and clinical focus	CHEERS <sup>a</sup> score	Effect measure	Effect improvement with telehealth?	Reason for lower cost in the telehealth group	Payback period
Chen et al (2013) [36], Taiwan	Remote biometric monitoring of patients with cardiovascular disease; out-of-range values trigger contact via phone from the clinical unit.	13	Hospital event rate	Yes	Reduced hospitalization, length of stay, and general medical costs to the health system when compared with similar patients without clinical monitoring and support.	Not calculable
Ho et al (2014) [42], Taiwan	Remote biometric monitoring of patients with cardiovascular disease; out-of-range values trigger contact via phone from the clinical unit.	21	Hospital event rate	Yes	Reduced hospitalization, length of stay, and general medical costs to the health system when compared with similar patients without clinical monitoring and support.	<1 year; however, due to the ongoing cost of remote monitoring, savings would need to continue at the same rate.
Fusco et al (2016) [41], Italy	Physiotherapy rehabilitation sessions delivered via videoconference to patients after orthopedic surgery.	24	Range of motion for relevant joints	Yes	Cost savings primarily due to reduced need for ambulatory government-funded travel when compared with in-person physiotherapy.	Not calculable
Richter et al (2015) [43], United States	Videoconference counselling sessions to support smoking cessation provided by primary care clinics.	17	Smoking cessation abstinence at 12 months	Equivalent	Compared with counselling provided over the phone, videoconference sessions were shorter and therefore cost less in staff wages.	Not calculable

<sup>a</sup>CHEERS: Consolidated Health Economic Evaluation Reporting Standards.

CEA studies use measures of effectiveness that reflect expected outcomes from the intervention. Many of these effects represent positive health gains and potentially medium- to long-term cost savings for the health system, such as avoided treatment of smoking-related diseases. However, the studies in which these effects have been demonstrated have not valued these gains in terms of cost.

## CUA

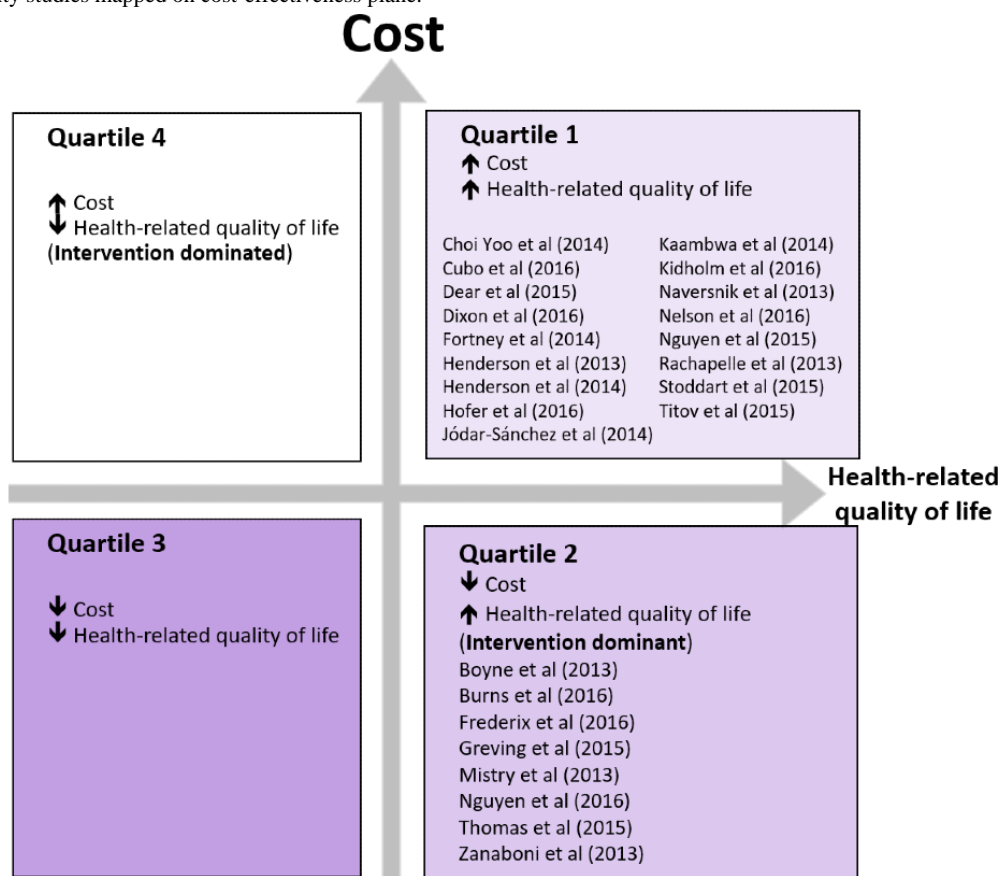
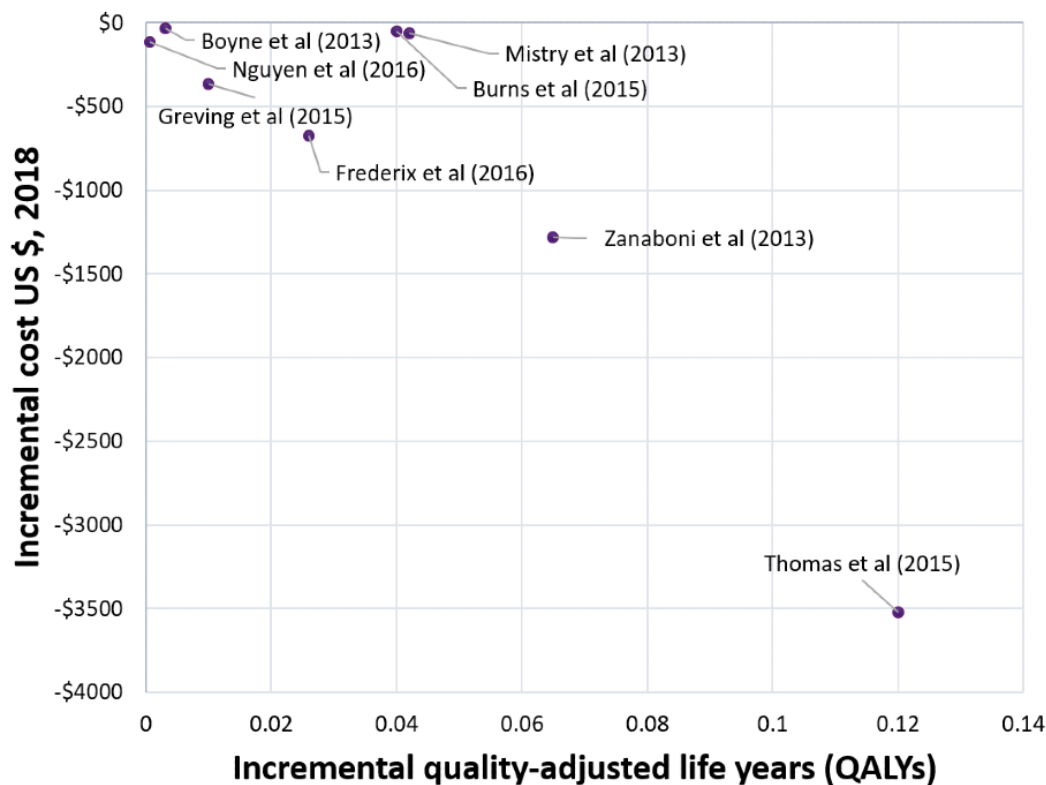
A total of 25 cost-utility studies from the past 5 years that reported costs from the perspective of the health system and changes in health-related quality of life (HRQoL) were identified [37,40,45-65]. Of these 25 studies, 8 (32%) studies were in quartile 2 of the health economics plane (Figures 2 and 4), as they demonstrated costs savings and increased or equivalent changes in effect as measured by HRQoL [45,47,49,50,57,60,65,66].

Telehealth interventions that save money and increase quality-adjusted life years (QALYs) should be considered for implementation. As shown in Figure 5, studies in the lower right quadrant of the plane (Q2) represent telehealth interventions that, when compared with usual care, represent a decrease in health care costs and a gain in QALYs. Each of these studies describes an intervention that, if implemented, could potentially save money for the health system.

The remaining studies (17.25, 68%) represent telehealth interventions that, when compared with usual care, represent an increase in health care costs and a gain in QALYs (quartile 1; Figure 4). Therefore, these interventions would only be implemented by decision makers if they were willing to pay for the relative increase in HRQoL.

All the cost-utility studies scored highly on quality as assessed using the CHEERS checklist [15], indicating the comprehensive reporting of study results.

An incidental finding of the review was that most telehealth interventions examined provided a small marginal improvement in the number of QALYs [37,40,45,46,48-68]. This small marginal change is likely due to the sensitivity of the instruments used to measure the quality of life and their ability to respond to the changes resulting from changing the service delivery model. Increases in QALYs for telehealth interventions compared with usual care were between 0.0006 and 0.12, irrespective of the clinical area or telehealth intervention type [37,40,45,46,48-68]. This is below what is considered clinically meaningful. Studies conducted from a health system perspective reported cost savings between US \$32 and US \$3523. All studies were located in the dominant quartile, demonstrating that the studies that reduced costs for the health system also demonstrated either equal or increased QALYs (Figure 5; Table 4).

**Figure 4.** Cost-utility studies mapped on cost-effectiveness plane.**Figure 5.** Quartile 2 incremental cost-utility values in 2019 US\$.

**Table 4.** Cost-utility analysis articles demonstrating lower cost from the perspective of the health system.

Reference	Telehealth modality and clinical focus	CHEERS <sup>a</sup> score	Cost (telehealth minus usual)	Utility (telehealth minus usual)	Cost (2019 US \$)	Health-related quality of life tool	Reason for lower cost in the telehealth group	Payback period
Boyne et al (2013) [45], Netherlands	In-home remote patient monitoring by a nurse. Patient's response to clinical questions aimed at identifying exacerbation of heart failure.	22	−31 (US \$, 2018)	0.0031	−31.71	EQ5D <sup>b</sup>	Reduction in in-person appointments and reduction in the use of physiotherapy services.	Not calculable
Frederix et al (2016) [49], Belgium	Remote monitoring for cardiovascular disease. Patients wear an accelerometer and receive feedback on their activity via email or SMS.	20	−564.4 (Euro, 2015)	0.026	−676.79	EQ5D	Reduction in rehospitalization costs.	Not calculable
Greving et al (2015) [50], Netherlands	Remote monitoring of vascular disease using patient-collected biometric information, with feedback from a remote nurse monitoring their data.	22	−219 (Euro, 2009)	0.01	−366.54	EQ5D	Reduction in paramedic support and hospital admissions.	Not calculable
Mistry et al (2013) [57], United Kingdom	Prenatal screening for congenital heart disease: store-and-forward images and videoconference consultations	23	−30 (UK £, 2009-10)	0.042	−60.1	Multiple literature sources	Economies of scale in telehealth screening compared with in-person screening.	Not calculable
Nguyen et al (2016) [60], Singapore	Store-and-forward diabetic retinopathy screening. Images captured by a nurse reviewed off-site, and the report is sent to the doctor.	20	−144 (Singapore \$, not specified)	0.0006	−114.65	Time trade off	Centralized image examination was lower cost when compared with distributed image examination. Additionally, the triage process reduced unnecessary referrals for appointments and procedures.	Not calculable
Thomas et al (2015) [66], United Kingdom	Store-and-forward ophthalmic images for glaucoma screening.	22	−3569.88 (Can \$, 2014)	0.12	−3523.28	Literature sources [69]	Reduction in travel and associated costs associated with travel (direct costs and staff time) and staff wages due to shorter appointments.	1-2 years for one site (saving per patient is \$3570 and current annual workload is 300 patients).
Zanaboni et al (2013) [65], Italy	Remote monitoring of biometric data from an implanted device to identify heart failure exacerbations.	19	−888.1 (Euro, 2010)	0.065	−1280.46	EQ-5D <sup>b</sup>	Substituting in-person clinic visits with lower cost virtual consults, and reduction in the emergency department and urgent clinic visits.	1-2 years

Reference	Telehealth modality and clinical focus	CHEERS <sup>a</sup> score	Cost (telehealth minus usual)	Utility (telehealth minus usual)	Cost (2019 US \$)	Health-related quality of life tool	Reason for lower cost in the telehealth group	Payback period
Burns et al (2016) [47], Australia	Videoconference speech pathology for rural patients with head and neck cancer.	17	−59 (Aus \$, 2015)	Equal	−48.94	Assessment of Quality of Life Instrument (4 dimensions)	Reduction in travel and associated costs and staff wages due to shorter appointments.	7-8 years for one site (saving per patient is \$59 and current annual workload is 82 patients) plus 30 patients on average to cover annual levy.

<sup>a</sup>CHEERS: Consolidated Health Economic Evaluation Reporting Standards

<sup>b</sup>EQ5D: EuroQol five dimensions

## Focus Group

The expert panel focus group ran for 1 hour, and 7 experts attended from 5 institutions. Two further experts from 2 different organizations were unable to attend and provided input via email or telephone instead. They covered telehealth expertise in academic research, health service provision, and economics.

Seven experts who were contacted did not respond. The videoconference session was recorded so that the investigators could review the comments during the writing of this report.

At the conclusion of the expert panel, the telehealth domains where health system costs could potentially be reduced were finalized. These domains are listed in [Table 5](#).

**Table 5.** Domains identified from expert feedback.

Domain	Description	Level of evidence [16]
Productivity gains	Optimization of staff time leading to productivity gains.	Level II-IV
Secondary care resource use	There is potential for telehealth to reduce secondary care resource (eg, emergency department presentation, hospitalization, and medical imaging) use with associated costs savings.	Level I-III
Alternative funding models	The commercialization of telehealth has resulted in direct-to-consumer models of care where patients pay for their services directly, rather than accessing subsidized care.	Evidence in this emerging field does not map to any NHMRC <sup>a</sup> levels
Telementoring	Telementoring of primary care can increase the skill level of clinicians, thereby reducing future referrals to specialists for similar cases.	Evidence in this emerging field does not map to any NHMRC levels

<sup>a</sup>NHMRC: National Health and Medical Research Council.

## Productivity Gains

The first domain identified by the expert focus group was productivity gains, which is when the productivity of a system is increased, cost savings can be realized due to the increased capacity of the system. In the case of telehealth, it is possible that a greater volume of patients can be managed with similar resources, which increases the productivity of the health system and reduces the marginal cost per patient. Evidence of increasing physician productivity with the use of telehealth exists in various other forms, including case control, RCT, CMA, and case series. Telehealth has the potential to increase health system productivity through a number of mechanisms, including reducing travel, reducing consultation time, and substituting for in-person service modalities.

## Reduction in Travel

When telehealth reduces or eliminates clinician travel time because service delivery occurs through videoconference, the

clinicians' productivity is increased because they can see more patients in the same time frame. One program in the United States that implemented a videoconference telehealth model to replace home visits found that nurses' caseload capacity more than doubled and over 14 months, 43,560 driving minutes were saved [70]. Similarly, an economic evaluation of The Northern Health Authority in Canada found that telehealth sessions replacing in-person sessions saved the health service Can \$65,520 (US \$49,584.14) in annual travel costs associated with clinical sessions [71].

## Change in Consultation Time

When in-person consultations are substituted for a video consultation of a shorter duration, the clinic has an increased capacity to see more patients in the same amount of time.

Studies that compared videoconference consultation time with in-person consultation time reported disparate results, finding that videoconference was less time efficient, equally time



efficient, and more time efficient. One study reported that teleconsultations may take more time than in-person appointments in cases such as assessing injuries by video as opposed to on-site [72]. Four studies reported that videoconference was equally as time efficient as in-person consultations. These included studies reporting results from studies examining services in dermatology [73], prostate cancer [74], pulmonary medicine [75], and orthopedics [76]. However, there is also evidence to suggest that videoconference consultations were more time efficient, as shown in a dermatology service from Norway [76,77] and a diabetes, antenatal, and cancer care service by Greenhalgh et al [77,78]. The small gains in clinician efficiency may be offset by increased administrative overhead associated with telehealth compared with in-person consultations [79].

### **Consultation Mode Substitution**

If in-person consultations are substituted for alternate consultation modalities such as asynchronous consultations (store-and-forward or virtual), productivity is often increased as more patients are able to be managed simultaneously [5,27,80-83]. Asynchronous consultations are when the patient and clinician are not localized to the same time point. Asynchronous consultations, therefore, represent a mechanism by which the overall productivity of the health system can be improved. For instance, a specialist may be able to review clinical notes and images for a large group of patients in lesser time than it would take them to see all the patients in person [29]. Such services include dermatology consultations where either a patient or primary carer sends clinical information and dermoscopic or regular images for review; patients are then either returned to their primary care, discharged from care, or scheduled for an in-person dermatologist appointment [5,80,81,84]. Consultations are typically quicker when they are asynchronous, for example, one study found that an asynchronous dermatology consultation takes 4 min, which is one tenth of the time for a traditional consultation [85]. This increases the system throughput and can optimize both waitlists and patient prioritization. Similarly, a case-control study showed that clinicians who used a web-based messaging platform to manage patients had a 10% increase in productivity compared with those who conducted in-person consultations and followed up with phone calls [86]. The increase in productivity resulted in an additional 2.54 patients per day being seen, which in a fixed funding model would lead to a reduction in marginal cost per patient.

A study by Liddy et al [82] demonstrated that when community clinicians could send asynchronous consultations to specialists, compared with the previous system where 50% of these cases would have resulted in the referral of the patient for an in-person consultation, only 18% of cases required an in-person consultation. Another recent study showed that 68% of in-person appointments were unnecessary when an asynchronous teledermatology triage model was used [86,87].

### **Failure to Attend**

When patients fail to attend appointments, it costs the health system money by increasing the marginal cost of all appointments in that service. Additionally, it represents an

opportunity cost as an alternative patient has to forego an appointment. One study postulated that when patients are treated by primary care providers in their own community, the chance of them missing the appointment may be less likely than if they were traveling to a hub site further from where they live [88].

### **Payment Model**

The reimbursement model needs to be considered when examining the reported productivity increases, as the marginal cost per patient will only decrease if the service provision costs remain constant as patient volume increases. A more productive clinician who is able to manage a larger number of patients will increase the cost to the health system under activity-based funding, fee-for-service, or a capitation reimbursement model. However, when clinician costs are fixed (eg, salaried), increased productivity will reduce the marginal cost per patient.

### **Summary**

Telehealth can increase clinician productivity, thereby increasing the volume of patients that can be managed by a health service. Assuming fixed costs, this can result in a reduced marginal cost per patient overall. Telehealth can enable a clinician to convert travel time to clinical time, thereby improving productivity. However, when a clinician does not have to travel, substituting in-person consultations with video consultations is unlikely to have a major impact on consultation time and resultant productivity and savings. Furthermore, the increased administrative overhead for scheduling video consultations may counteract any gains. Increased productivity is more likely to be achieved in store-and-forward and virtual consultations. Realizing savings from productivity gains is dependent on the funding model. Increased productivity under activity-based funding, fee-for-service, or a capitation reimbursement model will increase the cost for the health system.

### **Secondary Care Resource Use**

Secondary care involves services provided by specialist or tertiary centers and includes hospital admissions, specialist outpatient visits, and ED presentations. Avoidance of secondary care in favor of other methods of care has the potential to reduce health system costs [89]. The evidence was from RCTs, case-control prospective and observational studies, and reviews of health services. Most of the studies, except the CMA conducted by Pare et al [22], reported economic findings as a secondary result. This resulted in moderate quality studies examining changes in secondary care usage related to telehealth, but a level of extrapolation from the results was needed to interpret cost savings. The use of telehealth to reduce secondary care may be realized through a number of scenarios, including remote monitoring, hospital avoidance, and triage.

### **Remote Monitoring**

Remote monitoring, or telemonitoring, is an established modality of telehealth where patients are monitored from a distance. Remote monitoring is used most often to monitor chronic diseases (eg, hypertension, cardiac disease, pulmonary disease, and diabetes). Remote monitoring involves the continual in-home recording of targeted biometric readings (eg, blood pressure, glucose levels, weight, and spirometry) and in some services, patient-reported measures (eg, level of breathlessness).

The readings are subsequently transmitted to a clinician for review. Remote monitoring may be performed in conjunction with in-person consults [7] or video or audio conference consults [90]. The aim of remote monitoring is the early detection and management of exacerbations, which may obviate an ED presentation or hospital admission.

Findings on secondary care usage resulting from remote monitoring usage are mixed. Telemonitoring in France, the United Kingdom, the United States, and Australia have demonstrated a reduction in hospitalization resulting directly from the use of remote monitoring compared with patients who were not monitored remotely [91-94]. In the UK trial, remote monitoring of patients with chronic disease (eg chronic obstructive pulmonary disorder, diabetes, or heart failure) was associated with an overall reduction in hospital admissions [93]. These results were consistent at three trial sites using a variety of remote monitoring technologies, such as pulse oximeters, glucometers, and weighing scales [93]. Similarly, in the United States, when patient vital signs were transmitted to a nurse for review and intervention, rates of acute care hospitalization (1.7 vs 4.4 per 1000 home health days) and ED presentations (1.9 vs 5.3 per 1000 patient days) reduced [94]. In Australia, the study by Celler et al [92] found that when using remote monitoring, clinicians were able to predict and avoid 53% of admissions by conducting a low-cost intervention in a timely manner. Remote monitoring has been shown to reduce not only a patient's presentation to the hospital but also their length of stay once admitted. This may be due to the confidence remote monitoring gives clinicians, that when they discharge a patient, the patient is still under observation should any acute needs arise. For example, a hospital in Belgium monitored high-risk mothers at home instead of keeping them in the hospital [18]. There are, however, some studies that demonstrate in certain scenarios telemonitoring does not reduce secondary care use but, in fact, can increase secondary care usage [83,95].

When reduced secondary care usage is achieved, it would logically convert into the reduced cost for the health system. However, the economic analysis of remote monitoring does not conclusively report savings. Some studies report significant cost savings [18,22,90,96]. Other studies have found very small savings. Lew et al [97] found that for some patient groups, admission costs only reduced from US \$10,835 to US \$10,678. In the Whole System Demonstrator trial, cost savings amounted to a modest £188 per person per year [93]. Other studies report that remote monitoring resulted in equivalent or greater costs [97,98]. Many economic analyses of remote monitoring only report direct health care costs and do not report overall program costs such as amortization of equipment costs or the cost of running the service [7]. Other contextual factors may also influence the findings. For example, monitoring a single vital sign is less costly than monitoring multiple signs, and remote monitoring of hypertension and congestive cardiac failure is less costly than remote monitoring of respiratory diseases [7].

### **Hospital Avoidance**

Telehealth can be used to facilitate hospital avoidance, which can potentially reduce costs, particularly in reducing ED presentations. Emergency Medical Services in Houston, United

States, implemented a system where after an ambulance was dispatched and before the patient was transported to the ED, they conducted a videoconference with a physician [99]. Where appropriate, the patient was directed to primary care, the ED via a taxi or personal travel means or via ambulance as necessary [99]. Transports to the ED by ambulance were significantly reduced by over 50%, and the team was back in service for the next call 44 min faster [99]. A similar service in the context of residential aged care reported that the use of telehealth before transportation reduced ED presentations by 28% [100]. However, neither of these studies quantified savings.

### **Triage**

Similar to findings (previously described) from the economic analyses reviewed, a recent review identified that teletriage could reduce a substantial number of unnecessary specialist outpatient appointments. The review found that for dermatology, the reported rate of avoided in-person appointments ranged from 38% to 88% and for ophthalmology ranged from 16% to 48%. Single studies for ear, nose, and throat and vascular surgery/wound care reported an 89% and 18% reduction in in-person appointments, respectively [101]. However, no study has quantified the potential cost savings.

### **Summary**

Telehealth appears to have the potential to reduce secondary health care usage. However, while many studies demonstrate a reduction in secondary care, there are limited studies that quantify cost savings for the health system payer. Cost analyses do not always consider the overall costs of telehealth interventions and instead only compare the costs associated with direct health care utilization. A more accurate assessment would include program costs such as amortization of equipment costs or the cost of running the service. The use of telehealth for triage can reduce unnecessary specialist outpatient appointments. Although this would logically reduce costs for a health care provider, no study has quantified cost savings.

### **Alternative Funding Models**

Direct-to-consumer telehealth services are often funded by consumer payments. User pays funding models can potentially reduce costs, particularly if they substitute for government-funded or government-subsidized health services. Hence, savings to the health system are based on the assumption that individuals who access direct-to-consumer services would have otherwise accessed an equivalent health system service if the telehealth option was unavailable to them. In Australia, very few direct-to-consumer services are eligible for reimbursement from the Medicare Benefits Scheme. Consumers accessing these services are required to pay the full service fee as an out-of-pocket cost. This leads to cost savings for the government when consumers access these services instead of Medicare-funded services. Users may be willing to pay a higher out-of-pocket cost for these services because they offer elements that they value, for instance, either convenience or timely specialist access. However, due to the convenience of telehealth services, there is a risk that direct-to-consumer services may not reduce health system usage, but rather may increase overall health service utilization and costs to the health system [102].

That is, individuals who would not normally use health services may begin accessing telehealth services because of the convenience provided by the direct-to-consumer model. When these individuals are referred to a health system service provider or given a Pharmaceutical Benefits Scheme reimbursed prescription, they increase service utilization and costs to the overall health system.

There is limited academic literature that reports on savings to the health system from direct-to-consumer telehealth. Examples of commercial direct-to-consumer telehealth models include Qoctor (previously called DrSicknote), which is an Australian web-based general practitioner service. Qoctor provides a range of general practitioner services through their website, including medical certificates, prescriptions, and specialist referral letters. In January 2019, Qoctor reported \$1,040,566 in saved costs to Medicare to date by diverting appropriate patients from standard general practitioner clinics [103]. A further example is iDoc24, where consumers capture images of skin lesions (eg, a mole), transmit the image to a dermatologist, and receive a diagnosis within 24 hours [104].

### Summary

At this stage, it is difficult to quantify the cost outcomes of funding models where a consumer pays out of pocket for a commercial telehealth service. Most information is reported by the companies themselves and not through research studies. Assuming that individuals who choose to access commercial telehealth services would have instead accessed government-funded services, commercial services can reduce health system costs. As these funding models are still new, it is difficult to quantify the effect or anticipate all consequences.

### Telementoring

Deferring treatment to less qualified and, therefore, less expensive staff could potentially result in reduced health system costs. This can be facilitated by telementoring. Arguably, the most well-known telementoring program is the Project Extension for Community Healthcare Outcomes (Project ECHO), a model that was started in the United States but has been adopted and practiced internationally [105]. Using this model, primary care staff are upskilled using videoconference sessions with specialists in the form of weekly telementoring sessions. During these sessions, primary care providers can present cases and receive specialist advice on diagnosis, management, and treatment for their patients.

Although telementoring was identified by the focus group as a way to potentially reduce costs, limited evidence was found to support this view. A pre- and poststudy found that telementoring in the context of geriatric mental health resulted in a small reduction in per-patient cost when medication, specialist outpatient visits, hospitalization, and ED visits were quantified from insurance claim data [106]. Although overall costs were reduced, a subanalysis did reveal increased costs resulting from an increase in antipsychotic medication prescriptions. The time horizon for reported cost reduction was 6 months. The cost of setting up and running the telementoring was not considered in this analysis.

Project ECHO, a program that provides telementoring for primary care physicians, for the management of hepatitis C, was found to be cost-effective with an incremental cost-effectiveness ratio (ICER) of US \$10,351 per QALY [107]. This would indicate that telementoring for hepatitis C will increase costs to the health system but has the potential to increase population-related quality of life.

From another perspective, savings to the health system as a result of telementoring may be realized in the retention of staff and patients at remote medical practices [88].

### Summary

There is limited evidence due to only a very small number of studies analyzing the costs of telementoring. Telementoring can potentially reduce health system costs in both the short term and over a longer time horizon. However, at present, evidence to support this is lacking.

## Discussion

### Principal Findings

Telehealth was shown to reduce costs to the health system in the short to medium term in 53% of CMA, 50% of CEA, and 32% of CUA studies reviewed. The predominant reason for reduced costs was when the health system funded travel and either patient or clinician travel was reduced or avoided. In the remaining studies (not reviewed in detail here), telehealth increased costs but was also shown to improve care [37-40,42-44,46,48,51,52,54-56,58,59,61-64]. For example, evidence indicates that remote patient monitoring is currently a poor cost minimizer; however, it is very effective for improving overall health and reducing morbidity and hospitalization.

The models of the care and the contexts in which telehealth is used are heterogeneous. The question as to whether telehealth decreases the cost of health care delivery is complex, as are all economic evaluations in telehealth [7]. There are many compounding factors to consider, for example, the modality (real-time videoconferencing, remote patient monitoring, and store-and-forward) of telehealth used or the way telehealth consultations are remunerated. In Australia, Medicare reimburses provider telehealth consultations at 150% of an equivalent in-person consultation [108]. Furthermore, reimbursement for telehealth in Australia under Medicare and activity-based funding have both provider and patient-end payments, automatically making telehealth more expensive than conventional care. Although in the United States, state-by-state reimbursement under Medicaid means a telehealth consultation may be reimbursed at a lower or equal rate to an equivalent in-person consultation.

By improving accessibility, telehealth may also increase the cost of providing health care as populations served by traditional models of care who had limited or no access to care can now access services. Further, by improving the convenience of access, there is potential for excess use [109,110]. Furthermore, the potential for telehealth to become adjunctive to traditional in-person care (rather than substitutive) also increases the potential for increasing use and associated costs.



Telehealth is not implemented at scale in some jurisdictions (eg, Australia) [79,111], which may impact the ability to reduce costs. The marginal cost or average cost of a telehealth consultation was found to be less than an equivalent in-person consultation in a number of studies [19,32]. However, costs of running the telehealth service remained higher than the equivalent in-person service until a threshold number of consultations was reached. A threshold number of examinations required is necessary to counteract the implementation costs of telehealth, including technology and project costs. Telehealth technology costs are decreasing. Video conference equipment costs are reducing with a move from room-based systems to consumer videoconferencing platforms. A similar cost decrease of remote monitoring equipment and resultant cost decrease in remote monitoring costs per patient per year has also been reported [7]. Future studies may find a reduction in both the threshold number of examinations and the payback period resulting from a decrease in technology costs.

### Implications for Practice

Many telehealth services do not result in short- to medium-term savings. For this reason, health services considering implementing telehealth should be motivated by the benefits of telehealth other than cost reduction and recognize that implementing telehealth will require recurrent investment costs.

Asynchronous consultations are likely to reduce consultation time compared with equivalent in-person consultations, thereby improving clinician productivity. There is potential to increase the cost to the health system under fee-for-service models; hence, reimbursement for consultations where the patient is not present needs to account for increased productivity. One such example is a recent application from the Australasian College of Dermatologists for Medicare funding for store-and-forward dermatology, which sought a lower reimbursement than in-person consultations [112].

### Implications for Research

This study has identified a number of opportunities for future research. Many of the reviewed economic analyses of telehealth are older than 10 years. Contemporary economic analysis of telehealth is needed to consider changes in technology costs. Our study has also identified a gap in economic analyses related to teleremoting. Furthermore, the use of telehealth to triage referrals in dermatology, wound care, otolaryngology, and ophthalmology can be effective in reducing unnecessary specialist outpatient appointments; however, there are no published cost analyses. Additionally, the use of remote patient monitoring has been shown to reduce costs and increase costs in different contexts. Many remote monitoring studies did not report overall cost savings (eg, the cost of implementing and running the service). Analyses of remote monitoring services need to include overall costs rather than direct health costs alone to determine if remote patient monitoring reduces system costs.

### Limitations

Positive reporting bias is likely to result in an increased number of studies reporting cost-effectiveness and cost-minimization

of telehealth. Savings on travel resulting from telehealth are more beneficial to the health system in countries such as Australia. The large geographical areas and a public health system that funds travel favor such contexts. For this reason, the findings of cost savings are rarely generalizable. Additionally, it is unlikely that the studies examined were statistically powered for economic findings, as it is routine practice to power for the primary clinical outcome, meaning economic outcomes may not be precise. The aim of this study was to identify short- to medium-term cost savings; therefore, a further limitation of this study is that studies examining the long-term cost impact of telehealth interventions were not included.

### Conclusions

This study aimed to determine whether telehealth has the potential to improve the sustainability of health systems by reducing costs. Telehealth was shown to reduce costs to the health system in the short to medium term in 53%, 50%, and 32% of the cost-minimization, cost-effectiveness, and cost-utility studies reviewed, respectively. The predominant reason for reduced costs was when the health system-funded travel (patient or clinician) was reduced or avoided. In the remaining studies, telehealth increased costs, albeit with improved care.

The expert focus group identified 4 areas of potential savings from telehealth: productivity gains, reductions in secondary care use, emerging alternate funding models for care provision, and savings resulting from teleremoting effects. Telehealth can increase clinician productivity when it is used to convert travel time to clinical time. In terms of consultation time, there are unlikely to be productivity gains when substituting an in-person consultation with a video consultation. The use of asynchronous consultations as a substitute for in-person consultations is more likely to increase productivity by reducing consultation time. However, under activity-based funding mechanisms, there is a likelihood that these productivity gains could result in cost increases.

Mitigation of secondary care through remote patient monitoring, teleremote, and hospital avoidance has the potential to reduce costs to the provider. However, there is currently a lack of economic evidence to support this. Similarly, teleremoting has scant economic evaluations to demonstrate cost savings.

Alternate funding models from telehealth systems have the potential to save the health system money in situations where the consumers pay fully out of pocket to receive services, thereby mitigating the cost to the health system. The convenience of telehealth may influence consumers to pay out-of-pocket fees. This may be considered cost-shifting as opposed to cost saving.

The available evidence has indicated that telehealth does not always reduce the cost of care from the perspective of the health system in the short to medium term. Health services considering implementing telehealth should be motivated by benefits other than cost reduction.

## Acknowledgments

This research was conducted for the NHMRC Partnership Centre for Health System Sustainability (Grant ID #: 9100002) administered by the Australian Institute of Health Innovation, Macquarie University. Along with the NHMRC, the funding partners in this research collaboration are: The Bupa Health Foundation; NSW Ministry of Health; Department of Health, WA; and The University of Notre Dame Australia. Their generous support is gratefully acknowledged.

While the NHMRC, The Bupa Health Foundation, NSW Ministry of Health, Department of Health, WA and The University of Notre Dame Australia, have provided in-kind and financial support for this research, they have not reviewed the content and are not responsible for any injury, loss or damage however arising from the use of, or reliance on, the information provided herein. The published material is solely the responsibility of the authors and does not reflect the views of the NHMRC or its funding partners

## Conflicts of Interest

None declared.

## References

1. Australian Institute of Health and Welfare. Health expenditure Australia 2016-17. AIHW. Canberra: Australian Government; 2018 Sep 28. URL: <https://www.aihw.gov.au/reports/health-welfare-expenditure/health-expenditure-australia-2016-17/contents/data-visualisation> [accessed 2018-12-31] [WebCite Cache ID [www.aihw.gov.au/reports/health-welfare-expenditure/health-expenditure-australia-2016-17/contents/data-visualisation](http://www.aihw.gov.au/reports/health-welfare-expenditure/health-expenditure-australia-2016-17/contents/data-visualisation)]
2. Centers for Medicare and Medicaid Services. National Health Expenditure Data. CMS. 2018. URL: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html> [accessed 2019-11-26]
3. Bilinski A, Neumann P, Cohen J, Thorat T, McDaniel K, Salomon JA. When cost-effective interventions are unaffordable: integrating cost-effectiveness and budget impact in priority setting for global health programs. *PLoS Med* 2017 Oct;14(10):e1002397 [FREE Full text] [doi: [10.1371/journal.pmed.1002397](https://doi.org/10.1371/journal.pmed.1002397)] [Medline: [28968399](https://pubmed.ncbi.nlm.nih.gov/28968399/)]
4. Agha Z, Weinstein RS, Dunn BE. Cost minimization analysis of telepathology. *Am J Clin Pathol* 1999 Oct;112(4):470-478. [doi: [10.1093/ajcp/112.4.470](https://doi.org/10.1093/ajcp/112.4.470)] [Medline: [10510670](https://pubmed.ncbi.nlm.nih.gov/10510670/)]
5. Snoswell C, Finnane A, Janda M, Soyer HP, Whitty JA. Cost-effectiveness of store-and-forward teledermatology: a systematic review. *JAMA Dermatol* 2016 Jun 01;152(6):702-708. [doi: [10.1001/jamadermatol.2016.0525](https://doi.org/10.1001/jamadermatol.2016.0525)] [Medline: [27074289](https://pubmed.ncbi.nlm.nih.gov/27074289/)]
6. Bergmo TS. Using QALYs in telehealth evaluations: a systematic review of methodology and transparency. *BMC Health Serv Res* 2014 Aug 3;14:332 [FREE Full text] [doi: [10.1186/1472-6963-14-332](https://doi.org/10.1186/1472-6963-14-332)] [Medline: [25086443](https://pubmed.ncbi.nlm.nih.gov/25086443/)]
7. Peretz D, Arnaert A, Ponzoni NN. Determining the cost of implementing and operating a remote patient monitoring programme for the elderly with chronic conditions: a systematic review of economic evaluations. *J Telemed Telecare* 2018 Jan;24(1):13-21. [doi: [10.1177/1357633X16669239](https://doi.org/10.1177/1357633X16669239)] [Medline: [27650163](https://pubmed.ncbi.nlm.nih.gov/27650163/)]
8. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. *Telemed J E Health* 2015 Feb;21(2):81-85 [FREE Full text] [doi: [10.1089/tmj.2014.0053](https://doi.org/10.1089/tmj.2014.0053)] [Medline: [25474190](https://pubmed.ncbi.nlm.nih.gov/25474190/)]
9. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005 Feb 20;8(1):19-32 [FREE Full text] [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
10. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci* 2010 Sep 20;5:69 [FREE Full text] [doi: [10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69)] [Medline: [20854677](https://pubmed.ncbi.nlm.nih.gov/20854677/)]
11. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018 Oct 2;169(7):467-473. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
12. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. Methods for the economic evaluation of health care programmes. New York: Oxford University Press; 2015.
13. Snoswell C, Smith AC, Scuffham PA, Whitty JA. Economic evaluation strategies in telehealth: obtaining a more holistic valuation of telehealth interventions. *J Telemed Telecare* 2016 Oct 27;23(9):792-796. [doi: [10.1177/1357633x16671407](https://doi.org/10.1177/1357633x16671407)] [Medline: [27789615](https://pubmed.ncbi.nlm.nih.gov/27789615/)]
14. Robinson R. Costs and cost-minimisation analysis. *Br Med J* 1993 Sep 18;307(6906):726-728 [FREE Full text] [doi: [10.1136/bmj.307.6906.726](https://doi.org/10.1136/bmj.307.6906.726)] [Medline: [8401098](https://pubmed.ncbi.nlm.nih.gov/8401098/)]
15. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, CHEERS Task Force. Consolidated health economic evaluation reporting standards (CHEERS) statement. *Cost Eff Resour Alloc* 2013 Mar 25;11(1):6 [FREE Full text] [doi: [10.1186/1478-7547-11-6](https://doi.org/10.1186/1478-7547-11-6)] [Medline: [23531194](https://pubmed.ncbi.nlm.nih.gov/23531194/)]



16. Merlin T, Weston A, Tooher R. Extending an evidence hierarchy to include topics other than treatment: revising the Australian 'levels of evidence'. *BMC Med Res Methodol* 2009 Jun 11;9:34 [FREE Full text] [doi: [10.1186/1471-2288-9-34](https://doi.org/10.1186/1471-2288-9-34)] [Medline: [19519887](https://pubmed.ncbi.nlm.nih.gov/19519887/)]
17. Kovács G, Somogyvári Z, Maka E, Nagyjánosi L. Bedside ROP screening and telemedicine interpretation integrated to a neonatal transport system: economic aspects and return on investment analysis. *Early Hum Dev* 2017;106-107:1-5. [doi: [10.1016/j.earlhumdev.2017.01.007](https://doi.org/10.1016/j.earlhumdev.2017.01.007)] [Medline: [28171806](https://pubmed.ncbi.nlm.nih.gov/28171806/)]
18. Buysse H, De Moor G, Van Maele G, Baert E, Thienpont G, Temmerman M. Cost-effectiveness of telemonitoring for high-risk pregnant women. *Int J Med Inform* 2008 Jul;77(7):470-476. [doi: [10.1016/j.ijmedinf.2007.08.009](https://doi.org/10.1016/j.ijmedinf.2007.08.009)] [Medline: [17923433](https://pubmed.ncbi.nlm.nih.gov/17923433/)]
19. Xu CQ, Smith AC, Scuffham PA, Wootton R. A cost minimisation analysis of a telepaediatric otolaryngology service. *BMC Health Serv Res* 2008 Feb 04;8:30 [FREE Full text] [doi: [10.1186/1472-6963-8-30](https://doi.org/10.1186/1472-6963-8-30)] [Medline: [18241356](https://pubmed.ncbi.nlm.nih.gov/18241356/)]
20. Armstrong AW, Dorer DJ, Lugn NE, Kvedar JC. Economic evaluation of interactive teledermatology compared with conventional care. *Telemed J E Health* 2007 Apr;13(2):91-99. [doi: [10.1089/tmj.2006.0035](https://doi.org/10.1089/tmj.2006.0035)] [Medline: [17489695](https://pubmed.ncbi.nlm.nih.gov/17489695/)]
21. Smith AC, Scuffham P, Wootton R. The costs and potential savings of a novel telepaediatric service in Queensland. *BMC Health Serv Res* 2007 Mar 02;7:35 [FREE Full text] [doi: [10.1186/1472-6963-7-35](https://doi.org/10.1186/1472-6963-7-35)] [Medline: [17331259](https://pubmed.ncbi.nlm.nih.gov/17331259/)]
22. Paré G, Sicotte C, St-Jules D, Gauthier R. Cost-minimization analysis of a telehomecare program for patients with chronic obstructive pulmonary disease. *Telemed J E Health* 2006 Apr;12(2):114-121. [doi: [10.1089/tmj.2006.12.114](https://doi.org/10.1089/tmj.2006.12.114)] [Medline: [16620165](https://pubmed.ncbi.nlm.nih.gov/16620165/)]
23. Labiris G, Tsitlakidis C, Niakas D. Retrospective economic evaluation of the Hellenic Air Force Teleconsultation Project. *J Med Syst* 2005 Oct;29(5):493-500. [doi: [10.1007/s10916-005-6106-4](https://doi.org/10.1007/s10916-005-6106-4)] [Medline: [16180485](https://pubmed.ncbi.nlm.nih.gov/16180485/)]
24. Norum J, Bruland Ø, Spanne O, Bergmo T, Green T, Olsen DR, et al. Telemedicine in radiotherapy: a study exploring remote treatment planning, supervision and economics. *J Telemed Telecare* 2005;11(5):245-250. [doi: [10.1258/1357633054471858](https://doi.org/10.1258/1357633054471858)] [Medline: [16035967](https://pubmed.ncbi.nlm.nih.gov/16035967/)]
25. Scuffham PA, Steed M. An economic evaluation of the Highlands and Islands Teledentistry Project. *J Telemed Telecare* 2016 Nov 29;8(3):165-177. [doi: [10.1177/1357633x0200800307](https://doi.org/10.1177/1357633x0200800307)]
26. Björvig S, Johansen MA, Fossen K. An economic analysis of screening for diabetic retinopathy. *J Telemed Telecare* 2002 Feb 01;8(1):32-35. [doi: [10.1258/1357633021937433](https://doi.org/10.1258/1357633021937433)]
27. Harno K, Arajärvi E, Paavola T, Carlson C, Viikinkoski P. Clinical effectiveness and cost analysis of patient referral by videoconferencing in orthopaedics. *J Telemed Telecare* 2001;7(4):219-225. [doi: [10.1258/1357633011936435](https://doi.org/10.1258/1357633011936435)] [Medline: [11506757](https://pubmed.ncbi.nlm.nih.gov/11506757/)]
28. Bergmo TS. A cost-minimization analysis of a realtime teledermatology service in northern Norway. *J Telemed Telecare* 2000 Oct 01;6(5):273-277. [doi: [10.1258/1357633001935905](https://doi.org/10.1258/1357633001935905)]
29. Harno K, Paavola T, Carlson C, Viikinkoski P. Patient referral by telemedicine: effectiveness and cost analysis of an Intranet system. *J Telemed Telecare* 2000;6(6):320-329. [doi: [10.1258/1357633001935996](https://doi.org/10.1258/1357633001935996)] [Medline: [11265100](https://pubmed.ncbi.nlm.nih.gov/11265100/)]
30. McCue MJ, Mazmanian PE, Hampton CL, Marks TK, Fisher EJ, Parpart F, et al. Cost-minimization analysis: a follow-up study of a telemedicine program. *Telemed J* 1998;4(4):323-327. [doi: [10.1089/tmj.1.1998.4.323](https://doi.org/10.1089/tmj.1.1998.4.323)] [Medline: [10220472](https://pubmed.ncbi.nlm.nih.gov/10220472/)]
31. McCue MJ, Mazmanian PE, Hampton C, Marks TK, Fisher E, Parpart F, et al. The case of Powhatan Correctional Center/Virginia Department of Corrections and Virginia Commonwealth University/Medical College of Virginia. *Telemed J* 1997;3(1):11-17. [doi: [10.1089/tmj.1.1997.3.11](https://doi.org/10.1089/tmj.1.1997.3.11)] [Medline: [10166440](https://pubmed.ncbi.nlm.nih.gov/10166440/)]
32. Loane MA, Oakley A, Rademaker M, Bradford N, Fleischl P, Kerr P, et al. A cost-minimization analysis of the societal costs of realtime teledermatology compared with conventional care: results from a randomized controlled trial in New Zealand. *J Telemed Telecare* 2001;7(4):233-238. [doi: [10.1258/1357633011936453](https://doi.org/10.1258/1357633011936453)] [Medline: [11506759](https://pubmed.ncbi.nlm.nih.gov/11506759/)]
33. Pak HS, Datta SK, Triplett CA, Lindquist JH, Grambow SC, Whited JD. Cost minimization analysis of a store-and-forward teledermatology consult system. *Telemed J E Health* 2009 Mar;15(2):160-165. [doi: [10.1089/tmj.2008.0083](https://doi.org/10.1089/tmj.2008.0083)] [Medline: [19292625](https://pubmed.ncbi.nlm.nih.gov/19292625/)]
34. Snoswell CL, Taylor ML, Caffery LJ. The breakeven point for implementing telehealth. *J Telemed Telecare* 2019 Oct;25(9):530-536. [doi: [10.1177/1357633X19871403](https://doi.org/10.1177/1357633X19871403)] [Medline: [31631758](https://pubmed.ncbi.nlm.nih.gov/31631758/)]
35. Pearl R. Kaiser Permanente Northern California: current experiences with internet, mobile, and video technologies. *Health Aff (Millwood)* 2014 Feb;33(2):251-257. [doi: [10.1377/hlthaff.2013.1005](https://doi.org/10.1377/hlthaff.2013.1005)] [Medline: [24493768](https://pubmed.ncbi.nlm.nih.gov/24493768/)]
36. Chen YH, Lin YH, Hung CS, Huang CC, Yeih DF, Chuang PY, et al. Clinical outcome and cost-effectiveness of a synchronous telehealth service for seniors and nonseniors with cardiovascular diseases: quasi-experimental study. *J Med Internet Res* 2013 Apr 24;15(4):e87 [FREE Full text] [doi: [10.2196/jmir.2091](https://doi.org/10.2196/jmir.2091)] [Medline: [23615318](https://pubmed.ncbi.nlm.nih.gov/23615318/)]
37. Choi Yoo SJ, Nyman JA, Cheville AL, Kroenke K. Cost effectiveness of telecare management for pain and depression in patients with cancer: results from a randomized trial. *Gen Hosp Psychiatry* 2014;36(6):599-606 [FREE Full text] [doi: [10.1016/j.genhosppsy.2014.07.004](https://doi.org/10.1016/j.genhosppsy.2014.07.004)] [Medline: [25130518](https://pubmed.ncbi.nlm.nih.gov/25130518/)]
38. Chung LMY, Law QPS, Fong SSM, Chung JWY, Yuen PP. A cost-effectiveness analysis of teledietetics in short-, intermediate-, and long-term weight reduction. *J Telemed Telecare* 2015 Jul;21(5):268-275. [doi: [10.1177/1357633X15572200](https://doi.org/10.1177/1357633X15572200)] [Medline: [25761467](https://pubmed.ncbi.nlm.nih.gov/25761467/)]

39. Comans T, Mihala G, Sakzewski L, Boyd RN, Scuffham P. The cost-effectiveness of a web-based multimodal therapy for unilateral cerebral palsy: the Mitii randomized controlled trial. *Dev Med Child Neurol* 2017 Jul;59(7):756-761. [doi: [10.1111/dmcn.13414](https://doi.org/10.1111/dmcn.13414)] [Medline: [28247406](https://pubmed.ncbi.nlm.nih.gov/28247406/)]
40. Cubo E, Mariscal N, Solano B, Becerra V, Armesto D, Calvo S, et al. Prospective study on cost-effectiveness of home-based motor assessment in Parkinson's disease. *J Telemed Telecare* 2017 Feb;23(2):328-338. [doi: [10.1177/1357633X16638971](https://doi.org/10.1177/1357633X16638971)] [Medline: [27000142](https://pubmed.ncbi.nlm.nih.gov/27000142/)]
41. Fusco F, Turchetti G. Telerehabilitation after total knee replacement in Italy: cost-effectiveness and cost-utility analysis of a mixed telerehabilitation-standard rehabilitation programme compared with usual care. *BMJ Open* 2016 May 17;6(5):e009964. [doi: [10.1136/bmjopen-2015-009964](https://doi.org/10.1136/bmjopen-2015-009964)] [Medline: [27188803](https://pubmed.ncbi.nlm.nih.gov/27188803/)]
42. Ho YL, Yu JY, Lin YH, Chen YH, Huang CC, Hsu TP, et al. Assessment of the cost-effectiveness and clinical outcomes of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease. *J Med Internet Res* 2014 Jun 10;16(6):e145 [FREE Full text] [doi: [10.2196/jmir.3346](https://doi.org/10.2196/jmir.3346)] [Medline: [24915187](https://pubmed.ncbi.nlm.nih.gov/24915187/)]
43. Richter KP, Shireman TI, Ellerbeck EF, Cupertino AP, Catley D, Cox LS, et al. Comparative and cost effectiveness of telemedicine versus telephone counseling for smoking cessation. *J Med Internet Res* 2015 May 08;17(5):e113 [FREE Full text] [doi: [10.2196/jmir.3975](https://doi.org/10.2196/jmir.3975)] [Medline: [25956257](https://pubmed.ncbi.nlm.nih.gov/25956257/)]
44. Isetta V, Lopez-Agustina C, Lopez-Bernal E, Amat M, Vila M, Valls C, et al. Cost-effectiveness of a new internet-based monitoring tool for neonatal post-discharge home care. *J Med Internet Res* 2013 Feb 18;15(2):e38 [FREE Full text] [doi: [10.2196/jmir.2361](https://doi.org/10.2196/jmir.2361)] [Medline: [23419609](https://pubmed.ncbi.nlm.nih.gov/23419609/)]
45. Boyne JJ, Di Van Asselt A, Gorgels AP, Steuten LM, De Weerd G, Kragten J, et al. Cost-effectiveness analysis of telemonitoring versus usual care in patients with heart failure: The TEHAF-study. *J Telemed Telecare* 2016 Jul 09;19(5):242-248. [doi: [10.1177/1357633x13495478](https://doi.org/10.1177/1357633x13495478)]
46. Dear BF, Zou JB, Ali S, Lorian CN, Johnston L, Sheehan J, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of anxiety: a randomized controlled trial. *Behav Ther* 2015 Mar;46(2):206-217. [doi: [10.1016/j.beth.2014.09.007](https://doi.org/10.1016/j.beth.2014.09.007)] [Medline: [25645169](https://pubmed.ncbi.nlm.nih.gov/25645169/)]
47. Burns CL, Kularatna S, Ward EC, Hill AJ, Byrnes J, Kenny LM. Cost analysis of a speech pathology synchronous telepractice service for patients with head and neck cancer. *Head Neck* 2017 Dec;39(12):2470-2480. [doi: [10.1002/hed.24916](https://doi.org/10.1002/hed.24916)] [Medline: [28963804](https://pubmed.ncbi.nlm.nih.gov/28963804/)]
48. Fortney JC, Pyne JM, Burgess JF. Population-Level Cost-Effectiveness of Implementing Evidence-Based Practices into Routine Care. *Health Serv Res* 2014 Oct 18;49(6):1832-1851. [doi: [10.1111/1475-6773.12247](https://doi.org/10.1111/1475-6773.12247)]
49. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Effect of comprehensive cardiac telerehabilitation on one-year cardiovascular rehospitalization rate, medical costs and quality of life: a cost-effectiveness analysis. *Eur J Prev Cardiol* 2016 May;23(7):674-682. [doi: [10.1177/2047487315602257](https://doi.org/10.1177/2047487315602257)] [Medline: [26289723](https://pubmed.ncbi.nlm.nih.gov/26289723/)]
50. Greving JP, Kaasjager HA, Vernooij JW, Hovens MM, Wierdsma J, Grandjean HM, et al. Cost-effectiveness of a nurse-led internet-based vascular risk factor management programme: economic evaluation alongside a randomised controlled clinical trial. *BMJ Open* 2015 May 20;5(5):e007128. [doi: [10.1136/bmjopen-2014-007128](https://doi.org/10.1136/bmjopen-2014-007128)] [Medline: [25995238](https://pubmed.ncbi.nlm.nih.gov/25995238/)]
51. Henderson C, Knapp M, Fernández JL, Beecham J, Hirani SP, Beynon M, et al. Cost-effectiveness of telecare for people with social care needs: the Whole Systems Demonstrator cluster randomised trial. *Age Ageing* 2014 Nov;43(6):794-800 [FREE Full text] [doi: [10.1093/ageing/afu067](https://doi.org/10.1093/ageing/afu067)] [Medline: [24950690](https://pubmed.ncbi.nlm.nih.gov/24950690/)]
52. Henderson C, Knapp M, Fernández JL, Beecham J, Hirani SP, Cartwright M, Whole System Demonstrator evaluation team. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. *BMJ* 2013 Mar 20;346:f1035. [doi: [10.1136/bmj.f1035](https://doi.org/10.1136/bmj.f1035)] [Medline: [23520339](https://pubmed.ncbi.nlm.nih.gov/23520339/)]
53. Hofer F, Achelrod D, Stargardt T. Cost-utility analysis of telemonitoring interventions for patients with chronic obstructive pulmonary disease (COPD) in Germany. *Appl Health Econ Health Policy* 2016 Dec;14(6):691-701. [doi: [10.1007/s40258-016-0267-9](https://doi.org/10.1007/s40258-016-0267-9)] [Medline: [27475635](https://pubmed.ncbi.nlm.nih.gov/27475635/)]
54. Jódar-Sánchez F, Ortega F, Parra C, Gómez-Suárez C, Bonachela P, Leal S, et al. Cost-utility analysis of a telehealth programme for patients with severe chronic obstructive pulmonary disease treated with long-term oxygen therapy. *J Telemed Telecare* 2014 Sep;20(6):307-316 [FREE Full text] [doi: [10.1177/1357633X14544421](https://doi.org/10.1177/1357633X14544421)] [Medline: [25052387](https://pubmed.ncbi.nlm.nih.gov/25052387/)]
55. Kaambwa B, Bryan S, Jowett S, Mant J, Bray EP, Hobbs FDR, et al. Telemonitoring and self-management in the control of hypertension (TASMINH2): a cost-effectiveness analysis. *Eur J Prev Cardiol* 2014 Dec;21(12):1517-1530. [doi: [10.1177/2047487313501886](https://doi.org/10.1177/2047487313501886)] [Medline: [23990660](https://pubmed.ncbi.nlm.nih.gov/23990660/)]
56. Kidholm K, Rasmussen MK, Andreassen JJ, Hansen J, Nielsen G, Spindler H, et al. Cost-utility analysis of a cardiac telerehabilitation program: the teledialog project. *Telemed J E Health* 2016 Jul;22(7):553-563 [FREE Full text] [doi: [10.1089/tmj.2015.0194](https://doi.org/10.1089/tmj.2015.0194)] [Medline: [26713491](https://pubmed.ncbi.nlm.nih.gov/26713491/)]
57. Mistry H, Gardiner HM. The cost-effectiveness of prenatal detection for congenital heart disease using telemedicine screening. *J Telemed Telecare* 2013 Jun;19(4):190-196. [doi: [10.1258/jtt.2012.120418](https://doi.org/10.1258/jtt.2012.120418)] [Medline: [23576807](https://pubmed.ncbi.nlm.nih.gov/23576807/)]
58. Naveršnik K, Mrhar A. Cost-effectiveness of a novel e-health depression service. *Telemed J E Health* 2013 Feb;19(2):110-116. [doi: [10.1089/tmj.2012.0081](https://doi.org/10.1089/tmj.2012.0081)] [Medline: [23374036](https://pubmed.ncbi.nlm.nih.gov/23374036/)]

59. Nelson RE, Okon N, Lesko AC, Majersik JJ, Bhatt A, Baraban E. The cost-effectiveness of telestroke in the Pacific Northwest region of the USA. *J Telemed Telecare* 2016 Oct;22(7):413-421. [doi: [10.1177/1357633X15613920](https://doi.org/10.1177/1357633X15613920)] [Medline: [26541170](https://pubmed.ncbi.nlm.nih.gov/26541170/)]
60. Nguyen HV, Tan GSW, Tapp RJ, Mital S, Ting DSW, Wong HT, et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. *Ophthalmology* 2016 Dec;123(12):2571-2580. [doi: [10.1016/j.ophtha.2016.08.021](https://doi.org/10.1016/j.ophtha.2016.08.021)] [Medline: [27726962](https://pubmed.ncbi.nlm.nih.gov/27726962/)]
61. Nguyen KH, Smith AC, Armfield NR, Bensink M, Scuffham PA. Cost-effectiveness analysis of a mobile ear screening and surveillance service versus an outreach screening, surveillance and surgical service for indigenous children in Australia. *PLoS One* 2015;10(9):e0138369 [FREE Full text] [doi: [10.1371/journal.pone.0138369](https://doi.org/10.1371/journal.pone.0138369)] [Medline: [26406592](https://pubmed.ncbi.nlm.nih.gov/26406592/)]
62. Rachapelle S, Legood R, Alavi Y, Lindfield R, Sharma T, Kuper H, et al. The cost-utility of telemedicine to screen for diabetic retinopathy in India. *Ophthalmology* 2013 Mar;120(3):566-573. [doi: [10.1016/j.ophtha.2012.09.002](https://doi.org/10.1016/j.ophtha.2012.09.002)] [Medline: [23211635](https://pubmed.ncbi.nlm.nih.gov/23211635/)]
63. Stoddart A, van der Pol M, Pinnock H, Hanley J, McCloughan L, Todd A, et al. Telemonitoring for chronic obstructive pulmonary disease: a cost and cost-utility analysis of a randomised controlled trial. *J Telemed Telecare* 2015 Mar;21(2):108-118. [doi: [10.1177/1357633X14566574](https://doi.org/10.1177/1357633X14566574)] [Medline: [25586810](https://pubmed.ncbi.nlm.nih.gov/25586810/)]
64. Titov N, Dear BF, Ali S, Zou JB, Lorian CN, Johnston L, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: a randomized controlled trial. *Behav Ther* 2015 Mar;46(2):193-205. [doi: [10.1016/j.beth.2014.09.008](https://doi.org/10.1016/j.beth.2014.09.008)] [Medline: [25645168](https://pubmed.ncbi.nlm.nih.gov/25645168/)]
65. Zanaboni P, Landolina M, Marzegalli M, Lunati M, Perego GB, Guenzati G, et al. Cost-utility analysis of the EVOLVO study on remote monitoring for heart failure patients with implantable defibrillators: randomized controlled trial. *J Med Internet Res* 2013 May 30;15(5):e106 [FREE Full text] [doi: [10.2196/jmir.2587](https://doi.org/10.2196/jmir.2587)] [Medline: [23722666](https://pubmed.ncbi.nlm.nih.gov/23722666/)]
66. Thomas S, Hodge W, Malvankar-Mehta M. The cost-effectiveness analysis of teleglaucoma screening device. *PLoS One* 2015;10(9):e0137913 [FREE Full text] [doi: [10.1371/journal.pone.0137913](https://doi.org/10.1371/journal.pone.0137913)] [Medline: [26382956](https://pubmed.ncbi.nlm.nih.gov/26382956/)]
67. Dixon P, Hollinghurst S, Edwards L, Thomas C, Gaunt D, Foster A, et al. Cost-effectiveness of telehealth for patients with raised cardiovascular disease risk: evidence from the Healthlines randomised controlled trial. *BMJ Open* 2016 Aug 26;6(8):e012352. [doi: [10.1136/bmjopen-2016-012352](https://doi.org/10.1136/bmjopen-2016-012352)] [Medline: [27566642](https://pubmed.ncbi.nlm.nih.gov/27566642/)]
68. Isetta V, Negrín MA, Monasterio C, Masa JF, Feu N, Álvarez A, SPANISH SLEEP NETWORK. A Bayesian cost-effectiveness analysis of a telemedicine-based strategy for the management of sleep apnoea: a multicentre randomised controlled trial. *Thorax* 2015 Nov;70(11):1054-1061. [doi: [10.1136/thoraxjnl-2015-207032](https://doi.org/10.1136/thoraxjnl-2015-207032)] [Medline: [26310452](https://pubmed.ncbi.nlm.nih.gov/26310452/)]
69. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. *Med Care* 2000 Jun;38(6):583-637. [doi: [10.1097/00005650-200006000-00004](https://doi.org/10.1097/00005650-200006000-00004)] [Medline: [10843310](https://pubmed.ncbi.nlm.nih.gov/10843310/)]
70. Dimmick SL, Mustaleski C, Burgiss SG, Welsh T. Home Health Nurse 2000 Feb;18(2):124-135. [doi: [10.1097/00004045-200002000-00013](https://doi.org/10.1097/00004045-200002000-00013)] [Medline: [11040644](https://pubmed.ncbi.nlm.nih.gov/11040644/)]
71. Schaafsma J, Pantazi SV, Moehr JR, Anglin CR, Grimm NA. An economic evaluation of a telehealth network in British Columbia. *J Telemed Telecare* 2007 Jul 01;13(5):251-256. [doi: [10.1258/135763307781458877](https://doi.org/10.1258/135763307781458877)]
72. Benger JR, Noble SM, Coast J, Kendall JM. The safety and effectiveness of minor injuries telemedicine. *Emerg Med J* 2004 Jul;21(4):438-445. [Medline: [15208226](https://pubmed.ncbi.nlm.nih.gov/15208226/)]
73. Oakley AM, Kerr P, Duffill M, Rademaker M, Fleischl P, Bradford N, et al. Patient cost-benefits of realtime teledermatology--a comparison of data from Northern Ireland and New Zealand. *J Telemed Telecare* 2000;6(2):97-101. [doi: [10.1258/1357633001935112](https://doi.org/10.1258/1357633001935112)] [Medline: [10824377](https://pubmed.ncbi.nlm.nih.gov/10824377/)]
74. Viers BR, Lightner DJ, Rivera ME, Tollefson MK, Boorjian SA, Karnes RJ, et al. Efficiency, satisfaction, and costs for remote video visits following radical prostatectomy: a randomized controlled trial. *Eur Urol* 2015 Oct;68(4):729-735. [doi: [10.1016/j.eururo.2015.04.002](https://doi.org/10.1016/j.eururo.2015.04.002)] [Medline: [25900782](https://pubmed.ncbi.nlm.nih.gov/25900782/)]
75. Agha Z, Roter DL, Schapira RM. An evaluation of patient-physician communication style during telemedicine consultations. *J Med Internet Res* 2009 Sep 30;11(3):e36 [FREE Full text] [doi: [10.2196/jmir.1193](https://doi.org/10.2196/jmir.1193)] [Medline: [19793720](https://pubmed.ncbi.nlm.nih.gov/19793720/)]
76. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Quality of care for remote orthopaedic consultations using telemedicine: a randomised controlled trial. *BMC Health Serv Res* 2016 Sep 8;16(1):483. [doi: [10.1186/s12913-016-1717-7](https://doi.org/10.1186/s12913-016-1717-7)]
77. Nordal EJ, Moseng D, Kvammen B, Løchen M. A comparative study of teleconsultations versus face-to-face consultations. *J Telemed Telecare* 2001 Oct 01;7(5):257-265. [doi: [10.1258/1357633011936507](https://doi.org/10.1258/1357633011936507)]
78. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. *J Med Internet Res* 2018 Apr 17;20(4):e150 [FREE Full text] [doi: [10.2196/jmir.9897](https://doi.org/10.2196/jmir.9897)] [Medline: [29625956](https://pubmed.ncbi.nlm.nih.gov/29625956/)]
79. Wade V. How to Make Telehealth Work: Defining Telehealth Processes and Procedures.: Adelaide Unicare; 2013. URL: [https://www.peternac.org/sites/default/files/media-uploads/How to Make telehealth Work %20Victoria%20Wade%29.pdf](https://www.peternac.org/sites/default/files/media-uploads/How%20to%20Make%20telehealth%20Work%20Victoria%20Wade%20.pdf) [accessed 2019-11-26]
80. Blount Y, Gloet M. Adoption of ICT Enabled Telehealth Services in the Australian Context: Implications of Technology Use for Telehealth Workers. 2015 Presented at: 48th Hawaii International Conference on System Sciences; Jan. 5 2015 to Jan. 8 2015; Hawaii, USA. [doi: [10.1109/hicss.2015.430](https://doi.org/10.1109/hicss.2015.430)]



81. Eminović N, Dijkgraaf MG, Berghout RM, Prins AH, Bindels PJ, de Keizer NF. A cost minimisation analysis in teledermatology: model-based approach. *BMC Health Serv Res* 2010 Aug 25;10:251 [FREE Full text] [doi: [10.1186/1472-6963-10-251](https://doi.org/10.1186/1472-6963-10-251)] [Medline: [20738871](https://pubmed.ncbi.nlm.nih.gov/20738871/)]
82. Liddy C, Moroz I, Afkham A, Keely E. Evaluating the implementation of The Champlain BASE™ eConsult Service in a new region of Ontario, Canada: a cross-sectional study. *Healthc Policy* 2017 Nov 30;13(2):79-95. [doi: [10.12927/hcpol.2017.25320](https://doi.org/10.12927/hcpol.2017.25320)]
83. Steventon A, Ariti C, Fisher E, Bardsley M. Effect of telehealth on hospital utilisation and mortality in routine clinical practice: a matched control cohort study in an early adopter site. *BMJ Open* 2016 Feb 03;6(2):e009221. [doi: [10.1136/bmjopen-2015-009221](https://doi.org/10.1136/bmjopen-2015-009221)] [Medline: [26842270](https://pubmed.ncbi.nlm.nih.gov/26842270/)]
84. Snoswell CL, Caffery LJ, Whitty JA, Soyer HP, Gordon LG. Cost-effectiveness of skin cancer referral and consultation using teledermoscopy in Australia. *JAMA Dermatol* 2018 Jun 01;154(6):694-700 [FREE Full text] [doi: [10.1001/jamadermatol.2018.0855](https://doi.org/10.1001/jamadermatol.2018.0855)] [Medline: [29801161](https://pubmed.ncbi.nlm.nih.gov/29801161/)]
85. Loane M, Bloomer S, Corbett R, Eedy D, Hicks N, Lotery H, et al. A comparison of real-time and store-and-forward teledermatology: a cost-benefit study. *Br J Dermatol* 2000 Dec;143(6):1241-1247. [doi: [10.1046/j.1365-2133.2000.03895.x](https://doi.org/10.1046/j.1365-2133.2000.03895.x)] [Medline: [11122028](https://pubmed.ncbi.nlm.nih.gov/11122028/)]
86. Liederman EM, Lee JC, Baquero VH, Seites PG. The impact of patient-physician Web messaging on provider productivity. *J Healthc Inf Manag* 2005;19(2):81-86. [Medline: [15869217](https://pubmed.ncbi.nlm.nih.gov/15869217/)]
87. Cheung CMM, Muttardi K, Chinthapalli S, Ismail F. Pilot teledermatology service for assessing solitary skin lesions in a tertiary London dermatology center. *J Healthc Qual* 2019;41(1):e1-e6. [doi: [10.1097/JHQ.0000000000000142](https://doi.org/10.1097/JHQ.0000000000000142)] [Medline: [29634594](https://pubmed.ncbi.nlm.nih.gov/29634594/)]
88. Hager B, Hasselberg M, Arzubi E, Betlinski J, Duncan M, Richman J, et al. Leveraging behavioral health expertise: practices and potential of the Project ECHO approach to virtually integrating care in underserved areas. *Psychiatr Serv* 2018 Apr 01;69(4):366-369. [doi: [10.1176/appi.ps.201700211](https://doi.org/10.1176/appi.ps.201700211)] [Medline: [29446334](https://pubmed.ncbi.nlm.nih.gov/29446334/)]
89. World Health Organisation. Building the economic case for primary health care: a scoping review. 2018 May 08. URL: <https://www.who.int/publications/i/item/WHO-HIS-SDS-2018.48> [accessed 2019-11-26]
90. Comín-Colet J, Enjuanes C, Verdú-Rotellar JM, Linas A, Ruiz-Rodríguez P, González-Robledo G, et al. Impact on clinical events and healthcare costs of adding telemedicine to multidisciplinary disease management programmes for heart failure: Results of a randomized controlled trial. *J Telemed Telecare* 2016 Jul;22(5):282-295. [doi: [10.1177/1357633X15600583](https://doi.org/10.1177/1357633X15600583)] [Medline: [26350543](https://pubmed.ncbi.nlm.nih.gov/26350543/)]
91. Boulé S, Ninni S, Finat L, Botcherby EJ, Kouakam C, Klug D, et al. Potential role of antitachycardia pacing alerts for the reduction of emergency presentations following shocks in patients with implantable cardioverter-defibrillators: implications for the implementation of remote monitoring. *Europace* 2016 Dec;18(12):1809-1817. [doi: [10.1093/europace/euv392](https://doi.org/10.1093/europace/euv392)] [Medline: [26711741](https://pubmed.ncbi.nlm.nih.gov/26711741/)]
92. Celler B, Varnfield M, Sparks R, Li J, Nepal S, Jang-Jaccard J. Home Monitoring of Chronic Disease for Aged Care.: Australian e-Health Research Centre, CSIRO; 2016 May. URL: [https://www.csiro.au/~media/BF/Files/Telehealth-Trial-Final-Report-May-2016\\_3-Final.pdf](https://www.csiro.au/~media/BF/Files/Telehealth-Trial-Final-Report-May-2016_3-Final.pdf) [accessed 2019-11-26]
93. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *BMJ* 2012 Jun 21;344:e3874 [FREE Full text] [doi: [10.1136/bmj.e3874](https://doi.org/10.1136/bmj.e3874)] [Medline: [22723612](https://pubmed.ncbi.nlm.nih.gov/22723612/)]
94. Woods LW, Snow SW. The impact of telehealth monitoring on acute care hospitalization rates and emergency department visit rates for patients using home health skilled nursing care. *Home Healthc Nurse* 2013 Jan;31(1):39-45. [doi: [10.1097/NHH.0b013e3182778dd3](https://doi.org/10.1097/NHH.0b013e3182778dd3)] [Medline: [23238623](https://pubmed.ncbi.nlm.nih.gov/23238623/)]
95. Takahashi PY, Pecina JL, Upatising B, Chaudhry R, Shah ND, Van Houten H, et al. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. *Arch Intern Med* 2012 May 28;172(10):773-779 [FREE Full text] [doi: [10.1001/archinternmed.2012.256](https://doi.org/10.1001/archinternmed.2012.256)] [Medline: [22507696](https://pubmed.ncbi.nlm.nih.gov/22507696/)]
96. Jiménez-Marrero S, Yun S, Cainzos-Achirica M, Enjuanes C, Garay A, Farre N, et al. Impact of telemedicine on the clinical outcomes and healthcare costs of patients with chronic heart failure and mid-range or preserved ejection fraction managed in a multidisciplinary chronic heart failure programme: A sub-analysis of the iCOR randomized trial. *J Telemed Telecare* 2020;26(1-2):64-72. [doi: [10.1177/1357633X18796439](https://doi.org/10.1177/1357633X18796439)] [Medline: [30193564](https://pubmed.ncbi.nlm.nih.gov/30193564/)]
97. Lew SQ, Sikka N, Thompson C, Magnus M. Impact of remote biometric monitoring on cost and hospitalization outcomes in peritoneal dialysis. *J Telemed Telecare* 2019 Dec;25(10):581-586. [doi: [10.1177/1357633X18784417](https://doi.org/10.1177/1357633X18784417)] [Medline: [30001668](https://pubmed.ncbi.nlm.nih.gov/30001668/)]
98. Ehlert A, Oberschachtsiek D. Can telehealth reduce health care expenditure? A lesson from German health insurance data. *Int J Health Plann Manage* 2019 Oct;34(4):1121-1132. [doi: [10.1002/hpm.2764](https://doi.org/10.1002/hpm.2764)] [Medline: [30924194](https://pubmed.ncbi.nlm.nih.gov/30924194/)]
99. Langabeer JR, Gonzalez M, Alqusairi D, Champagne-Langabeer T, Jackson A, Mikhail J, et al. Telehealth-enabled emergency medical services program reduces ambulance transport to urban emergency departments. *West J Emerg Med* 2016 Nov;17(6):713-720 [FREE Full text] [doi: [10.5811/westjem.2016.8.30660](https://doi.org/10.5811/westjem.2016.8.30660)] [Medline: [27833678](https://pubmed.ncbi.nlm.nih.gov/27833678/)]

100. Gillespie SM, Shah MN, Wasserman EB, Wood NE, Wang HY, Noyes K, et al. Reducing emergency department utilization through engagement in telemedicine by senior living communities. *Telemed J E Health* 2016 Jun;22(6):489-496. [doi: [10.1089/tmj.2015.0152](https://doi.org/10.1089/tmj.2015.0152)] [Medline: [26741194](https://pubmed.ncbi.nlm.nih.gov/26741194/)]
101. Caffery LJ, Farjian M, Smith AC. Telehealth interventions for reducing waiting lists and waiting times for specialist outpatient services: a scoping review. *J Telemed Telecare* 2016 Dec;22(8):504-512. [doi: [10.1177/1357633X16670495](https://doi.org/10.1177/1357633X16670495)] [Medline: [27686648](https://pubmed.ncbi.nlm.nih.gov/27686648/)]
102. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-to-consumer telehealth may increase access to care but does not decrease spending. *Health Aff (Millwood)* 2017 Mar 01;36(3):485-491. [doi: [10.1377/hlthaff.2016.1130](https://doi.org/10.1377/hlthaff.2016.1130)] [Medline: [28264950](https://pubmed.ncbi.nlm.nih.gov/28264950/)]
103. Qoctor. Dr Sicknote rebrands as qoctor 2018. 2017. URL: [https://www.qoctor.com.au/drsicknote\\_rebrands\\_as\\_qoctor/](https://www.qoctor.com.au/drsicknote_rebrands_as_qoctor/) [accessed 2019-11-26]
104. iDoc. iDoc24 Research. 2016. URL: <https://www.idoc24.com/about/background-research/> [accessed 2019-11-26]
105. Zhou C, Crawford A, Serhal E, Kurdyak P, Sockalingam S. The impact of Project ECHO on participant and patient outcomes: a systematic review. *Acad Med* 2016 Oct;91(10):1439-1461. [doi: [10.1097/ACM.0000000000001328](https://doi.org/10.1097/ACM.0000000000001328)] [Medline: [27489018](https://pubmed.ncbi.nlm.nih.gov/27489018/)]
106. Fisher E, Hasselberg M, Conwell Y, Weiss L, Padrón NA, Tiernan E, et al. Telementoring primary care clinicians to improve geriatric mental health care. *Popul Health Manag* 2017 Oct;20(5):342-347. [doi: [10.1089/pop.2016.0087](https://doi.org/10.1089/pop.2016.0087)] [Medline: [28106511](https://pubmed.ncbi.nlm.nih.gov/28106511/)]
107. Rattay T, Dumont IP, Heinzow HS, Hutton DW. Cost-effectiveness of access expansion to treatment of hepatitis C virus infection through primary care providers. *Gastroenterology* 2017 Dec;153(6):1531-1543.e2. [doi: [10.1053/j.gastro.2017.10.016](https://doi.org/10.1053/j.gastro.2017.10.016)] [Medline: [29074450](https://pubmed.ncbi.nlm.nih.gov/29074450/)]
108. Wade V, Soar J, Gray L. Uptake of telehealth services funded by Medicare in Australia. *Aust Health Rev* 2014 Nov;38(5):528-532. [doi: [10.1071/AH14090](https://doi.org/10.1071/AH14090)] [Medline: [25219655](https://pubmed.ncbi.nlm.nih.gov/25219655/)]
109. Ford JA, Pereira A. Does teler dermatology reduces secondary care referrals and is it acceptable to patients and doctors?: a service evaluation. *J Eval Clin Pract* 2015 Aug;21(4):710-716. [doi: [10.1111/jep.12373](https://doi.org/10.1111/jep.12373)] [Medline: [25903046](https://pubmed.ncbi.nlm.nih.gov/25903046/)]
110. Jones MD, Etherage JR, Harmon SC, Okiishi JC. Acceptability and cost-effectiveness of military telehealth mental health screening. *Psychol Serv* 2012 May;9(2):132-143. [doi: [10.1037/a0026709](https://doi.org/10.1037/a0026709)] [Medline: [22662728](https://pubmed.ncbi.nlm.nih.gov/22662728/)]
111. Caffery LJ, Taylor M, Lucas K, Smith AC. Substitution rates of video consultations for traditional consultations at a tertiary public hospital. *J Telemed Telecare* 2016 Dec;22(8):453-458. [doi: [10.1177/1357633X16672767](https://doi.org/10.1177/1357633X16672767)] [Medline: [27799447](https://pubmed.ncbi.nlm.nih.gov/27799447/)]
112. Medical Services Advisory Committee. 1360.1 - Specialist dermatology services delivered by asynchronous store and forward technology (resubmission). Canberra: Australian Government, Department of Health; 2017 Aug 03. URL: <http://www.msac.gov.au/internet/msac/publishing.nsf/Content/1360.1-public> [accessed 2019-11-26]

## Abbreviations

**CEA:** cost-effectiveness analysis

**CHEERS:** Consolidated Health Economic Evaluation Reporting Standards

**CMA:** cost-minimization analysis

**CUA:** cost-utility analysis

**ED:** emergency department

**GDP:** gross domestic product

**HRQoL:** health-related quality of life

**PRISMA-ScR:** Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

**Project ECHO:** Project Extension for Community Healthcare Outcomes

**QALYs:** quality-adjusted life years

**RCTs:** randomized controlled trials

*Edited by G Eysenbach; submitted 04.12.19; peer-reviewed by B Smith, K Okamoto; comments to author 07.07.20; revised version received 02.08.20; accepted 17.08.20; published 19.10.20.*

*Please cite as:*

Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ

Determining if Telehealth Can Reduce Health System Costs: Scoping Review

*J Med Internet Res* 2020;22(10):e17298

URL: <http://www.jmir.org/2020/10/e17298/>

doi: [10.2196/17298](https://doi.org/10.2196/17298)

PMID: [33074157](https://pubmed.ncbi.nlm.nih.gov/33074157/)



©Centaine L Snoswell, Monica L Taylor, Tracy A Comans, Anthony C Smith, Leonard C Gray, Liam J Caffery. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 19.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Review

# Accuracy of Mobile Device–Compatible 3D Scanners for Facial Digitization: Systematic Review and Meta-Analysis

Hang-Nga Mai<sup>1</sup>, PhD; Du-Hyeong Lee<sup>1,2</sup>, PhD

<sup>1</sup>Institute for Translational Research in Dentistry, Kyungpook National University, Daegu, Republic of Korea

<sup>2</sup>Department of Prosthodontics, School of Dentistry, Kyungpook National University, Daegu, Republic of Korea

**Corresponding Author:**

Du-Hyeong Lee, PhD

Department of Prosthodontics, School of Dentistry

Kyungpook National University

2175 Dalgubeoldae-ro, Jung-Gu

Daegu, 41940

Republic of Korea

Phone: 82 53 600 7676

Fax: 82 53 427 0778

Email: [dewylee@knu.ac.kr](mailto:dewylee@knu.ac.kr)

## Abstract

**Background:** The accurate assessment and acquisition of facial anatomical information significantly contributes to enhancing the reliability of treatments in dental and medical fields, and has applications in fields such as craniomaxillofacial surgery, orthodontics, prosthodontics, orthopedics, and forensic medicine. Mobile device–compatible 3D facial scanners have been reported to be an effective tool for clinical use, but the accuracy of digital facial impressions obtained with the scanners has not been explored.

**Objective:** We aimed to review comparisons of the accuracy of mobile device–compatible face scanners for facial digitization with that of systems for professional 3D facial scanning.

**Methods:** Individual search strategies were employed in PubMed (MEDLINE), Scopus, Science Direct, and Cochrane Library databases to search for articles published up to May 27, 2020. Peer-reviewed journal articles evaluating the accuracy of 3D facial models generated by mobile device–compatible face scanners were included. Cohen *d* effect size estimates and confidence intervals of standardized mean difference (SMD) data sets were used for meta-analysis.

**Results:** By automatic database searching, 3942 articles were identified, of which 11 articles were considered eligible for narrative review, with 6 studies included in the meta-analysis. Overall, the accuracy of face models obtained using mobile device–compatible face scanners was significantly lower than that of face models obtained using professional 3D facial scanners (SMD 3.96 mm, 95% CI 2.81–5.10 mm;  $z=6.78$ ;  $P<.001$ ). The difference between face scanning when performed on inanimate facial models was significantly higher (SMD 10.53 mm, 95% CI 6.29–14.77 mm) than that when performed on living participants (SMD 2.58 mm, 95% CI 1.70–3.47 mm,  $P<.001$ ,  $df=12.94$ ).

**Conclusions:** Overall, mobile device–compatible face scanners did not perform as well as professional scanning systems in 3D facial acquisition, but the deviations were within the clinically acceptable range of  $<1.5$  mm. Significant differences between results when 3D facial scans were performed on inanimate facial objects and when performed on the faces of living participants were found; thus, caution should be exercised when interpreting results from studies conducted on inanimate objects.

(*J Med Internet Res* 2020;22(10):e22228) doi:[10.2196/22228](https://doi.org/10.2196/22228)

**KEYWORDS**

accuracy; facial digitization; facial scanners; systematic review; meta-analysis

## Introduction

Oral and facial rehabilitation involves comprehensive diagnosis and treatment planning [1,2]. Facial morphology assessment is

vital for the diagnosis of maxillofacial anomalies, surgery, fabrication of prostheses, and postoperative evaluation [2,3]. Esthetics and prognosis of treatment outcomes can be improved through simulation performed on the 3D facial models of

patients [4]. The conventional method for generating facial models of patients is physical facial impression, in which a replica of the face is fabricated using elastomeric materials and a gypsum cast [5,6]. However, the method is uncomfortable for patients because their face is covered with materials during the impression-taking process [6]. In addition, the dimensional accuracy of the physical facial impression model is affected by several factors, including the viscosity of the impression materials, setting time, storage conditions, and time interval from material mixing to stone pouring of the casts [7,8]. Furthermore, the human face is made up of complex anatomical structures with complicated skin textures and colors, which makes realistic replication of the face challenging.

Modern digital technologies have revolutionized the facial impression method by enabling 3D facial morphology to be captured using noncontact optical facial scanning devices [9,10]. Digital impression does not require conventional laboratory work or the use of impression materials, thus reducing the discomfort and chair time of the patients. Compared with facial stone casts, wherein only direct anthropometric measurements of the faces can be performed for facial analyses, virtually reconstructed models of the face can be utilized for multidisciplinary purposes [11-13]. Facial landmarks can easily be extracted from a digital facial model, and the digitized data format enables image merging and advanced dimensional analyses, such as surface-to-surface distance measurements and volume misfit evaluations, using analytical computer software [3,14-17]. In addition, digital facial scanning provides an efficient basis for dental education and facial recognition [18-20].

Stationary facial scanning systems based on stereophotogrammetry technology were first introduced in dentistry [21]. However, because of the encumbrance and high cost of this technology, handheld scanning systems using laser or structured-light technology were developed [21-23]. Although most professional handheld scanners are considered acceptable in terms of their scan image quality, they are expensive and often require considerable training time to learn their complex scanning protocols [3,24,25]. Alternatively, 3D sensor cameras based on structured-light technology have been developed for smartphone and tablet devices [15,26-28]. An advantage of using mobile devices for face scanning is their user-friendly operation; this reduces the training time for users [15,29]. Apps can be developed and customized for specific purposes by using open source scripts and software coding [15,29]. Moreover, when an external attachment-type 3D sensor camera is used, the position of the camera is controllable in the mobile-device system [27,29].

Facial scanning using a mobile device 3D sensor camera has been attracting a lot of interest in recent years because it is highly portable and cost-effective and because of the popularity of mobile devices [29]. Smartphone- and tablet-compatible 3D facial scanners have been reported to be an effective tool for clinical use in prosthodontic treatment [27,30-33]. However, the accuracy of the digital facial impression obtained with mobile device-compatible face scanners has not been explored. The purpose of this systematic review and meta-analysis was

to investigate the accuracy of mobile device-compatible face scanners for facial digitization.

## Methods

### Study Design

This study was designed based on PRISMA guidelines (Preferred Reporting Items For Systematic Reviews and Meta-Analyses) [34]. This review was not preregistered on PROSPERO. Accuracy was defined as a dimensional discrepancy between the digital facial impression made by a mobile device-compatible face scanning camera and reference image data set. The PICO (population, intervention, comparison, and outcomes) question was as follows: Are digital facial impressions (population) obtained with mobile device-compatible 3D facial scanning cameras (intervention) equivalent to those of professional handheld face scanners (comparison) in terms of accuracy (outcomes)?

### Search Strategy

Peer-reviewed studies published until May 27, 2020 were searched using the following formulated Boolean operator: (*digital facial impression OR 3D virtual face OR digital face*) AND (*optical scanner OR 3D scanner OR stereophotogrammetry OR structured light OR laser scanner OR depth sensor cameras OR depth-sensing cameras*) AND (*smart device OR mobile OR smartphone OR tablet OR notebook OR laptop*) AND (*validation OR comparison OR accuracy OR agreement OR reliability OR precision OR reproducibility*). The Boolean operator was applied in major electronic databases including PubMed (MEDLINE), Scopus, Science Direct, and Cochrane Library. The Google Scholar search engine was used to find additional articles by combining the related MeSH (Medical Subject Headings) terms and text words. No automatic limiter setting was used during the searches to prevent unwanted filtering of related articles. EndNote software (version 9.2, Clarivate Analytics Inc) was used to manage the articles' references.

### Inclusion and Exclusion Criteria

Inclusion and exclusion criteria were set based on the study design, objectives, interventions, and measurement results. The search was limited to articles published in English only. The inclusion criteria for meta-analysis were low risk of bias, low concern for applicability, and relevant numeric data for pool-weighted estimation using the Cohen *d* statistical method. Accordingly, randomized and nonrandomized controlled trials, cohort studies, case-control studies, and cross-sectional studies that were performed with human participants and on inanimate objects, reporting quantitative assessments of digital facial models obtained with 3D facial scanners and mobile device-compatible 3D facial scan cameras were included in this review. Conversely, conference papers, case reports, case letters, epidemiologic studies, and author or editorial opinion articles were excluded. Original studies that used only 2D images or did not include mobile device-compatible 3D facial scanners were not reviewed, and studies in which the accuracy could not be quantitatively determined were not considered for analysis.

## Data Collection

Two reviewers (H-NM and D-HL) independently participated in collecting, screening, and selecting the potential studies based on the information provided by the titles and abstracts. The full texts of relevant articles were assessed and reviewed by both reviewers. The papers that satisfied all the inclusion criteria were considered eligible for review. The following information was collected from full-text papers and recorded on an electronic spreadsheet (Office Excel, Microsoft Inc): authors, year of publication, study purpose, participant information (sample size, mean age, age range, and gender proportion), scanning methods (scanning device, capture technology, working condition, and scanning process), reference standard for validation (direct anthropometry or another 3D scanning device), types of measurement performed (linear distances or surface-to-surface deviation), number of measurements (number of landmarks, measurement times, and raters), measurement results (mean, estimation errors, and types of statistical analysis), and major conclusions. Articles with missing data or unreliable data were excluded from the meta-analysis. The agreement ( $\kappa$ ) between the 2 reviewers was calculated. In case of disagreement, a discussion between the 2 reviewers was conducted to resolve the issues.

## Quality Assessment and Meta-Analysis

The risk of bias and concern for applicability based on 4 bias domains—patient selection, index test, reference standard, and flow and timing—were assessed by the 2 reviewers using the Quality Assessment Tool for Diagnostic Accuracy Studies-2 (QUADAS-2) [35].

The random- or fixed-effects model was used to analyze the standardized mean difference (SMD) between the experimental and reference data sets to investigate the effect size estimate

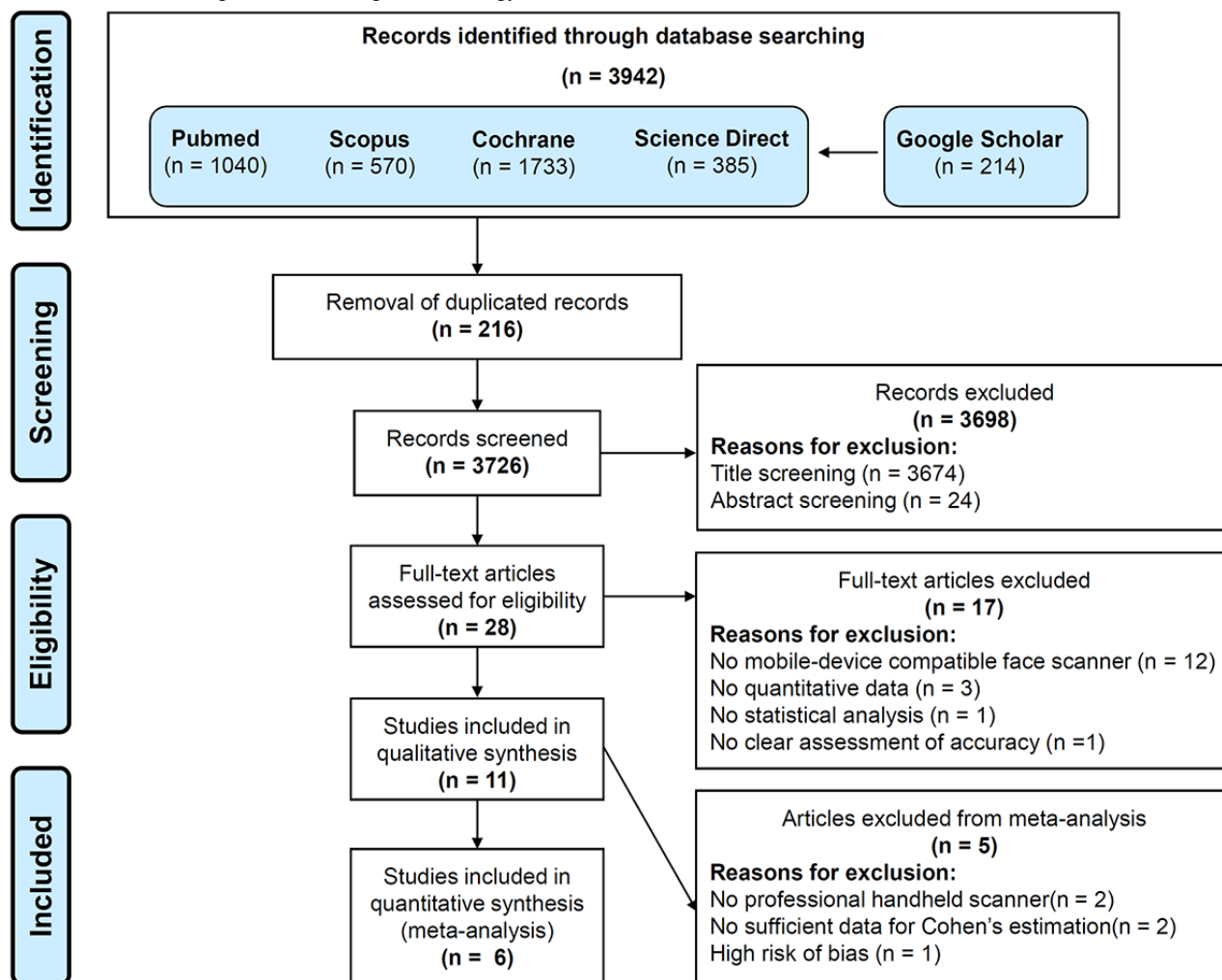
and the confidence intervals of SMDs using Cohen  $d$  [36]. Heterogeneity was evaluated using the Cochran  $Q$  test based on the Higgins  $I^2$  statistic [37], where a higher  $I^2$  value indicated a stronger heterogeneity. When the  $Q$  test indicated high heterogeneity across studies ( $P < .05$ ) or  $I^2 > 50\%$ , the random-effects model was selected, and subgroup analysis was performed [38]. The subgroup was defined based on the participants or inanimate objects investigated.

Publication bias was assessed using the Egger linear regression statistical test and visually inspected using funnel plots. Meta-analyses were performed using the meta package for R software (version 3.6.0, R Foundation for Statistical Computing Platform); the significance level was set at .05. The robvis package (version 0.3.0) was used to visualize the risk-of-bias assessment results [39].

## Results

### Search Results

The search resulted in a total of 3942 articles, which were reduced to 3726 articles after removing 216 duplicates. In the title screening process, 3674 articles that were outside the scope of this review were excluded, thereby leaving 52 articles for abstract screening. After the exclusion of 24 articles with irrelevant abstracts, the full texts of 28 articles were read and assessed, and 11 articles were considered eligible for this review. Of these, 6 articles were included in the global meta-analysis, 4 articles were included in the living person face subgroup analysis, and 3 articles in the inanimate face subgroup analysis. The results of the searching and screening process are summarized in Figure 1. There was substantial interrater agreement ( $\kappa=0.90$ ).

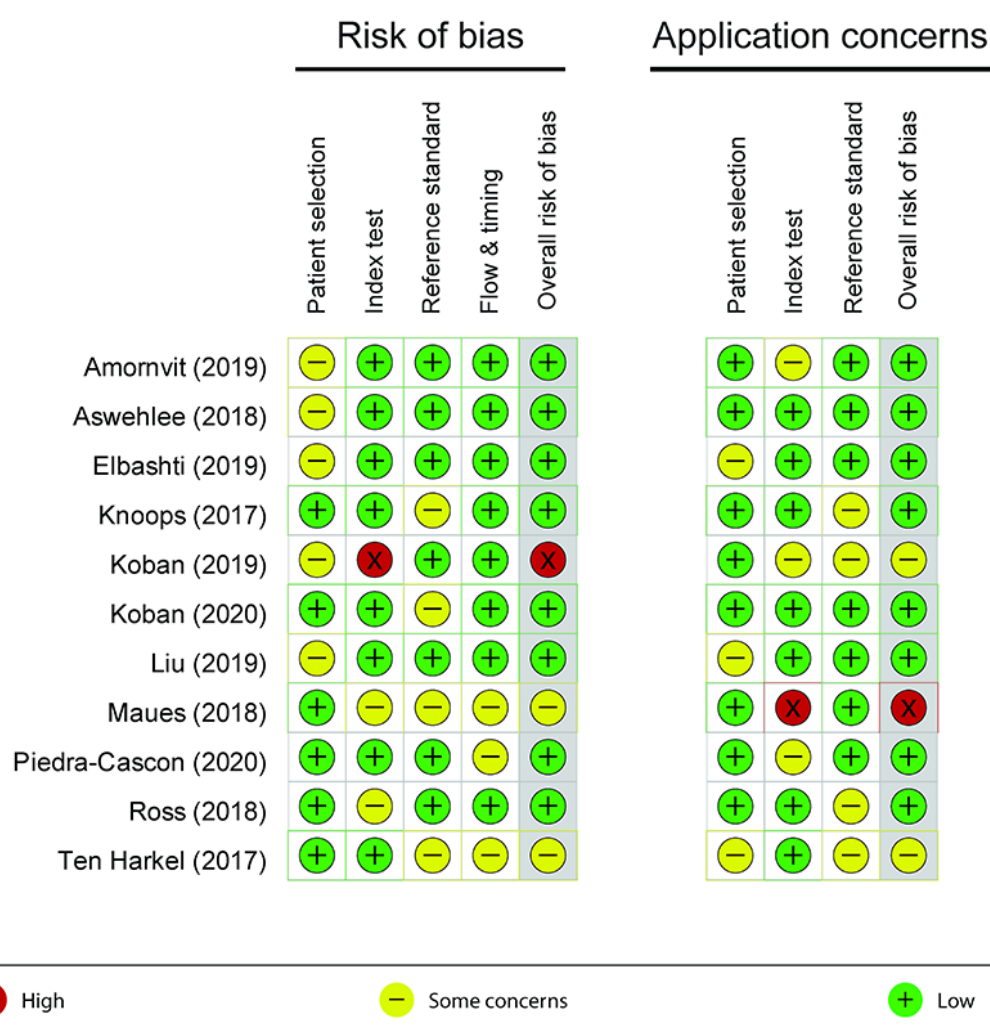
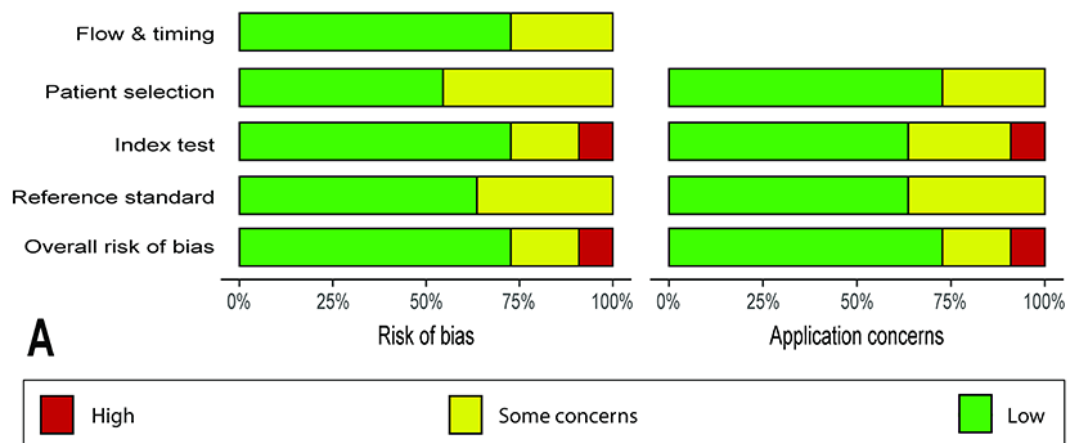
**Figure 1.** PRISMA flow diagram summarizing search strategy and search results.

### Quality Assessment and Applicability Concerns

The quality assessment results from the Quality Assessment Tool for Diagnostic Accuracy Studies-2 showed that among the 11 studies included, one study [40] had a high risk of bias, and another study [41] had a high concern for applicability (Figure 2). There were 2 studies [41,42] showing some risk of bias, and there were 2 studies [40,42] for which there were some concerns

for applicability. The patient selection and index test had a higher risk of bias than those of other domains in some studies because of unclear statements regarding the methods employed for random sampling [28,43] or the small number of participants included [5,15]. For applicability, the major concerns arose in the index test domain because several studies did not describe the scanning procedures in detail or did not provide sufficient information about the scanning devices [27,28,40,41].



**Figure 2.** Quality assessment results according to Quality Assessment Tool for Diagnostic Accuracy Studies-2 guidelines.

### Study Characteristics

Extracted data were organized according to the characteristics of the studies (Table 1). The characteristics of the mobile

device-compatible face scanners that were investigated are summarized in Multimedia Appendix 1.

**Table 1.** Characteristics of the included studies.

Study	Participant or specimen	Mobile device, face scanner	Reference	Landmark, n	Measurement	Major findings
Amornvit (2019)[28]	1 mannequin head	iPhone X (Apple Inc), FaceApp (Bellus3D Inc)	Manual measurement	N/A <sup>a</sup>	$\Delta x$ , $\Delta y$ , and $\Delta z$	$\Delta x$ : range 10-50 mm; $\Delta y$ : range 50-120 mm; failure to record the details in the z-axis
Aswehlee (2018)[5]	1 impression cast	Scanify (Fuel 3D Technologies Ltd)	CT <sup>b</sup>	3D point clouds	RMSE <sup>c</sup>	The most accurate noncontact 3D digitizer for maxillofacial defects was Vivid 910 (Minolta Corp), followed by Danae (NEC Engineering), 3dMD (3dMD LLC), and Scanify ( $P<.001$ ).
Elbashti (2019)[15]	1 impression cast	iPhone 6 (Apple Inc), 123D Catch App (Autodesk Inc)	CT	3D point clouds	RMSE	Smartphone 3D modeling was not as accurate as that of the commercially available laser scanning, with higher RMSE values in the defect area representing the depth of the defect.
Knoops (2017)[29]	8 (4 male, 4 female)	Structure Sensor (Occipital Inc)	SP <sup>d</sup>	3D point clouds; 4	RMSE	RMSE of the Structure Sensor was significantly higher than that of M4D Scan (Rodin4D) ( $P=.008$ ). Structure Sensor lacks hardware and software to accurately characterize areas with complex shape and high curvature but is good at describing general facial forms.
Koban (2019)[40]	4 cadaver heads (N/A)	Sense (3D Systems Inc); iSense (3D Systems Inc)	N/A	3D point clouds	RMSE	Artec Eva (Artec Group) provided significantly more accurate results than those of the Sense ( $P<.001$ ) and the iSense devices ( $P<.001$ ). The Sense was more accurate than the iSense scanner; however, the difference was not significant ( $P=.12$ ).
Koban (2020)[44]	30 (15 male, 15 female), 1 mannequin head	Sense (3D Systems Inc)	SP	3D point clouds	RMSE	Whole face $<1.0$ mm (RMSE 0.516, SD 0.109 mm).
Liu (2019)[43]	2 impression cast (male)	Scanify (Fuel 3D Technologies Ltd)	CT	13	11 linear deviations ( $\Delta x$ , $\Delta y$ , and $\Delta z$ )	Overall, linear deviations $<1$ mm for Scanify. The mean overall difference $<0.3$ mm between Scanify (mean 0.74, SD 0.089 mm) and Vectra (mean 0.15, SD 0.015 mm) images.
Maues (2018)[41]	10 (5 male, 5 female)	Kinect (Microsoft Inc)	SP	10	7 linear distances (mean difference)	Mean difference between scanning methods was 0.3 (SD 2.03 mm), showing reasonable accuracy. The mean difference between the images taken with Kinect was 0.1 (SD 0.6 mm; $P<.05$ ) showing good accuracy. Kinect appears to be an interesting and promising resource for facial analysis.
Piedra-Cascón (2020)[27]	10 (2 male, 8 female)	Face Camera Pro (Bellus3D Inc)	Manual measurement	6	RMSE	Face Camera Pro exhibited a trueness RMSE of 0.91 mm and a precision RMSE of 0.32 mm.
Ross (2018)[45]	16 (8 male, 8 female)	iPhone 7 (Apple Inc), Camera+ app (tap tap tap LLC); RealSense (Intel Corp)	Structured light	3D point clouds	RMSE	No significant differences in RMSE values between iPhone scans with 90 photographs (RMSE 1.4, SD 0.6 mm), 60 photographs (RMSE 1.2, SD 0.2 mm), or 30 photographs (RMSE 1.2, SD 0.3 mm). RealSense had significantly higher RMSE than the iPhone experimental groups ( $P<.001$ ).
Ten Harkel (2017)[42]	34 (10 male, 24 female)	RealSense (Intel Corp)	SP	3D point clouds	RMSE	RealSense depth accuracy was not affected by facial palsy (RMSE 1.48, SD 0.28 mm) compared to a healthy face (RMSE 1.46, SD 0.26 mm) or Sunnybrook poses <sup>e</sup> ( $P=.76$ ). However, distance of the patients to the RealSense device was shown to affect accuracy, where the highest depth accuracy (1.07 mm) was measured at a distance of 35 cm.

<sup>a</sup>N/A: not applicable.<sup>b</sup>CT: computed tomography.<sup>c</sup>RMSE: root-mean-square error (surface-to-surface).<sup>d</sup>SP: stereophotogrammetry.<sup>e</sup>Sunnybrook poses are a facial grading system for evaluating facial movement outcomes, both at rest and through 5 facial expressions based on voluntary movements (forehead wrinkle, gentle eye closure, open mouth smile, snarl, and lip pucker) [42].

Among the 11 studies included, 6 were conducted on adult volunteers or patients [27,29,41,42,44,45] with a mean age of 35.50 years (SD 8.50; range 24-59). The number of participants in these studies ranged from 8 to 34, with 2 to 15 male and 4 to 15 female participants. The other 5 studies were conducted using inanimate objects such as impression casts of the face [5,15,43] or mannequin heads [28,44], and 1 study [40] was conducted on human cadaver heads. Stereophotogrammetry [29,41,42,44], computed tomography [5,15,43], and high-resolution structured-light handheld scanning [40,45] were used as the reference measurements for comparison, and 2 studies [27,28] used manual interlandmark distance as the reference measurement.

For the evaluation, most studies [5,15,27,29,40-42,44,45] measured the global surface-to-surface deviation between the reference and test images by calculating the root-mean-square error (RMSE) of the superimposed 3D images using analytical computer software, with a higher RMSE value indicating a higher surface deviation; however, 3 studies [28,41,43] compared the distances between facial landmarks on a digitized face with those between respective landmarks on a physical model obtained using the manual measurement method. Among them, 1 study [41] evaluated both the global surface-to-surface

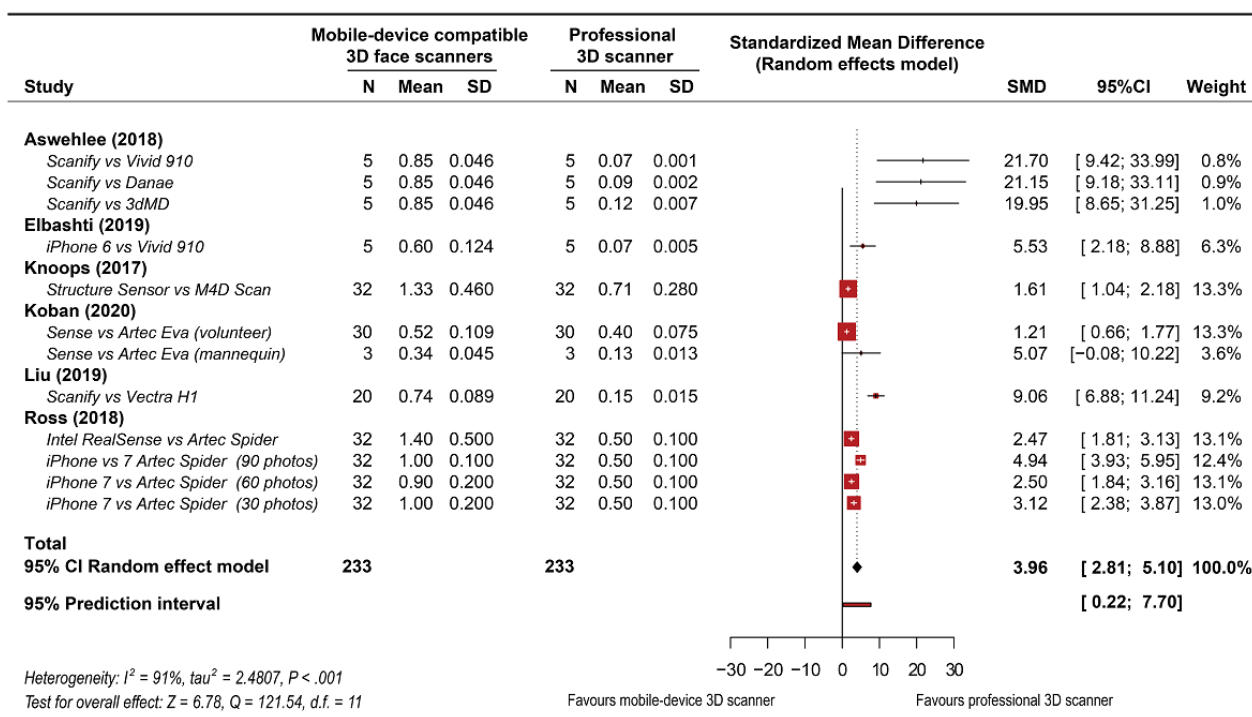
deviation and interlandmark linear distances, and the deviation was assessed along the *x*-axis (horizontal length), *y*-axis (vertical length), and *z*-axis (depth) in another study [28].

## Meta-Analysis

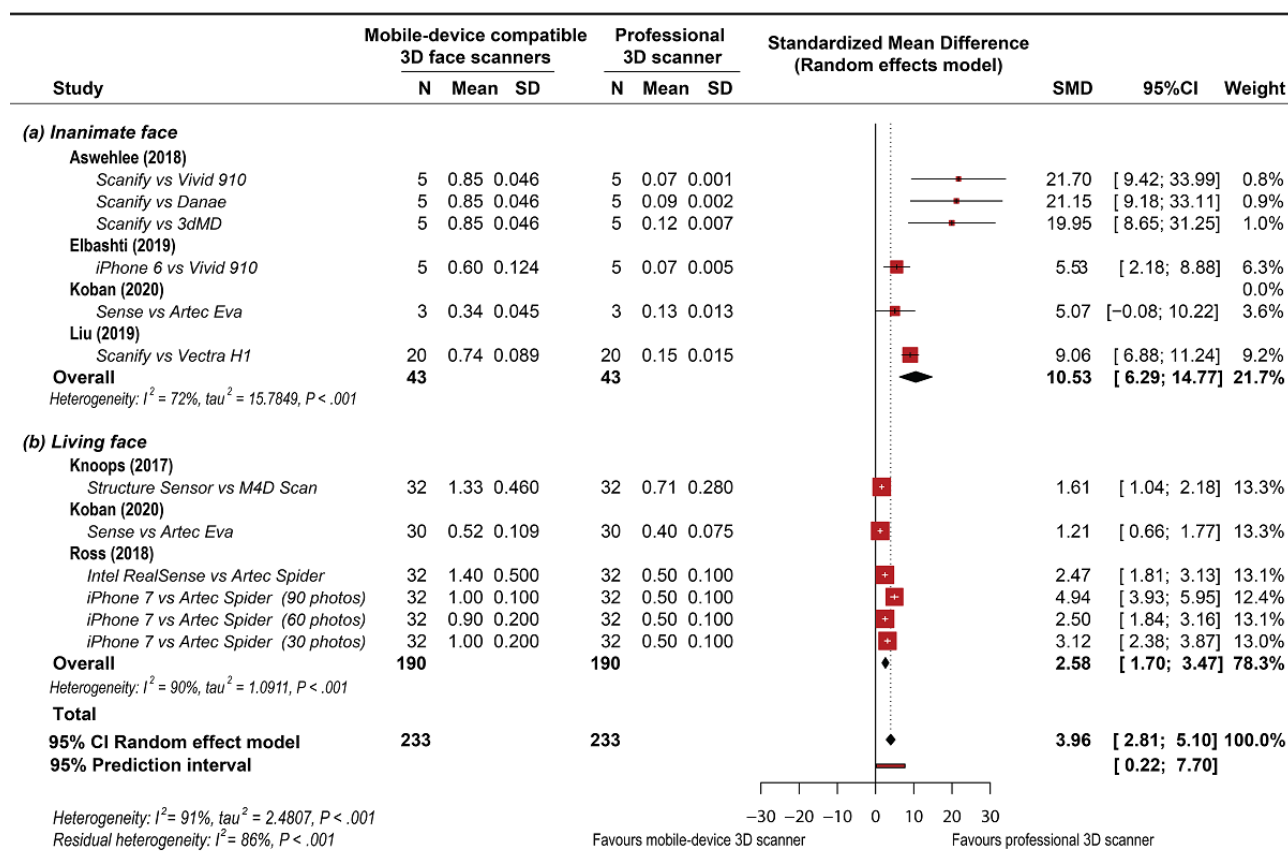
The global analysis revealed heterogeneity ( $I^2=91\%$   $P<.001$ ). Random-effects models were selected for both global and subgroup meta-analyses based on the heterogeneity among the studies. In general, the accuracy of facial models obtained with mobile device-compatible face scanners was significantly lower than that of facial models obtained using professional face scanners (SMD 3.96 mm, 95% CI 2.81-5.10 mm,  $z=6.78$ ,  $P<.001$ ; Figure 3). Results from the subgroup analysis revealed a significant difference between the subgroups (Figure 4). The difference between the mobile device-compatible and professional face scanners was significantly higher for the face scans of inanimate facial objects (SMD 10.53 mm, 95% CI 6.29-14.77 mm) than for those of living participants (SMD 2.58 mm, 95% CI 1.70-3.47 mm,  $P<.001$ ,  $df=12.94$ ).

The funnel plot showed asymmetry arising from 3 distinct points with different effect estimates (Figure 5). Regarding publication bias, Egger test results showed an intercept of 3.9 (95% CI 1.94-5.86,  $t=3.792$ ,  $P=.004$ ).

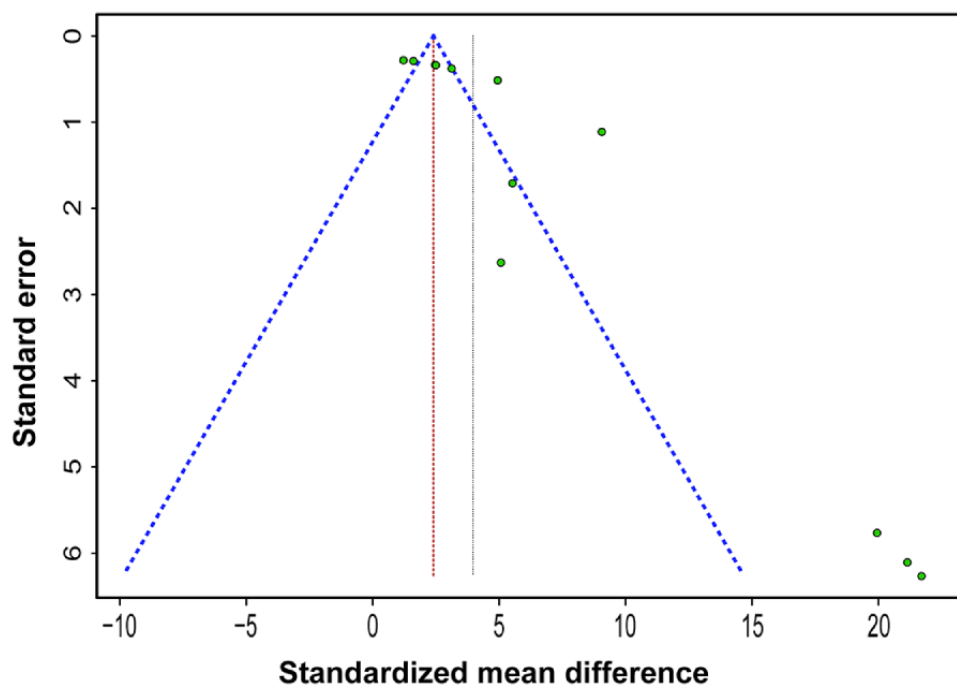
**Figure 3.** Global meta-analysis results of comparison of facial models obtained using mobile device-compatible face scanners versus professional face scanners.



**Figure 4.** Subgroup meta-analysis results of comparison of facial models obtained using mobile device-compatible face scanners versus professional face scanners. (a) 3D facial scans performed on inanimate objects, (b) 3D facial scans performed on living persons.



**Figure 5.** Funnel plot showing of publication bias assessment.



## Discussion

### Principal Findings

We aimed to investigate the accuracy of mobile device-compatible face scanners in facial digitization. Mean discrepancy values of the digitized face obtained using mobile device-compatible 3D facial scanners ranged from 0.34 to 1.40 mm in articles included in this systematic review. The meta-analysis revealed that mobile device-compatible 3D facial scanners were less accurate than their professional 3D counterparts. The reliability of a digital face scanner can be classified into 4 categories: highly reliable (deviation <1.0 mm), reliable (deviation 1.0 mm-1.5 mm), moderately reliable (deviation 1.5 mm-2.0 mm), and unreliable (deviation >2.0 mm) [46]. For clinical application, deviations <1.5 mm were considered acceptable [3,47,48]. Based on the classifications, mobile device-compatible 3D facial scanners were considered acceptable for clinical use even though their accuracies were lower than those of the professional 3D facial scanners. Amornvit et al [28] and Liu et al [43] reported that mobile device-compatible face scanners are comparable to professional 3D facial scanners when scanning simple and flat areas of the face such as the forehead, cheeks, and chin. However, scanning accuracy was relatively low when mobile device-compatible face scanners were used to capture complex facial regions, such as the external ears, eyelids, nostril, and teeth [28,44,45]. Higher inaccuracy was found in the facial areas with defects, depending on the depth of the defect [15]. Thus, careful consideration in accordance with the purpose and the person might be needed when using mobile device-compatible face scanners.

In the preliminary stages, smartphone-based 3D scanners used a multiphotogrammetry approach that captured several photographs of the object from different views and matched common features in the photographs to establish a 3D model of the object by using dedicated smartphone software apps [15,45]. The resolution of a 3D image depended on the number of reconstructed polygons that were calculated by the software algorithm based on the resolution of the captured images [49]. The working principle is similar to that of professional stereophotogrammetry facial scanning systems; however, professional systems usually use digital single-lens reflex cameras that have higher pixel densities with better noise reduction software and higher ISO settings compared with those of smartphone cameras [50]. The accuracy of smartphone multiphotogrammetry in facial data acquisition was reported as 0.605 (SD 0.124) mm by Elbashti et al [15]. In another study by Ross et al [45], the mean discrepancy of scan data obtained using smartphones ranged from 0.9 mm to 1.0 mm, depending on the number of photographs taken during scanning. In recent years, infrared structured-light depth-sensing cameras have been incorporated in mobile devices to facilitate 3D optical scans [51]. 3D depth-sensing cameras work by the time-of-flight principle, measuring the time taken for light emitted by an illumination unit to travel to an object and back to the sensor array [52,53]. The 3D images are then reconstructed based on a depth map of the object and surroundings [54]. Although smartphone depth-sensing cameras share similar working principles with professional laser scanning systems, laser

systems are more sensitive to depths because they are built with higher sensitivity sensors [15,23]. Amornvit et al [28] reported that the 3D depth-sensing sensor scanner of a smartphone is reliable in linear measurement at the frontal plane, but it has less accuracy in depth measurement compared with that of professional face scanners. Depth-sensing cameras can also be used separately and attached or plugged into smartphones, tablets, or laptop computers to acquire 3D scans [27,29,40-42,44]. Because the quality of facial scanning is also affected by the performance of compatible mobile devices when external depth-sensing cameras are used, the resulting accuracy might vary widely and should be evaluated for each combination of depth-sensing camera and mobile device.

Subgroup meta-analysis showed that the accuracy of 3D facial scans performed on living persons was significantly different compared with those performed on inanimate objects. This result implies that the outcomes of in vitro or laboratory studies could be different from those obtained from people. Thus, based on the findings of this review, we recommend using living persons for related research on mobile device-compatible face scanning. Caution should be exercised when scanning the orbital, nasolabial, and oral regions on the face of a living person to minimize the discrepancies caused by motion artifacts [16,24]. Subconscious nose breathing, eye blinking, and lip twitching should also be carefully considered as these are the main sources of involuntary facial movements [16]. Ozsoy et al [17] reported that changes in facial expressions could affect the reproducibility and reliability of a scan, with the highest error values observed for a frightened facial expression and the lowest value observed for neutral facial expression. To reduce motion artifacts, the person should be instructed to maintain a neutral facial expression and avoid any head movement during image acquisition [55]. Another concern is that human faces contain complex skin textures, pores, freckles, scars, and wrinkles. Some artifacts or missing scan data appear as holes, originating from surfaces that are difficult to capture, such as eyebrows, eyelashes, and hairlines [29]. Small empty holes can be repaired using image processing software that uses neighboring areas that are morphologically similar; however, large defects can cause difficulties in the stitching process because of the lack of reference [24]. In addition, human faces vary in shape and are not perfectly symmetric, thus may appear different in different angles of view [56]. This phenomenon might cause some artifacts when the multiphotogrammetry approach is used because the 3D model of an object is reconstructed by matching common facial features in the captured photographs.

A limitation of this review is that the review protocol was not preregistered on PROSPERO. Most included studies are not directly correlated with clinical treatment outcomes due to the difficulty of performing clinical studies to assess the accuracy of scanners. However, the findings of this review show great promise for the clinical use of mobile device-compatible face scanners. Another limitation of this systematic review is the small number of included studies. The limited number of studies show high heterogeneity and funnel plot asymmetry. Regarding publication bias, the Egger test result was significant ( $P=.004$ ). Heterogeneity can cause funnel plot asymmetry when a correlation between intervention effects and study sizes is



present [57]. Further examination was performed on the eligibility of a study that showed distinctly larger effect estimates [5], and we included the study [5] in the meta-analysis because it was conducted in an environment of a scanning intervention and was methodologically scientific. Although the inclusion of this study [5] increased the heterogeneity among studies and funnel plot asymmetry, the results were fundamentally attributed to a small number of articles [58]. All eligible papers included in the review were published between 2017 and 2020 due to the novelty of the research topic. A random-effects model is often used in meta-analyses for studies with heterogeneity. Random effects meta-analyses weigh studies more equally than fixed-effect analyses by incorporating the variance between studies [58]. Therefore, in this review, based on heterogeneity and funnel plot asymmetry, random-effects

models were selected for global and subgroup analyses. Additional controlled in vitro and randomized clinical trials will be needed to reinforce the impact of review articles. Moreover, considering the rapid development of face scanning in the medical field, diverse investigations with newly developed devices and systems need to be continuously performed.

## Conclusions

Overall, the accuracy of mobile device-compatible face scanners in 3D facial acquisition was not comparable to that of professional optical scanning systems, but it was still within the clinically acceptable range of <1.5 mm in dimensional deviation. There were significant differences between 3D facial scans performed on inanimate objects and living persons; thus, caution should be exercised when interpreting the results from studies conducted on inanimate objects.

## Acknowledgments

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation funded by the Korean government (2020R1I1A1A01062967).

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Commercial mobile device-compatible face scanners investigated in the studies.

[DOCX File, 18 KB - [jmir\\_v22i10e22228\\_app1.docx](https://www.jmir.org/2020/10/e22228_app1.docx)]

## References

1. Spear FM, Kokich VG. A multidisciplinary approach to esthetic dentistry. *Dent Clin North Am* 2007 Apr;51(2):487-505, x-xi. [doi: [10.1016/j.cden.2006.12.007](https://doi.org/10.1016/j.cden.2006.12.007)] [Medline: [17532924](https://pubmed.ncbi.nlm.nih.gov/17532924/)]
2. Jazayeri HE, Kang S, Masri RM, Kuhn L, Fahimipour F, Vanevenhoven R, et al. Advancements in craniofacial prosthesis fabrication: A narrative review of holistic treatment. *J Adv Prosthodont* 2018 Dec;10(6):430-439 [FREE Full text] [doi: [10.4047/jap.2018.10.6.430](https://doi.org/10.4047/jap.2018.10.6.430)] [Medline: [30584472](https://pubmed.ncbi.nlm.nih.gov/30584472/)]
3. Zhao YJ, Xiong YX, Wang Y. Three-dimensional accuracy of facial scan for facial deformities in clinics: a new evaluation method for facial scanner accuracy. *PLoS One* 2017;12(1):e0169402 [FREE Full text] [doi: [10.1371/journal.pone.0169402](https://doi.org/10.1371/journal.pone.0169402)] [Medline: [28056044](https://pubmed.ncbi.nlm.nih.gov/28056044/)]
4. Bidra AS. Three-dimensional esthetic analysis in treatment planning for implant-supported fixed prosthesis in the edentulous maxilla: review of the esthetics literature. *J Esthet Restor Dent* 2011 Aug;23(4):219-236. [doi: [10.1111/j.1708-8240.2011.00428.x](https://doi.org/10.1111/j.1708-8240.2011.00428.x)] [Medline: [21806753](https://pubmed.ncbi.nlm.nih.gov/21806753/)]
5. Aswehlee AM, Elbashti ME, Hattori M, Sumita Y, Taniguchi H. Feasibility and accuracy of noncontact three-dimensional digitizers for geometric facial defects: an in vitro comparison. *Int J Prosthodont* 2018;31(6):601-606. [doi: [10.11607/ijp.5855](https://doi.org/10.11607/ijp.5855)] [Medline: [30339162](https://pubmed.ncbi.nlm.nih.gov/30339162/)]
6. Alsiyabi AS, Minsley GE. Facial moulage fabrication using a two-stage poly (vinyl siloxane) impression. *J Prosthodont* 2006;15(3):195-197. [doi: [10.1111/j.1532-849X.2006.00100.x](https://doi.org/10.1111/j.1532-849X.2006.00100.x)] [Medline: [16681502](https://pubmed.ncbi.nlm.nih.gov/16681502/)]
7. Thongthammachat S, Moore B, Barco M, Hovijitra S, Brown D, Andres C. Dimensional accuracy of dental casts: influence of tray material, impression material, and time. *J Prosthodont* 2002 Jun;11(2):98-108. [doi: [10.1053/jopr.2002.125192](https://doi.org/10.1053/jopr.2002.125192)] [Medline: [12087547](https://pubmed.ncbi.nlm.nih.gov/12087547/)]
8. Gonçalves F, Popoff D, Castro C, Silva G, Magalhães C, Moreira A. Dimensional stability of elastomeric impression materials: a critical review of the literature. *Eur J Prosthodont Restor Dent* 2011 Dec;19(4):163-166. [doi: [10.1922/EJPRD\\_998Silva04](https://doi.org/10.1922/EJPRD_998Silva04)]
9. Gibelli D, Dolci C, Cappella A, Sforza C. Reliability of optical devices for three-dimensional facial anatomy description: a systematic review and meta-analysis. *Int J Oral Maxillofac Surg* 2020 Aug;49(8):1092-1106. [doi: [10.1016/j.ijom.2019.10.019](https://doi.org/10.1016/j.ijom.2019.10.019)] [Medline: [31786104](https://pubmed.ncbi.nlm.nih.gov/31786104/)]

10. Bohner L, Gamba DD, Hanisch M, Marcio BS, Tortamano Neto P, Laganá DC, et al. Accuracy of digital technologies for the scanning of facial, skeletal, and intraoral tissues: a systematic review. *J Prosthet Dent* 2019 Feb;121(2):246-251. [doi: [10.1016/j.prosdent.2018.01.015](https://doi.org/10.1016/j.prosdent.2018.01.015)] [Medline: [30017156](https://pubmed.ncbi.nlm.nih.gov/30017156/)]
11. Douglas TS. Image processing for craniofacial landmark identification and measurement: a review of photogrammetry and cephalometry. *Comput Med Imaging Graph* 2004 Oct;28(7):401-409. [doi: [10.1016/j.compmedimag.2004.06.002](https://doi.org/10.1016/j.compmedimag.2004.06.002)] [Medline: [15464879](https://pubmed.ncbi.nlm.nih.gov/15464879/)]
12. Plooi J, Maal T, Haers P, Borstlap W, Kuijpers-Jagtman A, Bergé SJ. Digital three-dimensional image fusion processes for planning and evaluating orthodontics and orthognathic surgery: a systematic review. *Int J Oral Maxillofac Surg* 2011 Apr;40(4):341-352. [doi: [10.1016/j.ijom.2010.10.013](https://doi.org/10.1016/j.ijom.2010.10.013)] [Medline: [21095103](https://pubmed.ncbi.nlm.nih.gov/21095103/)]
13. Kook M, Jung S, Park H, Oh H, Ryu S, Cho J, et al. A comparison study of different facial soft tissue analysis methods. *J Craniomaxillofac Surg* 2014 Jul;42(5):648-656. [doi: [10.1016/j.jcms.2013.09.010](https://doi.org/10.1016/j.jcms.2013.09.010)] [Medline: [24954528](https://pubmed.ncbi.nlm.nih.gov/24954528/)]
14. Lippold C, Liu X, Wangdo K, Drerup B, Schreiber K, Kirschneck C, et al. Facial landmark localization by curvature maps and profile analysis. *Head Face Med* 2014 Dec 08;10:54 [FREE Full text] [doi: [10.1186/1746-160X-10-54](https://doi.org/10.1186/1746-160X-10-54)] [Medline: [25488063](https://pubmed.ncbi.nlm.nih.gov/25488063/)]
15. Elbashti M, Sumita Y, Aswehlee A, Seelaus R. Smartphone application as a low-cost alternative for digitizing facial defects: is it accurate enough for clinical application? *Int J Prosthodont* 2019;32(6):541-543. [doi: [10.11607/ijp.6347](https://doi.org/10.11607/ijp.6347)] [Medline: [31664272](https://pubmed.ncbi.nlm.nih.gov/31664272/)]
16. Camison L, Bykowski M, Lee W, Carlson J, Roosenboom J, Goldstein J, et al. Validation of the Vectra H1 portable three-dimensional photogrammetry system for facial imaging. *Int J Oral Maxillofac Surg* 2018 Mar;47(3):403-410 [FREE Full text] [doi: [10.1016/j.ijom.2017.08.008](https://doi.org/10.1016/j.ijom.2017.08.008)] [Medline: [28919165](https://pubmed.ncbi.nlm.nih.gov/28919165/)]
17. Özsoy U, Sekerci R, Hizay A, Yildirim Y, Uysal H. Assessment of reproducibility and reliability of facial expressions using 3D handheld scanner. *J Craniomaxillofac Surg* 2019 Jun;47(6):895-901. [doi: [10.1016/j.jcms.2019.03.022](https://doi.org/10.1016/j.jcms.2019.03.022)] [Medline: [30975563](https://pubmed.ncbi.nlm.nih.gov/30975563/)]
18. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med* 2010 Oct;85(10):1589-1602. [doi: [10.1097/ACM.0b013e3181edfe13](https://doi.org/10.1097/ACM.0b013e3181edfe13)] [Medline: [20703150](https://pubmed.ncbi.nlm.nih.gov/20703150/)]
19. Ko B. A brief review of facial emotion recognition based on visual information. *Sensors (Basel)* 2018 Jan 30;18(2):401 [FREE Full text] [doi: [10.3390/s18020401](https://doi.org/10.3390/s18020401)] [Medline: [29385749](https://pubmed.ncbi.nlm.nih.gov/29385749/)]
20. Jeon B, Jeong B, Jee S, Huang Y, Kim Y, Park GH, et al. A facial recognition mobile app for patient safety and biometric identification: design, development, and validation. *JMIR Mhealth Uhealth* 2019 Apr 08;7(4):e11472 [FREE Full text] [doi: [10.2196/11472](https://doi.org/10.2196/11472)] [Medline: [30958275](https://pubmed.ncbi.nlm.nih.gov/30958275/)]
21. Burke P, Beard L. Stereophotogrammetry of the face. A preliminary investigation into the accuracy of a simplified system evolved for contour mapping by photography. *Am J Orthod Dentofacial Orthop* 1967 Oct;53(10):769-782. [doi: [10.1016/0002-9416\(67\)90121-2](https://doi.org/10.1016/0002-9416(67)90121-2)] [Medline: [5233929](https://pubmed.ncbi.nlm.nih.gov/5233929/)]
22. Gwilliam JR, Cunningham SJ, Hutton T. Reproducibility of soft tissue landmarks on three-dimensional facial scans. *Eur J Orthod* 2006 Oct;28(5):408-415. [doi: [10.1093/ejo/cjl024](https://doi.org/10.1093/ejo/cjl024)] [Medline: [16901962](https://pubmed.ncbi.nlm.nih.gov/16901962/)]
23. Kovacs L, Zimmermann A, Brockmann G, Baurecht H, Schwenzer-Zimmerer K, Papadopoulos N, et al. Accuracy and precision of the three-dimensional assessment of the facial surface using a 3-D laser scanner. *IEEE Trans Med Imaging* 2006 Jun;25(6):742-754. [doi: [10.1109/tmi.2006.873624](https://doi.org/10.1109/tmi.2006.873624)] [Medline: [16768239](https://pubmed.ncbi.nlm.nih.gov/16768239/)]
24. Bakirman T, Gumusay MU, Reis HC, Selbesoglu MO, Yosmaoglu S, Yaras MC, et al. Comparison of low cost 3D structured light scanners for face modeling. *Appl Opt* 2017 Feb 01;56(4):985-992. [doi: [10.1364/AO.56.000985](https://doi.org/10.1364/AO.56.000985)] [Medline: [28158103](https://pubmed.ncbi.nlm.nih.gov/28158103/)]
25. Ma L, Xu T, Lin J. Validation of a three-dimensional facial scanning system based on structured light techniques. *Comput Methods Programs Biomed* 2009 Jun;94(3):290-298. [doi: [10.1016/j.cmpb.2009.01.010](https://doi.org/10.1016/j.cmpb.2009.01.010)] [Medline: [19303659](https://pubmed.ncbi.nlm.nih.gov/19303659/)]
26. Piccirilli M, Doretto G, Ross A, Adjeroh D. A mobile structured light system for 3D face acquisition. *IEEE Sensors J* 2016 Apr;16(7):1854-1855. [doi: [10.1109/jsen.2015.2511064](https://doi.org/10.1109/jsen.2015.2511064)]
27. Piedra-Cascón W, Meyer MJ, Methani MM, Revilla-León M. Accuracy (trueness and precision) of a dual-structured light facial scanner and interexaminer reliability. *J Prosthet Dent* 2020 Jan 06;S0022-3913(19):30689-30684. [doi: [10.1016/j.prosdent.2019.10.010](https://doi.org/10.1016/j.prosdent.2019.10.010)] [Medline: [31918895](https://pubmed.ncbi.nlm.nih.gov/31918895/)]
28. Amornvit P, Sanohkan S. The accuracy of digital face scans obtained from 3D Scanners: an in vitro study. *Int J Environ Res Public Health* 2019 Dec 12;16(24):5061 [FREE Full text] [doi: [10.3390/ijerph16245061](https://doi.org/10.3390/ijerph16245061)] [Medline: [31842255](https://pubmed.ncbi.nlm.nih.gov/31842255/)]
29. Knoops PG, Beaumont CA, Borghi A, Rodriguez-Florez N, Breakey RW, Rodgers W, et al. Comparison of three-dimensional scanner systems for craniomaxillofacial imaging. *J Plast Reconstr Aesthet Surg* 2017 Apr;70(4):441-449. [doi: [10.1016/j.bjps.2016.12.015](https://doi.org/10.1016/j.bjps.2016.12.015)] [Medline: [28161205](https://pubmed.ncbi.nlm.nih.gov/28161205/)]
30. Hassan B, Greven M, Wismeijer D. Integrating 3D facial scanning in a digital workflow to CAD/CAM design and fabricate complete dentures for immediate total mouth rehabilitation. *J Adv Prosthodont* 2017 Oct;9(5):381-386 [FREE Full text] [doi: [10.4047/jap.2017.9.5.381](https://doi.org/10.4047/jap.2017.9.5.381)] [Medline: [29142646](https://pubmed.ncbi.nlm.nih.gov/29142646/)]
31. Lo Russo L, Salamini A, Troiano G, Guida L. Digital dentures: a protocol based on intraoral scans. *J Prosthet Dent* 2020 Apr 21;S0022-3913(20):30139-30136 (forthcoming). [doi: [10.1016/j.prosdent.2020.02.006](https://doi.org/10.1016/j.prosdent.2020.02.006)] [Medline: [32331785](https://pubmed.ncbi.nlm.nih.gov/32331785/)]

32. Hong S, Noh K. Setting the sagittal condylar inclination on a virtual articulator by using a facial and intraoral scan of the protrusive interocclusal position: a dental technique. *J Prosthet Dent* 2020 Mar 24;S0022-3913(20):30101-30103 (forthcoming). [doi: [10.1016/j.prosdent.2020.01.031](https://doi.org/10.1016/j.prosdent.2020.01.031)] [Medline: [32220398](https://pubmed.ncbi.nlm.nih.gov/32220398/)]
33. Revilla-León M, Raney L, Piedra-Cascón W, Barrington J, Zandinejad A, Özcan M. Digital workflow for an esthetic rehabilitation using a facial and intraoral scanner and an additive manufactured silicone index: a dental technique. *J Prosthet Dent* 2020 Apr;123(4):564-570. [doi: [10.1016/j.prosdent.2019.03.014](https://doi.org/10.1016/j.prosdent.2019.03.014)] [Medline: [31353112](https://pubmed.ncbi.nlm.nih.gov/31353112/)]
34. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
35. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011 Oct 18;155(8):529-536. [doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)] [Medline: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)]
36. Cohen J. Statistical Power Analysis. *Curr Dir Psychol Sci* 1992;1(3):98-101. [doi: [10.1111/1467-8721.ep10768783](https://doi.org/10.1111/1467-8721.ep10768783)]
37. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: [10.1136/bmj.327.7414.557](https://doi.org/10.1136/bmj.327.7414.557)] [Medline: [12958120](https://pubmed.ncbi.nlm.nih.gov/12958120/)]
38. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007 Feb;28(2):105-114. [doi: [10.1016/j.cct.2006.04.004](https://doi.org/10.1016/j.cct.2006.04.004)] [Medline: [16807131](https://pubmed.ncbi.nlm.nih.gov/16807131/)]
39. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. *Res Synth Methods* 2020 Apr 26:1-7 (forthcoming). [doi: [10.1002/jrsm.1411](https://doi.org/10.1002/jrsm.1411)] [Medline: [32336025](https://pubmed.ncbi.nlm.nih.gov/32336025/)]
40. Koban K, Cotofana S, Frank K, Green J, Etzel L, Li Z, et al. Precision in 3-dimensional surface imaging of the face: a handheld scanner comparison performed in a cadaveric model. *Aesthet Surg J* 2019 Mar 14;39(4):NP36-NP44. [doi: [10.1093/asj/sjy242](https://doi.org/10.1093/asj/sjy242)] [Medline: [30239592](https://pubmed.ncbi.nlm.nih.gov/30239592/)]
41. Maués CPR, Casagrande M, Almeida R, Almeida M, Carvalho F. Three-dimensional surface models of the facial soft tissues acquired with a low-cost scanner. *Int J Oral Maxillofac Surg* 2018 Sep;47(9):1219-1225. [doi: [10.1016/j.ijom.2018.03.028](https://doi.org/10.1016/j.ijom.2018.03.028)] [Medline: [29705405](https://pubmed.ncbi.nlm.nih.gov/29705405/)]
42. Ten Harkel TC, Speksnijder CM, van der Heijden F, Beurskens CHG, Ingels KJAO, Maal TJJ. Depth accuracy of the RealSense F200: low-cost 4D facial imaging. *Sci Rep* 2017 Nov 24;7(1):16263 [FREE Full text] [doi: [10.1038/s41598-017-16608-7](https://doi.org/10.1038/s41598-017-16608-7)] [Medline: [29176666](https://pubmed.ncbi.nlm.nih.gov/29176666/)]
43. Liu C, Artopoulos A. Validation of a low-cost portable 3-dimensional face scanner. *Imaging Sci Dent* 2019 Mar;49(1):35-43 [FREE Full text] [doi: [10.5624/isd.2019.49.1.35](https://doi.org/10.5624/isd.2019.49.1.35)] [Medline: [30941286](https://pubmed.ncbi.nlm.nih.gov/30941286/)]
44. Koban KC, Perko P, Etzel L, Li Z, Schenck TL, Giunta RE. Validation of two handheld devices against a non-portable three-dimensional surface scanner and assessment of potential use for intraoperative facial imaging. *J Plast Reconstr Aesthet Surg* 2020 Jan;73(1):141-148. [doi: [10.1016/j.bjps.2019.07.008](https://doi.org/10.1016/j.bjps.2019.07.008)] [Medline: [31519501](https://pubmed.ncbi.nlm.nih.gov/31519501/)]
45. Ross MT, Cruz R, Brooks-Richards TL, Hafner LM, Powell SK, Woodruff MA. Comparison of three-dimensional surface scanning techniques for capturing the external ear. *Virtual and Physical Prototyping* 2018 Jul 05;13(4):255-265. [doi: [10.1080/17452759.2018.1493803](https://doi.org/10.1080/17452759.2018.1493803)]
46. Aung S, Ngim R, Lee S. Evaluation of the laser scanner as a surface measuring tool and its accuracy compared with direct facial anthropometric measurements. *Br J Plast Surg* 1995 Dec;48(8):551-558. [doi: [10.1016/0007-1226\(95\)90043-8](https://doi.org/10.1016/0007-1226(95)90043-8)] [Medline: [8548155](https://pubmed.ncbi.nlm.nih.gov/8548155/)]
47. Secher JJ, Darvann TA, Pinholt EM. Accuracy and reproducibility of the DAVID SLS-2 scanner in three-dimensional facial imaging. *J Craniomaxillofac Surg* 2017 Oct;45(10):1662-1670. [doi: [10.1016/j.jcms.2017.07.006](https://doi.org/10.1016/j.jcms.2017.07.006)] [Medline: [28847623](https://pubmed.ncbi.nlm.nih.gov/28847623/)]
48. Ye H, Lv L, Liu Y, Liu Y, Zhou Y. Evaluation of the accuracy, reliability, and reproducibility of two different 3D face-scanning systems. *Int J Prosthodont* 2016;29(3):213-218. [doi: [10.11607/ijp.4397](https://doi.org/10.11607/ijp.4397)] [Medline: [27148978](https://pubmed.ncbi.nlm.nih.gov/27148978/)]
49. Rangel FA, Maal TJ, Bergé SJ, van Vlijmen OJ, Plooi JM, Schutyser F, et al. Integration of digital dental casts in 3-dimensional facial photographs. *Am J Orthod Dentofacial Orthop* 2008 Dec;134(6):820-826. [doi: [10.1016/j.ajodo.2007.11.026](https://doi.org/10.1016/j.ajodo.2007.11.026)] [Medline: [19061810](https://pubmed.ncbi.nlm.nih.gov/19061810/)]
50. Lane C, Harrell W. Completing the 3-dimensional picture. *Am J Orthod Dentofacial Orthop* 2008 Apr;133(4):612-620. [doi: [10.1016/j.ajodo.2007.03.023](https://doi.org/10.1016/j.ajodo.2007.03.023)] [Medline: [18405826](https://pubmed.ncbi.nlm.nih.gov/18405826/)]
51. Yao H, Ge C, Xue J, Zheng N. A high spatial resolution depth sensing method based on binocular structured light. *Sensors (Basel)* 2017 Apr 08;17(4):805 [FREE Full text] [doi: [10.3390/s17040805](https://doi.org/10.3390/s17040805)] [Medline: [28397759](https://pubmed.ncbi.nlm.nih.gov/28397759/)]
52. Sarbolandi H, Lefloch D, Kolb A. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Computer Vision and Image Understanding* 2015 Oct;139:1-20. [doi: [10.1016/j.cviu.2015.05.006](https://doi.org/10.1016/j.cviu.2015.05.006)]
53. Jia T, Zhou Z, Gao H. Depth measurement based on infrared coded structured light. *Journal of Sensors* 2014;2014:1-8. [doi: [10.1155/2014/852621](https://doi.org/10.1155/2014/852621)]
54. Alfaro-Santafé J, Gómez-Bernal A, Lanuza-Cerzócimo C, Alfaro-Santafé JV, Pérez-Morcillo A, Almenar-Arasanz AJ. Three-axis measurements with a novel system for 3D plantar foot scanning: iPhone X. *Footwear Science* 2020 Mar 05;12(2):123-131. [doi: [10.1080/19424280.2020.1734867](https://doi.org/10.1080/19424280.2020.1734867)]

55. Verhulst A, Hol M, Vreeken R, Becking A, Ulrich D, Maal T. Three-dimensional imaging of the face: a comparison between three different imaging modalities. *Aesthet Surg J* 2018 May 15;38(6):579-585. [doi: [10.1093/asj/sjx227](https://doi.org/10.1093/asj/sjx227)] [Medline: [29360971](https://pubmed.ncbi.nlm.nih.gov/29360971/)]
56. Zaidel DW, Hessamian M. Asymmetry and symmetry in the beauty of human faces. *Symmetry* 2010 Feb 23;2(1):136-149. [doi: [10.3390/sym2010136](https://doi.org/10.3390/sym2010136)]
57. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *J Clin Epidemiol* 2005 Sep;58(9):894-901. [doi: [10.1016/j.jclinepi.2005.01.006](https://doi.org/10.1016/j.jclinepi.2005.01.006)] [Medline: [16085192](https://pubmed.ncbi.nlm.nih.gov/16085192/)]
58. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011 Jul 22;343:d4002. [doi: [10.1136/bmj.d4002](https://doi.org/10.1136/bmj.d4002)] [Medline: [21784880](https://pubmed.ncbi.nlm.nih.gov/21784880/)]

## Abbreviations

**RMSE:** root-mean-square error

**SMD:** standardized mean difference

*Edited by G Eysenbach; submitted 06.07.20; peer-reviewed by M Behzadifar, KM Kuo, J Vermeir; comments to author 15.09.20; revised version received 29.09.20; accepted 30.09.20; published 23.10.20.*

*Please cite as:*

*Mai HN, Lee DH*

*Accuracy of Mobile Device-Compatible 3D Scanners for Facial Digitization: Systematic Review and Meta-Analysis*

*J Med Internet Res* 2020;22(10):e22228

URL: <http://www.jmir.org/2020/10/e22228/>

doi: [10.2196/22228](https://doi.org/10.2196/22228)

PMID: [33095178](https://pubmed.ncbi.nlm.nih.gov/33095178/)

©Hang-Nga Mai, Du-Hyeong Lee. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Review

# Toward a Digital Platform for the Self-Management of Noncommunicable Disease: Systematic Review of Platform-Like Interventions

Sarah A Tighe<sup>1,2</sup>, BSc, MSc; Kylie Ball<sup>1</sup>, PhD; Finn Kensing<sup>3</sup>, Dr Scient; Lars Kayser<sup>2</sup>, MD, PhD; Jonathan C Rawstorn<sup>1</sup>, PhD; Ralph Maddison<sup>1</sup>, PhD

<sup>1</sup>Institute for Physical Activity and Nutrition (IPAN), School of Exercise and Nutrition Sciences, Deakin University, Burwood, Australia

<sup>2</sup>Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

<sup>3</sup>Department of Computer Science, Faculty of Science, University of Copenhagen, Copenhagen, Denmark

**Corresponding Author:**

Sarah A Tighe, BSc, MSc

Institute for Physical Activity and Nutrition (IPAN)

School of Exercise and Nutrition Sciences

Deakin University

221 Burwood Highway

Burwood, VIC 3125

Australia

Phone: 61 92468383 ext 95590

Email: [stighe@deakin.edu.au](mailto:stighe@deakin.edu.au)

## Abstract

**Background:** Digital interventions are effective for health behavior change, as they enable the self-management of chronic, noncommunicable diseases (NCDs). However, they often fail to facilitate the specific or current needs and preferences of the individual. A proposed alternative is a digital platform that hosts a suite of discrete, already existing digital health interventions. A platform architecture would allow users to explore a range of evidence-based solutions over time to optimize their self-management and health behavior change.

**Objective:** This review aims to identify digital platform-like interventions and examine their potential for supporting self-management of NCDs and health behavior change.

**Methods:** A literature search was conducted in January 2020 using EBSCOhost, PubMed, Scopus, and EMBASE. No digital platforms were identified, so criteria were broadened to include digital platform-like interventions. Eligible platform-like interventions offered a suite of discrete, evidence-based health behavior change features to optimize self-management of NCDs in an adult population and provided digitally supported guidance for the user toward the features best suited to their needs and preferences. Data collected on interventions were guided by the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist, including evaluation data on effectiveness and process outcomes. The quality of the included literature was assessed using the Mixed Methods Appraisal Tool.

**Results:** A total of 7 studies were included for review. Targeted NCDs included cardiovascular diseases (CVD; n=3), diabetes (n=3), and chronic obstructive pulmonary disease (n=1). The mean adherence (based on the number of follow-up responders) was 69% (SD 20%). Of the 7 studies, 4 with the highest adherence rates (80%) were also guided by behavior change theories and took an iterative, user-centered approach to development, optimizing intervention relevance. All 7 interventions presented algorithm-supported user guidance tools, including electronic decision support, smart features that interact with patterns of use, and behavior change stage-matching tools. Of the 7 studies, 6 assessed changes in behavior. Significant effects in moderate-to-vigorous physical activity were reported, but for no other specific health behaviors. However, positive behavior change was observed in studies that focused on comprehensive behavior change measures, such as self-care and self-management, each of which addresses several key lifestyle risk factors (eg, medication adherence). No significant difference was found for psychosocial outcomes (eg, quality of life). Significant changes in clinical outcomes were predominately related to disease-specific, multifaceted measures such as clinical disease control and cardiovascular risk score.

**Conclusions:** Iterative, user-centered development of digital platform structures could optimize user engagement with self-management support through existing, evidence-based digital interventions. Offering a palette of interventions with an



appropriate degree of guidance has the potential to facilitate disease-specific health behavior change and effective self-management among a myriad of users, conditions, or stages of care.

(*J Med Internet Res* 2020;22(10):e16774) doi:[10.2196/16774](https://doi.org/10.2196/16774)

## KEYWORDS

noncommunicable diseases; chronic disease; web-based intervention; mobile health; self-management; health behavior; mobile phone

## Introduction

### Background

Noncommunicable diseases (NCDs) are the leading cause of death and disability worldwide [1,2]. Although complex [3-5], approximately 80% of NCDs can be accounted for by modifiable risk behaviors, such as physical inactivity, unhealthy diets, smoking, harmful alcohol consumption, and stress management [3,6,7]. The comprehensive management of risk behaviors [8,9] through the implementation of self-management strategies [10-12] is critical to fostering a life-long approach to secondary prevention [13]. According to Bandura [14], to achieve successful and sustainable self-management, *people have to learn to monitor their health behavior and the circumstances under which it occurs*.

Secondary prevention in an NCD context usually involves referral to a structured hospital or community-based programs, such as cardiac rehabilitation [13,15,16], pulmonary rehabilitation [17], or diabetes education [18-21]. Participation in center-based programs improves health-related quality of life [12,22-24] and clinical outcomes [23] and lowers the risk of hospitalization [12,22,24], recurring adverse health-related events, and all-cause mortality [9], compared with usual care.

Despite the proven effectiveness of secondary prevention [12,22-24], uptake and adherence to structured face-to-face programs is suboptimal [25-29]. Personal factors associated with low attendance include cultural, financial, and psychological barriers (eg, readiness or willingness to attend) [17,27,30]. Key operational barriers to active participation relate to availability of program resources, including limited program enrollment capacity [29,31], restrictive hours of operation [27], and limitations in program suitability (eg, for people self-managing multiple NCDs). Accessibility barriers include inadequate transportation and a lack of services within remote rural areas [17,27,29,30,32-36]. In response, there has been a necessary shift toward flexible and convenient home-based services [32,37-41], offering an evidence-based alternative to nonattendance and reducing the distance of care [42].

Digital health is a contemporary advancement in home-based self-management of NCDs. Widespread internet connectivity to at least 55% of the global population [43], over 5 billion mobile phone users [44] and the availability of thousands of mobile health applications [45] has provided unprecedented access to the digital delivery of home-based support. Content, structures, and modes of delivery for these digital interventions are wide ranging, and their potential effects are well documented.

A systematic review and meta-analysis of internet-based interventions targeting health behavior change in NCD groups (n=43,236) showed significant improvements in risk-related behavior, such as physical activity (effect size 0.24; 95% CI 0.09-0.38), dietary behavior (effect size 0.20; 95% CI 0.02-0.37), and alcohol consumption (effect size 0.14; 95% CI 0.00-0.27) [46]. The findings showed that offering multiple modes of delivery within an intervention (eg, internet-based plus text messaging) had a greater effect on behavior (effect size 0.81; 95% CI 0.14-1.49) [46].

Reviews of communication technologies used to deliver health services and facilitate patient and health care professional interaction [47] have reported that improvements in self-management of NCDs are not inferior to the positive changes produced by structured center-based programs or usual care [42,48]. Similarly, reviews of mobile-based text messaging interventions have shown sustained health behavior change [49-53] and the potential to overcome barriers associated with traditional center-based models via simple to use, flexible, cost-effective digital health [52,53].

Collectively, the literature indicates that evidence-based digital health interventions have the potential to discretely impact the self-management of NCDs while also complementing one another. Despite the advantages of these discrete interventions, no single program meets the needs of all users, as self-management of NCDs and user preferences are complex and multifaceted in nature. Discrete digital health solutions could be complimentary when combined for use. The REMOTE-CR program [54] and CAP-CR program [55] draw on combinations of web-based and mobile-based features through smartphone-enabled software to facilitate digitally assisted self-management. Outcomes from a randomized control trial of the CAP-CR intervention (n=120) indicated that uptake, adherence, and completion rates at 6 months were significantly higher than those of center-based cardiac rehabilitation programs [55]. REMOTE-CR integrates the use of smartphones, wearable sensors, and web apps to provide real-time remote exercise monitoring. A randomized control trial (n=162) found REMOTE-CR to be an effective and cost-efficient alternative to center-based programs [54].

The constraints of disparate digital health interventions mean that an undue amount of time is required to seek out interventions offering the content, features, or delivery mode best suited to the current preferences of the individual user. Moreover, when a person's health status or general lifestyle habits change, seeking practices would have to be repeated, which may reduce the sustainability of a discrete digital program intervention and long-term user adherence [56].

Therefore, we need flexible and versatile solutions in the digital health space. This requires a shift *away* from the creation of restrictive digital interventions *toward* a paradigm that facilitates optimal engagement through centralized choice.

## Objective

We propose a digital platform that would capitalize on existing digital self-management interventions that have been evaluated for effectiveness in a specific context (eg, NCDs) through a comprehensive experimental design [57]. The platform would host a digitally supported palette of discrete, evidence-based digital health interventions and incorporate a digital guidance tool to direct users to the intervention best suited to their current individual needs and preferences, optimizing personal relevance and user experience [47].

Existing disparate digital health interventions may offer modest positive effects, but engagement can be varied. Thus, a digital platform could potentially optimize engagement while also lessening the burden of care associated with irrelevant content, user ambivalence, and time-consuming seeking processes. Such an approach not only facilitates personalization but also encourages user autonomy through the self-selection of a combination of program components, which is associated with long-lasting positive effects on disease management and patient empowerment [37,58,59]. There is significant potential for this body of work, as a plethora of underutilized evidence-based digital health interventions exists, involving rigorous development processes and gold standard evaluations.

## Methods

### Identifying Digital Platforms

A preliminary pilot of this review was conducted in August 2018, but no digital platforms that matched our criteria were

identified. Therefore, the focus of this review was extended to include digital *platform-like* interventions. Details of the protocol for this review were registered on PROSPERO (International Prospective Register of Systematic Reviews) in 2018 (PROSPERO 2018, Registration Number: CRD42018102095). A follow-up literature search was conducted in January 2020.

For this review, a digital platform-like intervention was defined as *a digital solution that allows users to choose from a suite of discrete, evidence-based health behavior change features to support NCD self-management. Offers a digital tool for guidance toward intervention features that are most suited to the user's needs and preferences.* This broadened definition allowed us to best provide an assessment of the potential of digital platforms in self-management of NCDs, despite the absence of existing literature.

### Aims

The aims of this study were to identify digital platform-like interventions for the self-management of NCDs and to examine the potential for digital platform-like interventions to support self-management of NCDs through effectiveness and process outcome measures.

### Eligibility Criteria

The eligibility criteria are outlined in Table 1 according to an adapted version of the Population, Intervention, Context, Outcome approach [60,61]. Studies eligible for inclusion were published in English between January 1990 and January 2020; this period coincides with the activation of commercial internet service providers. All study designs were eligible for inclusion because of the emergent nature of the research topic.

**Table 1.** Framing the review aims according to an adapted Population, Intervention, Context, Outcome approach.

Factor	Description
<b>Population</b>	
Adults	Aged $\geq 18$ years
Noncommunicable chronic disease	Intervention supported participants in self-management of at least one lifestyle-related, noncommunicable, chronic disease (eg, cardiovascular disease, chronic obstructive pulmonary disease, and type 2 diabetes)
<b>Intervention</b>	
Health behavior	Intervention targeted at least one health behavior (eg, physical activity or diet)
Discrete behavior change components	Offered at least two evidence-based BC <sup>a</sup> components or features (eg, self-monitoring, goal setting, or feedback)
Digital delivery	Compatible with any modern computing devices (eg, web based or mobile based)
Guidance	Offered a choice between BC components. Provided users with a digitally assisted guidance tool to assist with feature selection.
<b>Context</b>	
Secondary prevention and self-management	Primarily facilitating outpatient, home-based stages of NCD <sup>b</sup> care. Focus on participants' own self-management of NCD.
<b>Outcomes</b>	
Intervention description	Mode of delivery, participant information, comparators, intervention features and components, theoretical frameworks or tools, development processes.
Effectiveness	Significant improvements in health behavior, clinical outcomes, or evidence-based psychosocial outcomes.
Process	Intervention use (log-in data), adherence (completion of follow-up data collection), and user satisfaction (quantitative or qualitative follow-up).

<sup>a</sup>BC: behavior change.<sup>b</sup>NCD: noncommunicable disease.

## Information Sources and Search Strategy

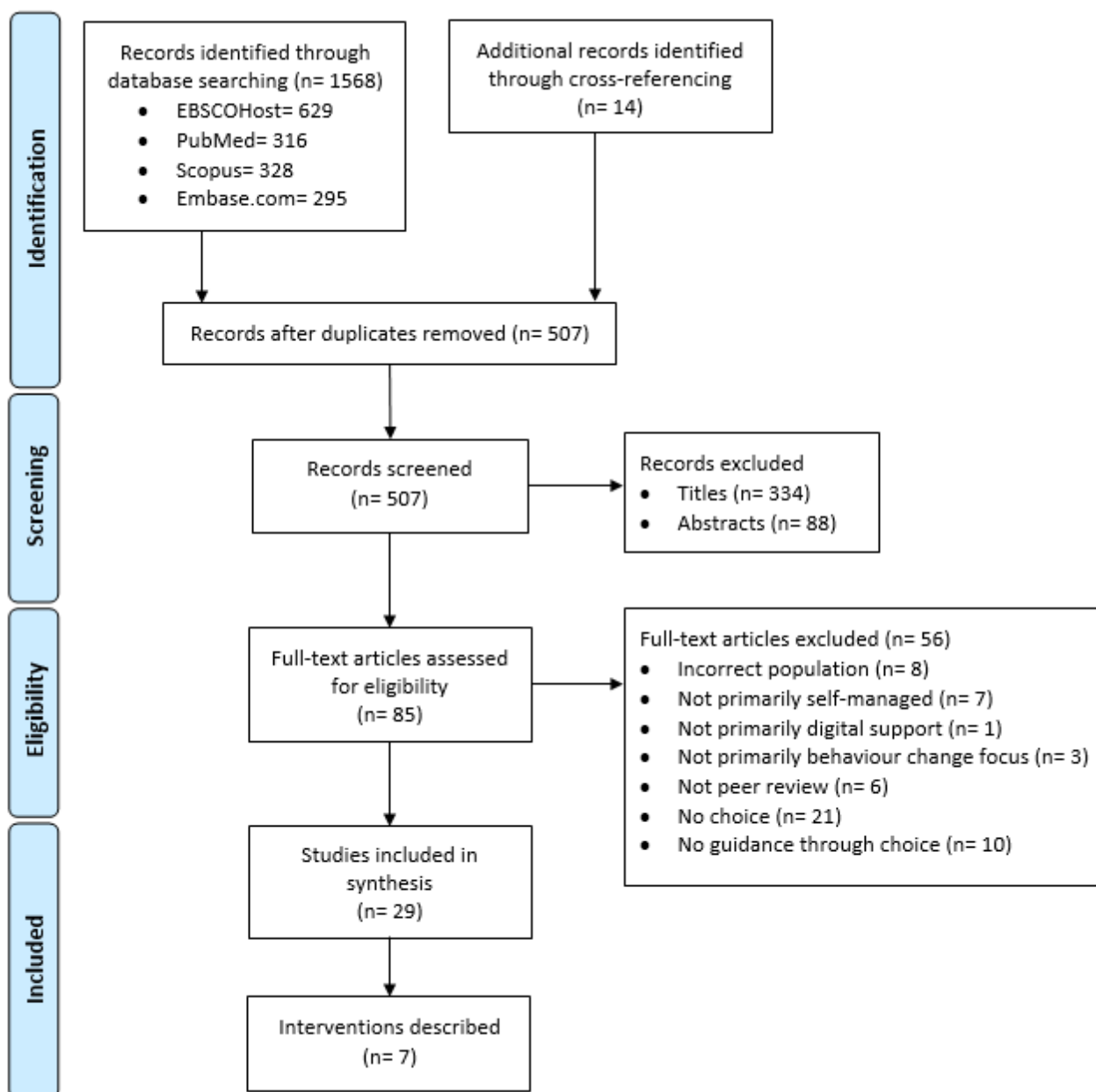
A systematic literature search of 9 electronic databases was conducted in January 2020: EBSCOhost (Academic Search Complete, Applied Science & Technology Source, Cumulative Index to Nursing and Allied Health Literature Complete, Global Health, Health Business Elite, and PsycINFO), PubMed (MEDLINE), Scopus, and EMBASE. Reference lists of included study publications and related conference proceedings were hand searched to identify additional publications that may not have appeared in database searches. Peer-reviewed systematic reviews and meta-analyses, in which the included studies were cited, were explored to identify any similar studies that did not appear in database searches.

Individual search strategies were developed for each database (example shown in [Multimedia Appendix 1](#)) and included search

terms derived from 3 main categories of interest: chronic disease, digital technology, and self-management of health behavior.

## Screening and Selection

All results were exported to a reference manager (EndNote X9, Clarivate Analytics), where duplicates were removed before screening for eligibility. Search results have been reported using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [62]. Each stage of the screening process is presented in [Figure 1](#). Titles, abstracts, and full texts were independently assessed by 2 reviewers (ST and JR), whereas the third and fourth reviewers (KB and RM) were consulted to collectively reach a consensus on studies with questionable eligibility.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

## Data Collection

### Quality

Methodological quality of the included literature was assessed using the Mixed Methods Appraisal Tool (MMAT; version 18) [63]. The MMAT allows for the simultaneous evaluation of qualitative, quantitative, and mixed-method research, which makes it an appropriate critical appraisal tool for this review.

Data were extracted from the publications using a specifically designed MMAT tool. Publications were not excluded based on quality assessment because of little empirical evidence to support this practice [63]. Instead, we decided to report on the quality of the reviewed studies.

The quality of each eligible study was rated according to 7 quality criteria. To determine the overall quality score for each study, the number of criteria met was divided by the total

number of criteria (7) and expressed as a percentage. It is discouraged to simply calculate an overall score [63]; thus, a more detailed report of the criteria was included to better inform the quality assessment of the included studies.

### Digital Platform-Like Interventions

The extraction of intervention data was guided by the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth) checklist [64], which outlines reporting guidelines for digital health interventions. Data on mode of delivery, participant information, comparators, intervention features and components, and theoretical frameworks or tools were collected. Data were also collected on the development processes of the digital platforms.

## Evaluation

According to highly regarded reports on monitoring and evaluation frameworks for digital health interventions [57,65-67], summative and process evaluations are typically carried out in line with the stage of maturity or specific characterization of a digital platform.

For this review, impact measures for effectiveness were classified as significant improvements in health behavior, clinical outcomes, or evidence-based psychosocial outcomes. Outcome measures for process evaluations were intervention use [68] (log-in data), adherence [68] (completion of follow-up data collection), and user satisfaction (quantitative or qualitative follow-up).

## Results

### Study Selection

The combined search strategy identified 1582 records, of which 1568 were identified through electronic database searching and 14 through other sources. Duplicates were removed, and the remaining 507 records were screened for eligibility through titles and abstracts. Full-text was obtained for the 85 remaining records, of which 29 publications outlining 7 digital platform-like interventions met the inclusion criteria (Table 2) [38,69-95]. The primary reasons for the exclusion of full-text records are shown in Figure 1. Table 2 shows how the 29 publications were related to each of the 7 studies, sorted according to the evaluation stage [57]. Owing to the heterogeneous nature of the included studies, a meta-analysis was not possible.

**Table 2.** Overview of included publications (n=29).

Study name	Proposal or protocol (n=7)		Formative evaluation		Summative evaluation						Formative evaluation	
	Year	Reference	Needs assess		Iterative development			Trial			Process evaluation	
			Qualitative (n=4)		Usability or feasibility (n=5)			Pilot or efficacy (n=5)			Mixed methods (n=4)	
			Year	Reference	Year	Reference	Year	Reference	Year	Reference	Year	Reference
Antypas and Wang-berg [71]	2012	[69]	2014a	[70]	— <sup>a</sup>	—	—	—	2014b	[71] <sup>b</sup>	—	—
Murray et al [88]	2015	[96]	2018 (Pal)	[38]	2019 (Dack)	[91]	2016 (Hofmann)	[86]	2017	[88] <sup>b</sup>	2017 (Alkal-di)	[87]
	—	—	—	—	—	—	—	—	—	—	2018 (Li)	[89]
	—	—	—	—	—	—	—	—	—	—	2018 (Podu-val)	[90]
Poppe et al [93]	2019a	[92]	2017	[72]	2018	[73]	2019b	[93] <sup>b</sup>	—	—	—	—
Sakakibara et al [94]	2019	[95]	—	—	—	—	2017	[94] <sup>b</sup>	—	—	—	—
Voncken-Brewster et al [77]	2013a	[74]	—	—	2013b	[75]	2014	[76]	2015	[77] <sup>b</sup>	2017	[78]
Walsh et al [82]	2017 (Claes)	[79]	2018	[80]	2019	[81]	2020 (Claes)	[82] <sup>b</sup>	—	—	—	—
Yu et al [85]	2012	[83]	—	—	2014a	[84]	—	—	2014b	[85] <sup>b</sup>	—	—

<sup>a</sup>Publication type not applicable to this study

<sup>b</sup>Parent publication for the study.

### Study Characteristics

This review identified 7 diverse digital platform-like interventions [38,69-95]. Of the 7 interventions, 3 targeted diabetes [85,88,93], 3 were intended for cardiovascular diseases

(CVDs) [71,82,94], and only one intervention supported people with chronic obstructive pulmonary disease (COPD) [77].

Of the 7 studies, 4 evaluated the effectiveness of their interventions [71,77,85,88], including 3 randomized control trials and 1 single-arm pre-post cohort design [85]. Sample sizes



ranged from 69 to 1325, and mean participant ages ranged from 57 to 65 years. Of the 4 studies, 2 reported gender imbalance within participant groups (23% [71] and 31% [88] female participants). Trial periods ranged from 3 to 12 months, and the control groups in the 3 randomized control designs received the usual care [77], accessed simple web-based information [88], or a variation of the described intervention [71].

Of the 7 studies, 3 had not evaluated effectiveness at the time of this review [82,93,94], but all 3 had completed a pilot evaluation of their intervention. Pilot sample sizes ranged from 35 to 120 and had relatively shorter trial periods of 5 weeks [93], 10 weeks [94], and 6 months [82]. One of the studies subsequently published a protocol for a planned randomized control trial [95] to be completed by December 2020 (clinical trial registration: NCT03159325; international registered report identifier: DERR1-10.2196/12322).

### Quality Appraisal

Parent publications for each study [71,77,82,85,88,93,94] were quality assessed; thus, the MMAT was completed once per study and not for each publication. Companion publications were assessed where a *Can't Tell* response was coded [38,69-96]. Overall, 6 of the digital platform-like interventions were intended to be used ad libitum. Therefore, the criterion "2.5. Did the participants adhere to the assigned intervention?" was not appropriate. This section was alternatively appraised based on the participants' completion of follow-up data collection.

The details of the assessment criteria can be found in [Multimedia Appendix 2](#) [71,77,82,85,88,93,94]. Overall, the scores for the included studies ranged from 3 to 7 out of a possible maximum of 7. This translates to a varied methodological quality range of 2 to 4 stars. The mixed-method studies were of the highest quality, achieving 3 [94] and 4 stars [85]. The most frequent limitations were incomplete outcome data sets [71,77,82,88,93], poor follow-up response rates, [71,77,82,88,93], and insufficient

integration of quantitative and qualitative components in mixed-method studies [94].

### Digital Platform-Like Interventions

The key characteristics of the included interventions are outlined in [Table 3](#). All interventions included an intermediary user interface, which allowed participants to explore and self-select optional components based on their individual needs and preferences. All 7 delivered a choice of at least two evidence-based behavior change techniques or strategies (eg, self-monitoring and goal setting) [97] targeting a lifestyle-related behavior (eg, physical inactivity and smoking). Physical activity was the most frequently targeted health behavior, addressed in 6 of the 7 digital health interventions [71,77,82,85,88,93]. Of the 7 interventions, 2 focused on disease-specific self-care as their key targeted behavior, which incorporates a wide range of health behaviors [85,94].

All 7 interventions presented algorithm-supported interventions that facilitated user guidance [71,77,82,85,88,93,94]. Specific features were included to optimize the sustainability of this automated approach, such as providing recommendations to users through computer-generated responses to questionnaires or self-reported data (ie, eDecision support tools) [73,77,82,88,93,94], smart recommendation widgets that offered advice based on individual patterns of use [85], and application of the transtheoretical model (TTM) of behavior change to select stage-matched intervention features [71].

Support from a health care professional was described in all 7 interventions: typically, cardiac rehabilitation nurse practitioners, exercise specialists, or a member of the research team. The primary role of health care professionals was to facilitate trial and data collection [71,77,82,85,88,93,94] or conduct initial familiarization sessions [71,82,88]. Overall, health care professionals did not provide digital support or enhance existing intervention features, as this approach was not considered sustainable [71,83].

**Table 3.** Overview of digital platform-like interventions.

Intervention name	Country	Mode of delivery	Target NCD <sup>a</sup>	Targeted health behaviors	Intended use
<b>Antypas and Wangberg [69-71]</b>					
Drupal	Norway	Web based and mobile based	CVD <sup>b</sup>	Physical activity, medication adherence, diet/nutrition, and smoking cessation	Ad libitum
<b>Murray et al [38,86-91,96]</b>					
HeLP-Diabetes	England	Web based and mobile based	T2D <sup>c</sup>	Physical activity, medication adherence, diet/nutrition, and smoking cessation	Ad libitum
<b>Poppe et al [72,73,92,93]</b>					
My Plan 2.0	The Netherlands	Web based and mobile based	T2D	Physical activity and sedentary behavior	1 session per week
<b>Sakakibara et al [94,95]</b>					
Healing Circles	Canada	Web based and mobile based	CVD	Multiple, unspecified: related to CVD <sup>a</sup>	Ad libitum
<b>Voncken-Brewster et al [74-78]</b>					
MasterYour Breath	The Netherlands	Web based	Chronic obstructive pulmonary disease	Physical activity, smoking cessation	Ad libitum
<b>Walsh et al [79-82]</b>					
PATHway	Ireland, Belgium	Web based and sensor based	CVD	Physical activity (primary) and other CVD-related	Ad libitum
<b>Yu et al [83-85]</b>					
Diabetes Online Companion	Canada	Web based	T2D	Multiple, unspecified: related to T2D	Ad libitum

<sup>a</sup>NCD: noncommunicable disease<sup>b</sup>CVD: cardiovascular disease.<sup>c</sup>T2D: type 2 diabetes.

## Development Processes

Of the 7 interventions, 6 had a clear theoretical basis to support the concept and development, including behavior change theories such as social cognitive theory (SCT) [81,83,88], social ecological model [88], the Health Action Process Approach (HAPA) [69,73], and the wide-ranging integrated change model (iChange) [69,74]. Of the 7 interventions, 1 was not explicit about the theoretical underpinning but was influenced by aspects of SCT [95]. Of the 7 studies, 3 [82,88,93] outlined platform-like intervention components according to the behavior change technique taxonomy of Michie et al [97].

Of the 7 studies, 4 conducted an early stage needs assessment before commencing development [38,70,72,80] to verify the unmet requirements of their target NCD population [38,70,80] and validate their intervention concept [72]. These needs assessments took the form of focus groups [38,70], semistructured interviews [80], or think-aloud sessions with existing digital infrastructures [72].

Of the 7 digital interventions, 4 took an iterative development approach [75,81,84,91], which involved target participants with NCD throughout the design process of the digital platform-like intervention. This approach included at least three iterative

cycles [75,81,84,91], and in some cases, the process was guided by participatory design principles [81,91].

In contrast, just one of the 7 interventions did not involve participants in the development process [94], choosing to focus on key theories and researcher expertise to conceptualize and create the Healing Circles intervention.

## Evaluation

All 7 digital platform-like interventions were evaluated (Multimedia Appendix 3; [71,77,82,85,88,93,94]) through either preliminary pilot investigations [82,93,94] or effectiveness trials [71,77,85,88]. Of the 7 interventions, 2 used evidence-based self-report measures for data collection [71,77]; 2 other studies collected data during researcher visits to participants [82,93] and from hip-worn ActiGraph accelerometers; and 3 collected data through a combination of existing medical records [85,94], self-report measures [85,88,94], and researcher visits to participants [85,88].

## Behavioral Outcome Measures

Of the 7 studies, 4 reported on physical activity behavior change [71,77,82,93], one of which demonstrated significant improvements in overall physical activity (median 5613 metabolic equivalents of task-minutes [MET-min] per week,

IQR 2828) compared with the control group at 3 months (median 1356 MET-min per week, IQR 2937) [71]. For specific intensities, changes in physical activity behavior for the intervention group were significantly better than that for the control for walking only (+453.8 MET-min per week;  $P=.05$ ) [71]. Accelerometer-assessed moderate-to-vigorous physical activity (MVPA) behavior was significantly improved from baseline in 2 of the digital platform-like interventions [82,93]. PATHway reported an increase of approximately 14 min MVPA per day ( $P=.04$ ) at 6 months [82], and the My Plan 2.0 intervention group increased MVPA from baseline by approximately 8 min per day [93]. Both interventions resulted in medium-to-large-interaction effects between the intervention and control groups [82,93].

Change in self-care behavior was the primary outcome for one of the studies that targeted diabetes [85]. The Summary of Diabetes Self-Care Activities measure [98] indicated that significant improvements in diabetes self-care behavior and sustained changes (9 months from baseline) were positively correlated with age (+0.04 per year, 95% CI 0.02-0.06;  $P<.001$ ) [85].

Improvements in self-monitoring behavior were observed in 2 of the 7 interventions [93,94]. My Plan 2.0 found a significant time-group intervention effect favoring the intervention group for self-monitoring (effect size 0.54;  $P=.008$ ) [93]. At 10 weeks, the Healing Circles intervention also reported a time effect for the intervention group ( $z=-2.04$ ;  $P=.04$ ) [94]. The effect of Healing Circles on self-monitoring behavior was emulated by the self-management domains of health behaviors ( $z=-2.11$ ;  $P=.04$ ) and social support ( $z=-2.58$ ,  $P=.01$ ), all 3 of which were assessed through the Health Education Impact Questionnaire to measure the changes in self-management behavior [99].

Smoking cessation [77], dietary changes [82], medication adherence [82], alcohol use [82], and stages and mediators of health behavior change [71,77,82] were examined in 3 of the 7 studies, but no significant effects were found.

### Clinical Outcome Measures

Disease-specific clinical outcomes were reported in 4 studies at 6 months [77,82], 9 months [85], and 12 months [88]. One of the publications demonstrated significant improvements in clinical disease control ( $-0.06$ , 95% CI  $-0.11$  to  $-0.01$ ;  $P=.01$ ) using the 10-item Clinical COPD Questionnaire [100]; however, this effect was not maintained when corrected for participant baseline characteristics, such as age, sex, and disease status [77]. At 6 months, a CVD-focused study [82] reported medium-sized group interactions in favor of the intervention for diastolic blood pressure (effect size  $-0.49$ ;  $P=.004$ ) and cardiovascular risk (effect size  $-0.36$ ;  $P=.03$ ; Framingham Risk Score [101]). One platform-like intervention for diabetes reported lowered glycated hemoglobin ( $HbA_{1c}$ ) in the intervention group at 12 months compared with the control group who had access to web-based information [88]. Further causal analyses of *high-usage* participants within the intervention group indicated that intervention use for greater than or equal to a median of 4 days could potentially reduce  $HbA_{1c}$  levels by 0.44% over 12 months [88]. Another study

measured weight, diabetes-related blood markers, and blood pressure, but no overall positive effects were identified [85].

### Psychosocial Outcome Measures

Health-related quality of life, disease-related quality of life [77,82,85,94], and self-efficacy [71,82,85,88,93] were the most prevalent psychosocial outcomes collected across the 7 studies, but 6 of the 7 reported no significant change. Significant improvements were recorded for disease-related quality of life (using Diabetes Distress Scale [102]) in the diabetes companion intervention group [85] when comparing users ( $n=70$ ) with nonusers ( $n=11$ ) at 9 months ( $-4.7$  vs  $-0.9$ ;  $P<.001$ ).

### Process Evaluation

Log-in data were measured in 6 of the 7 interventions [71,73,78,82,85,88], and adherence (completion of follow-up data collection) was measured in all 7 interventions [71,77,82,85,88,93,94]. One of the 7 interventions reported that 86% of participants logged in to the intervention at least once over 9 months [85]. Of these participants, 75% were classified as infrequent users ( $<2$  log-ins per month) with an average of one log in per user per month [85] and an average time of 6 min spent per log in. One of the 7 interventions explained that although participant log-in rates averaged 8 sessions per month (range 0.6-3.6 sessions per week) [82], usage dropped significantly in the final 2 months of the 6-month trial ( $P<.001$ ). Another one of the 7 interventions [78] recorded a baseline log-in of 59.5% for the intervention group, which was significantly higher than the initiation rates for the control group program (12.8%). This disparity between the intervention and control groups was also reported in another intervention [88]. Mean log-in values were significantly higher in the intervention group (18.7 vs 4.8;  $P<.001$ ), averaging about 1.5 log-ins per month over the 12-month trial [88]. One of the intervention protocols required users to log in 5 times over 5 weeks [73]. Overall, 92% of the participants logged in at least once, and the average total time spent using the intervention was about 49 min [73]. One of the 7 interventions intended to measure log-in rates [71], but unforeseen technical issues meant that some data were unreliable.

Overall, the mean adherence to intervention trials at follow-up was 69% (SD 20%). In 4 of the 7 interventions, approximately 80% of participants successfully completed follow-up data collection at 6 months [77,82], 9 months [85], and 12 months [88]. For 2 shorter trial periods, over 60% of participants adhered to data collection following a 5-week [93] and 10-week [94] intervention. In contrast, another intervention demonstrated a low responder rate of just 27% 3 months from baseline [71]. Of the 7 studies, 4 used email and SMS reminders to secure follow-up data [71,77,88,94], 3 of which reported adherence rates of over 60% [77,88,94].

Of the 7 interventions, 6 reported on user satisfaction, which was evaluated through both qualitative and quantitative measures 71,73,78,85,88,94

Of the 7 studies, 4 used semistructured exit interviews to gauge user satisfaction [73,78,85,94] and conducted a quantitative follow-up for satisfaction at the end of the trial by asking the user if they would recommend the intervention to a friend

[71,78] or by assessing changes in the diabetes treatment satisfaction questionnaire [88]. One of the interventions used a mixed-method approach [78] using both data collection methods.

Participants commented that they were satisfied with the intervention layout, navigation, and ease of use [73,78,85]. The interventions were seen as evidence-based, authoritative sources [78,85], which enhanced users' accountability for their self-management and behavior change goals [73,85]. One participant group noted that accountability could have been further enhanced by including more social support (eg, intervention access for family members or friends) [73], which was supported by user satisfaction for a social support component of another intervention (peer-to-peer web-based interaction) [85]. The personal relevance of the platform-like interventions was noted as a key contributor to overall user satisfaction [73,78,85], driven by users' individual contexts and circumstances. Moreover, users were also satisfied with the interventions' ability to accommodate a broad range of user needs [73,85]. Although users valued the personal relevance of the interventions, there was some dissatisfaction with the burden of a high volume of questions associated with digital guidance tools [73,78]. User dissatisfaction was also reported for interventions that did not offer a mobile delivery option [78,85] and by users who were self-managing more than one chronic condition [78,85]. User satisfaction outcomes were mixed for the 3 quantitative evaluations. Overall, 57% [78] and 68% [71] of participants would reportedly recommend the intervention to a friend, but the HeLP-Diabetes intervention [88] reported no significant difference between intervention and control group satisfaction outcomes.

## Discussion

### Principal Findings

This review is the first of its kind to systematically examine the literature on digital platform-like interventions for the self-management of NCDs and health behavior change. Although no digital platforms were identified, an examination of digital platform-like interventions has contributed knowledge to the conceptualization and potential value of using digital platform architectures to support self-management of NCDs.

A total of 7 digital platform-like interventions were included in the review. Evaluations for effectiveness were disparate, and so a consensus on the overall effect could not be reached. Nevertheless, positive effects were reported for physical activity, disease-specific self-care, and self-monitoring behaviors, which is a promising finding in support of digital platform use for the self-management of NCDs. As a result of the findings mentioned earlier, 3 dominant themes emerged: development, optimizing change, and support and guidance for users.

### Development

This review found that comprehensive and systematic development processes were implemented for most platform-like interventions. Iterative, user-centered approaches are highly regarded in the field of digital health research because the proactive engagement of patients can be beneficial in the development of digital interventions [103,104]. It allows

researchers to gain an in-depth understanding of the psychosocial context of the potential end users of the intervention [105] and shortens the communicative distance between the researcher and the user [106]. In this review, 4 of the 7 interventions [75,81,84,91] adopted an iterative, user-centered approach. Reporting methods across the included interventions were varied, so it was difficult to determine whether incorporating user-centered development was associated with greater effectiveness. However, research suggests that such development methodologies have the most potential for developing a sustainable solution [107]. An average of 80% adherence at follow-up stages was found in the same 4 studies [77,82,85,88], which could be attributed to the comprehensive iterative development approaches used in all 4 of the interventions. The level of user engagement with digital health technology tends to wane over time [56,108,109], which has been associated with a lack of perceived value by the user and increased burden because of irrelevant material [110,111]. Placing the user at the center of development considerations and including them in decision-making processes may have affected the creation of fitting platform-like structures, appropriate for those living with NCDs and delivering relevant, usable content or features.

All 4 of the aforementioned interventions [75,81,84,91] also specified a clear theoretical underpinning to their development (eg, TTM), which is consistent with research recommendations that intervention content and features associated with a solid theory base are more likely to be effective in changing behavior [46]. A strong theoretical framework may also have affected the generalizability and adaptability of the platform-like interventions, further adding to the potential for efficacy and longevity [66,112,113] as seen through high adherence rates. Only 3 of the studies outlined platform-like intervention components according to a taxonomy [82,88,93], which made identifying behavior change features difficult because of vague reporting. We anticipate that digital platform content would have superior clarity for users and health care professionals alike because of the stand-alone, discrete digital interventions offered.

### Optimizing Change

The results presented in this review indicate the potential for digital platforms to affect behavior change, such as disease-specific self-care behavior. Significant differences were reported for sustained diabetes self-care behavior between users and nonusers of the Diabetes Online Companion intervention [85].

The self-care outcome measure covers a range of health behaviors, which address several key lifestyle risk factors (eg, dietary behavior, physical activity, glucose monitoring, and smoking) [98]. Thus, our findings suggest that modest changes in multiple health behaviors (ie, changes in self-care) may have a better overall and sustained effect on the self-management of NCDs than larger effects in one single health behavior. This suggestion supports the opinion that affecting comprehensive lifestyle change may be a better approach to the self-management of NCDs, as NCDs are complex conditions influenced by several interconnected lifestyle risk factors



[3,114]. Similar to self-care, improvements in self-monitoring behavior were identified through the HAPA, which incorporates six personal determinants of behavior change.

Lifestyle-related risk factors and their corresponding health behaviors are heavily influenced by one another, and thus modifying one health behavior using a digital platform may not necessarily generate sustainable improvements in the self-management of NCDs. For example, smoking behavior has been inversely linked to physical activity behavior, meaning that a lack of engagement with smoking cessation may inhibit physical activity progress [115-117]. Similarly, poor stress management may have a negative effect on engagement with healthy behaviors, such as smoking cessation [118], physical activity [119,120], and maintaining a healthy diet [121]. Single interventions that independently target diet and physical activity behavior may improve those isolated behaviors, but sustained changes in associated risk factors (eg, weight loss and maintenance) are more probable when interventions simultaneously target both health behaviors [114]. The results of our review showed limited to no changes in isolated behaviors (eg, overall physical activity and smoking) [77,82], but modest improvements were apparent in overall disease-specific self-care and multiple self-management domains [85,93,94].

Focusing on modest overall improvements in health behavior change may produce further consequential effects on sustained NCD self-management. Self-regulatory and self-efficacy theories for promoting self-management suggest that the greatest improvements in self-care of chronic conditions typically occur following some initial success in changing behavior (ie, mastery experience) [14]. Thus, seeking a modest change in a more comprehensive and generalizable outcome measure such as disease-specific self-care may generate a cascade of behavior change improvements moving forward.

A comprehensive approach could also provide a greater scope for successful self-management throughout the unpredictable health trajectories of NCDs, which are relatively unique to each individual [122]. This hypothesis is in line with primary care digital health frameworks that dynamically adapt services to the clinical care pathway of the individual, which cannot be predetermined and changes regularly [123]. This approach supports our hypothesis that overly specific digital health interventions may not be the most appropriate solutions to comprehensive self-management of NCDs. A digital platform could accommodate not only the stage of NCD but also the current health state of the person living with that NCD, by providing the self-management tools required to suit individual circumstances. It would create an opportunity for users to experiment and *tinker* with evidence-based interventions [124], and experience successes and failures on their path to successful self-management, leading to new insights and skills in self-care. The evaluation of a digital platform should extend further than simply adhering to the discrete digital interventions hosted within the platform and should explore how they are used and actively incorporated into everyday life [104]. Longer evaluation trial periods may better facilitate these dynamic user health trajectories to accommodate periods of exploration and changing mindsets.

## Support and Guidance

The digital platform-like interventions included in this review had good adherence for up to 9 months, which could be attributed to the freedom for users to explore an array of digital features and identify the most relevant components for them as an individual. However, given that a digital platform would provide access to an extensive choice of discrete digital health interventions, it is important to note that the intention is not to overburden users [125]. It is possible that free navigation through a platform with various features to negotiate could be a cumbersome process [110].

Research indicates that using all available components is not necessarily more effective [56], and presenting a platform not matched to user needs could be detrimental to the overall success of the platform [126]. Thus, evidence suggests that the addition of human support in digital health can enhance user engagement, as users value reassurance and expert knowledge to guide their decision making [66,127,128]. Human support is resource intensive and can increase the overall cost and burden on health care systems, which may not be a feasible solution. Therefore, it is important to establish the extent to which further value can be added to a multicomponent digital health intervention by supplementing it with human support [127].

In view of this, our review has recognized the potential of applying existing theories and knowledge to create efficient, cost-effective automated guidance for users to make informed choices about their engagement. The digitally supported guidance offered to participants within the 7 interventions provided them with a supported pathway to choose behavior change components. This is in line with behavioral research that indicates that actively engaging participants in decision-making pathways for care can improve health outcomes [129] through factors such as improved autonomy [37], empowerment, and mastery [130]. These factors are critical in supporting people to self-manage their disease and in promoting a more *digitally engaged* patient [104].

One reason that people value human support is a sense of accountability [127]. However, this review showed that a sense of accountability was successfully acquired by several participants using automated platform-like interventions. Providing clear associations between a digital platform and *expert* health care professional advice may be enough to satisfy the desire for accountability. This is reinforced by research that suggests a key factor in promoting engagement with digital health interventions is clear endorsement by respected clinicians or *expert* organizations [110,127]. A digital platform would seek to use existing, evidence-based digital health interventions that could be put through a rigorous selection process before inclusion, driven by an extensive list of key stakeholders (eg, health care professionals, potential end users, and family members) to further endorse the trustworthiness of the platform [110].

The comprehensive development of these evidence-based, automated guiding decision pathways is another way to incorporate *expert* advice and guidance to support the trustworthiness of a digital platform. Automated eDecision support tools are currently being developed and validated to



assist with meaningful adherence to interventions and health behavior by providing individualized, real-time assistance [81,131]; however, no validated tools have been described in the included literature to support navigation through complex digital platforms. This may have an impact on the usability or trustworthiness of the included digital platform-like interventions. The variation in guidance systems presented in this review points to a requirement for more research on an operative level of support to optimize engagement and create a user-centered experience for the individual.

In summary, the adaptive nature of a digital platform accommodates the requirement for flexibility in self-management support, which could facilitate a diverse range of users and life circumstances [66,132]. Longitudinal and detailed evaluations of digital platforms must be carried out to influence the lifetime of positive health-related choices and behaviors. It is important not only to evaluate the longevity of participating in such an intervention but also to explore the diverse patterns of engagement by a wide range of users.

### Limitations

This review was prospectively registered, used comprehensive search strategies across multiple databases, and reported according to PRISMA. Searches and publication of results have been conducted in an up-to-date and timely fashion. The inclusive nature of this review accommodated a broad range of NCDs, which enhances the generalizability of the findings among a wider population.

This review was not without limitations. The novel concept of a digital platform was introduced for the first time, which made it difficult to identify relevant literature. Having broadened the inclusion criterion, discrepancies in terminologies could still have led to key studies being omitted from this literature search or the misinterpretation of intervention content. Current research is focused on the development of a comprehensive eHealth taxonomy, but this is not yet wide ranging enough [133]. In

response to this, a systematic approach and extensive use of terminology in search strategies were implemented to ensure that pertinent literature was included for review.

Another limitation was that none of the interventions included in this review were readily available on the web to the reviewer, none provided a digital preservation URL, and intervention descriptions were not always sufficient to identify key features. There has been a call for improved reporting of digital health interventions to improve examination and evaluation of intervention characteristics. This limitation is an important reminder that future developments should use reputable reporting frameworks and guidelines to outline their work [64,134].

### Conclusions

We have identified a gap in the research on comprehensive and flexible digital health for the self-management of NCDs. Thus, we proposed the contemporary concept of a digital platform, which supports the innovative use of already existing digital interventions for health behavior change. Initial searches indicated that no such digital platforms currently exist, which may indicate a missed opportunity to optimize user engagement with already developed, evidence-based digital interventions. In response, this review focused on digital platform-like interventions to provide an understanding of the development and contextual considerations required to optimally construct a digital platform.

Iterative, user-centered development may be associated with improved adherence and sustained use. Offering a palette of evidence-based interventions with an appropriate degree of guidance has the potential to facilitate disease-specific health behavior change and effective self-management among a myriad of users, conditions, or stages of care. This review and the novel concept of digital platform-like interventions contribute new knowledge to the rhetoric of digital health for the self-management of NCDs.

---

### Acknowledgments

This work was funded by the Institute for Physical Activity and Nutrition and the Deakin University Postgraduate Research Scholarship fund.

---

### Authors' Contributions

ST contributed to the study design, acquisition, and analysis of data as well as leading the writing of the manuscript. RM was involved in the conceptualization of the study. ST and JR independently screened literature, with RM and KB resolving discrepancies. RM, JR, KB, LK, and FK each reviewed and provided significant feedback on the manuscript. All authors have read and approved the final version of the manuscript.

---

### Conflicts of Interest

None declared.

---

#### Multimedia Appendix 1

Search strategy for EbscoHost.

[PDF File (Adobe PDF File), 111 KB - [jmir\\_v22i10e16774\\_app1.pdf](#)]

---

#### Multimedia Appendix 2

Star Rating Quality scoring of included studies. Mixed Methods Appraisal Tool (MMAT)-Version 2018.

[[PDF File \(Adobe PDF File\), 116 KB - jmir\\_v22i10e16774\\_app2.pdf](#)]

Multimedia Appendix 3

Overview of summative evaluations.

[[PDF File \(Adobe PDF File\), 155 KB - jmir\\_v22i10e16774\\_app3.pdf](#)]

## References

1. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017 Sep 16;390(10100):1151-1210 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(17\)32152-9](https://doi.org/10.1016/S0140-6736(17)32152-9)] [Medline: [28919116](https://pubmed.ncbi.nlm.nih.gov/28919116/)]
2. GBD 2016 DALYs HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017 Sep 16;390(10100):1260-1344 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(17\)32130-X](https://doi.org/10.1016/S0140-6736(17)32130-X)] [Medline: [28919118](https://pubmed.ncbi.nlm.nih.gov/28919118/)]
3. GBD 2016 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017 Sep 16;390(10100):1345-1422 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(17\)32366-8](https://doi.org/10.1016/S0140-6736(17)32366-8)] [Medline: [28919119](https://pubmed.ncbi.nlm.nih.gov/28919119/)]
4. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, ESC Scientific Document Group. *Eur Heart J* 2016 Dec 01;37(29):2315-2381 [[FREE Full text](#)] [doi: [10.1093/eurheartj/ehw106](https://doi.org/10.1093/eurheartj/ehw106)] [Medline: [27222591](https://pubmed.ncbi.nlm.nih.gov/27222591/)]
5. Barquera S, Pedroza-Tobías A, Medina C, Hernández-Barrera L, Bibbins-Domingo K, Lozano R, et al. Global Overview of the Epidemiology of Atherosclerotic Cardiovascular Disease. *Arch Med Res* 2015 Jul;46(5):328-338. [doi: [10.1016/j.arcmed.2015.06.006](https://doi.org/10.1016/j.arcmed.2015.06.006)] [Medline: [26135634](https://pubmed.ncbi.nlm.nih.gov/26135634/)]
6. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364(9438):937-952. [doi: [10.1016/S0140-6736\(04\)17018-9](https://doi.org/10.1016/S0140-6736(04)17018-9)] [Medline: [15364185](https://pubmed.ncbi.nlm.nih.gov/15364185/)]
7. Ezzati M, Vander Hoorn S, Lawes CMM, Leach R, James WPT, Lopez AD, et al. Rethinking the “Diseases of Affluence” Paradigm: Global Patterns of Nutritional Risks in Relation to Economic Development. *PLoS Med* 2005 May;2(5):e133 [[FREE Full text](#)] [doi: [10.1371/journal.pmed.0020133](https://doi.org/10.1371/journal.pmed.0020133)] [Medline: [15916467](https://pubmed.ncbi.nlm.nih.gov/15916467/)]
8. Rich MW, Beckham V, Wittenberg C, Leven CL, Freedland KE, Carney RM. A multidisciplinary intervention to prevent the readmission of elderly patients with congestive heart failure. *N Engl J Med* 1995 Nov 02;333(18):1190-1195. [doi: [10.1056/NEJM199511023331806](https://doi.org/10.1056/NEJM199511023331806)] [Medline: [7565975](https://pubmed.ncbi.nlm.nih.gov/7565975/)]
9. Dalal HM, Doherty P, Taylor RS. Cardiac rehabilitation. *BMJ* 2015 Sep 29;351:h5000. [Medline: [26419744](https://pubmed.ncbi.nlm.nih.gov/26419744/)]
10. Gardetto NJ. Self-management in heart failure: where have we been and where should we go? *J Multidiscip Healthc* 2011 Mar 31;4:39-51 [[FREE Full text](#)] [doi: [10.2147/JMDH.S8174](https://doi.org/10.2147/JMDH.S8174)] [Medline: [21544247](https://pubmed.ncbi.nlm.nih.gov/21544247/)]
11. Norris SL, Engelgau MM, Narayan KM. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. *Diabetes Care* 2001 Mar;24(3):561-587. [doi: [10.2337/diacare.24.3.561](https://doi.org/10.2337/diacare.24.3.561)] [Medline: [11289485](https://pubmed.ncbi.nlm.nih.gov/11289485/)]
12. Zwerink M, Brusse-Keizer M, van der Valk PDLPM, Zielhuis GA, Monninkhof EM, van der Palen J, et al. Self management for patients with chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2014 Mar 19;3(3):CD002990 [[FREE Full text](#)] [doi: [10.1002/14651858.CD002990.pub3](https://doi.org/10.1002/14651858.CD002990.pub3)] [Medline: [24665053](https://pubmed.ncbi.nlm.nih.gov/24665053/)]
13. Piepoli MF, Corrà U, Adamopoulos S, Benzer W, Bjarnason-Wehrens B, Cupples M, et al. *Eur J Prev Cardiol* 2014 Jun;21(6):664-681. [doi: [10.1177/2047487312449597](https://doi.org/10.1177/2047487312449597)] [Medline: [22718797](https://pubmed.ncbi.nlm.nih.gov/22718797/)]
14. Bandura A. Health promotion by social cognitive means. *Health Educ Behav* 2004 Apr;31(2):143-164. [doi: [10.1177/1090198104263660](https://doi.org/10.1177/1090198104263660)] [Medline: [15090118](https://pubmed.ncbi.nlm.nih.gov/15090118/)]
15. Buckley J, Doherty P, Furze G, Jones J, Hinton S, Hayward J. Standards and core components for cardiovascular disease prevention and rehabilitation. 2017 Jul. URL: [https://www.bacpr.com/resources/BACPR\\_Standards\\_and\\_Core\\_Components\\_2017.pdf](https://www.bacpr.com/resources/BACPR_Standards_and_Core_Components_2017.pdf) [accessed 2018-12-19]
16. Woodruffe S, Neubeck L, Clark RA, Gray K, Ferry C, Finan J, et al. Australian Cardiovascular Health and Rehabilitation Association (ACRA) core components of cardiovascular disease secondary prevention and cardiac rehabilitation 2014. *Heart Lung Circ* 2015 May;24(5):430-441. [doi: [10.1016/j.hlc.2014.12.008](https://doi.org/10.1016/j.hlc.2014.12.008)] [Medline: [25637253](https://pubmed.ncbi.nlm.nih.gov/25637253/)]
17. Spruit MA, Singh SJ, Garvey C, ZuWallack R, Nici L, Rochester C, ATS/ERS Task Force on Pulmonary Rehabilitation. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. *Am J Respir Crit Care Med* 2013 Oct 15;188(8):e13-e64. [doi: [10.1164/rccm.201309-1634ST](https://doi.org/10.1164/rccm.201309-1634ST)] [Medline: [24127811](https://pubmed.ncbi.nlm.nih.gov/24127811/)]
18. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Fischl AH, et al. Diabetes Self-management Education and Support in Type 2 Diabetes: A Joint Position Statement of the American Diabetes Association, the American Association

- of Diabetes Educators, and the Academy of Nutrition and Dietetics. *Diabetes Educ* 2017 Feb;43(1):40-53. [doi: [10.1177/0145721716689694](https://doi.org/10.1177/0145721716689694)] [Medline: [28118121](https://pubmed.ncbi.nlm.nih.gov/28118121/)]
19. World Health Organization. Global Report on Diabetes. In: Roglic G, editor. <https://www.who.int/diabetes/global-report/en/>. Geneva, Switzerland: WHO Press; 2016:1-86.
  20. Funnell MM, Brown TL, Childs BP, Haas LB, Hosey GM, Jensen B, et al. National standards for diabetes self-management education. *Diabetes Care* 2009 Jan;32 Suppl 1:S87-S94 [FREE Full text] [doi: [10.2337/dc09-S087](https://doi.org/10.2337/dc09-S087)] [Medline: [19118294](https://pubmed.ncbi.nlm.nih.gov/19118294/)]
  21. Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, et al. Management of hyperglycemia in type 2 diabetes, 2015: a patient-centered approach: update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* 2015 Jan;38(1):140-149. [doi: [10.2337/dc14-2441](https://doi.org/10.2337/dc14-2441)] [Medline: [25538310](https://pubmed.ncbi.nlm.nih.gov/25538310/)]
  22. Anderson L, Oldridge N, Thompson DR, Zwisler A, Rees K, Martin N, et al. Exercise-Based Cardiac Rehabilitation for Coronary Heart Disease: Cochrane Systematic Review and Meta-Analysis. *J Am Coll Cardiol* 2016 Jan 05;67(1):1-12 [FREE Full text] [doi: [10.1016/j.jacc.2015.10.044](https://doi.org/10.1016/j.jacc.2015.10.044)] [Medline: [26764059](https://pubmed.ncbi.nlm.nih.gov/26764059/)]
  23. Deakin T, McShane CE, Cade JE, Williams RDRR. Group based training for self-management strategies in people with type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2005 Apr 18(2):CD003417. [doi: [10.1002/14651858.CD003417.pub2](https://doi.org/10.1002/14651858.CD003417.pub2)] [Medline: [15846663](https://pubmed.ncbi.nlm.nih.gov/15846663/)]
  24. Jordan RE, Majothi S, Heneghan NR, Blissett DB, Riley RD, Sitch AJ, et al. Supported self-management for patients with moderate to severe chronic obstructive pulmonary disease (COPD): an evidence synthesis and economic analysis. *Health Technol Assess* 2015 May;19(36):1-516 [FREE Full text] [doi: [10.3310/hta19360](https://doi.org/10.3310/hta19360)] [Medline: [25980984](https://pubmed.ncbi.nlm.nih.gov/25980984/)]
  25. Bjarnason-Wehrens B, McGee H, Zwisler A, Piepoli MF, Benzer W, Schmid J, et al. Cardiac rehabilitation in Europe: results from the European Cardiac Rehabilitation Inventory Survey. *Eur J Cardiovasc Prev Rehabil* 2010 Aug;17(4):410-418. [doi: [10.1097/HJR.0b013e328334f42d](https://doi.org/10.1097/HJR.0b013e328334f42d)] [Medline: [20300001](https://pubmed.ncbi.nlm.nih.gov/20300001/)]
  26. Dunlay SM, Pack QR, Thomas RJ, Killian JM, Roger VL. Participation in cardiac rehabilitation, readmissions, and death after acute myocardial infarction. *Am J Med* 2014 Jun;127(6):538-546 [FREE Full text] [doi: [10.1016/j.amjmed.2014.02.008](https://doi.org/10.1016/j.amjmed.2014.02.008)] [Medline: [24556195](https://pubmed.ncbi.nlm.nih.gov/24556195/)]
  27. Horgan G, Davies M, Findlay-White F, Chaney D, Coates V. Reasons why patients referred to diabetes education programmes choose not to attend: a systematic review. *Diabet Med* 2017 Jan;34(1):14-26. [doi: [10.1111/dme.13120](https://doi.org/10.1111/dme.13120)] [Medline: [26996982](https://pubmed.ncbi.nlm.nih.gov/26996982/)]
  28. Doherty PJ, Harrison AS. The National Audit of Cardiac Rehabilitation: Annual Statistical Report 2017. British Heart Foundation. 2017. URL: <https://www.bhf.org.uk/informationsupport/publications/statistics/national-audit-of-cardiac-rehabilitation-annual-statistical-report-2017> [accessed 2019-02-19]
  29. Rochester CL, Vogiatzis I, Holland AE, Lareau SC, Marciniuk DD, Puhan MA, et al. An Official American Thoracic Society/European Respiratory Society Policy Statement: Enhancing Implementation, Use, and Delivery of Pulmonary Rehabilitation. *Am J Respir Crit Care Med* 2015 Dec 1;192(11):1373-1386. [doi: [10.1164/rccm.201510-1966ST](https://doi.org/10.1164/rccm.201510-1966ST)] [Medline: [26623686](https://pubmed.ncbi.nlm.nih.gov/26623686/)]
  30. Ades PA, Keteyian SJ, Wright JS, Hamm LF, Lui K, Newlin K, et al. Increasing Cardiac Rehabilitation Participation From 20% to 70%: A Road Map From the Million Hearts Cardiac Rehabilitation Collaborative. *Mayo Clin Proc* 2017 Feb;92(2):234-242 [FREE Full text] [doi: [10.1016/j.mayocp.2016.10.014](https://doi.org/10.1016/j.mayocp.2016.10.014)] [Medline: [27855953](https://pubmed.ncbi.nlm.nih.gov/27855953/)]
  31. Turk-Adawi K, Sarrafzadegan N, Grace SL. Global availability of cardiac rehabilitation. *Nat Rev Cardiol* 2014 Jul 15;11(10):586-596. [doi: [10.1038/nrcardio.2014.98](https://doi.org/10.1038/nrcardio.2014.98)]
  32. Neubeck L, Freedman SB, Clark AM, Briffa T, Bauman A, Redfern J. Participating in cardiac rehabilitation: a systematic review and meta-synthesis of qualitative data. *Eur J Prev Cardiol* 2012 Jun;19(3):494-503. [Medline: [22779092](https://pubmed.ncbi.nlm.nih.gov/22779092/)]
  33. Davies P, Taylor F, Beswick A, Wise F, Moxham T, Rees K, et al. Promoting patient uptake and adherence in cardiac rehabilitation. *Cochrane Database Syst Rev* 2010 Jul 07(7):CD007131 [FREE Full text] [doi: [10.1002/14651858.CD007131.pub2](https://doi.org/10.1002/14651858.CD007131.pub2)] [Medline: [20614453](https://pubmed.ncbi.nlm.nih.gov/20614453/)]
  34. Suaya JA, Shepard DS, Normand ST, Ades PA, Protas J, Stason WB. Use of cardiac rehabilitation by Medicare beneficiaries after myocardial infarction or coronary bypass surgery. *Circulation* 2007 Oct 9;116(15):1653-1662. [doi: [10.1161/CIRCULATIONAHA.107.701466](https://doi.org/10.1161/CIRCULATIONAHA.107.701466)] [Medline: [17893274](https://pubmed.ncbi.nlm.nih.gov/17893274/)]
  35. Clark AM, King-Shier KM, Thompson DR, Spaling MA, Duncan AS, Stone JA, et al. A qualitative systematic review of influences on attendance at cardiac rehabilitation programs after referral. *Am Heart J* 2012 Dec;164(6):835-45.e2. [doi: [10.1016/j.ahj.2012.08.020](https://doi.org/10.1016/j.ahj.2012.08.020)] [Medline: [23194483](https://pubmed.ncbi.nlm.nih.gov/23194483/)]
  36. Harrison AS, Doherty P, Phillips A. An analysis of barriers to entry of cardiac rehabilitation in patients with diabetes: Using data from the National Audit of Cardiac Rehabilitation. *Diab Vasc Dis Res* 2018 Mar;15(2):145-149. [doi: [10.1177/1479164117744490](https://doi.org/10.1177/1479164117744490)] [Medline: [29205066](https://pubmed.ncbi.nlm.nih.gov/29205066/)]
  37. Clark RA, Conway A, Poulsen V, Keech W, Tirimacco R, Tideman P. Alternative models of cardiac rehabilitation: a systematic review. *Eur J Prev Cardiol* 2015 Jan;22(1):35-74. [doi: [10.1177/2047487313501093](https://doi.org/10.1177/2047487313501093)] [Medline: [23943649](https://pubmed.ncbi.nlm.nih.gov/23943649/)]
  38. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital Health Interventions for Adults With Type 2 Diabetes: Qualitative Study of Patient Perspectives on Diabetes Self-Management Education and Support. *J Med Internet Res* 2018 Jan 29;20(2):e40 [FREE Full text] [doi: [10.2196/jmir.8439](https://doi.org/10.2196/jmir.8439)] [Medline: [29463488](https://pubmed.ncbi.nlm.nih.gov/29463488/)]

39. Lahham A, McDonald CF, Mahal A, Lee AL, Hill CJ, Burge AT, et al. Home-based pulmonary rehabilitation for people with COPD: A qualitative study reporting the patient perspective. *Chron Respir Dis* 2018 May;15(2):123-130 [FREE Full text] [doi: [10.1177/1479972317729050](https://doi.org/10.1177/1479972317729050)] [Medline: [28868892](https://pubmed.ncbi.nlm.nih.gov/28868892/)]
40. Holland AE, Mahal A, Hill CJ, Lee AL, Burge AT, Cox NS, et al. Home-based rehabilitation for COPD using minimal resources: a randomised, controlled equivalence trial. *Thorax* 2017 Dec;72(1):57-65. [doi: [10.1136/thoraxjnl-2016-208514](https://doi.org/10.1136/thoraxjnl-2016-208514)] [Medline: [27672116](https://pubmed.ncbi.nlm.nih.gov/27672116/)]
41. Anderson L, Sharp GA, Norton RJ, Dalal H, Dean SG, Jolly K, et al. Home-based versus centre-based cardiac rehabilitation. *Cochrane Database Syst Rev* 2017 Dec 30(6):CD007130. [doi: [10.1002/14651858.CD007130.pub4](https://doi.org/10.1002/14651858.CD007130.pub4)] [Medline: [28665511](https://pubmed.ncbi.nlm.nih.gov/28665511/)]
42. Rawstorn JC, Gant N, Direito A, Beckmann C, Maddison R. Telehealth exercise-based cardiac rehabilitation: a systematic review and meta-analysis. *Heart* 2016 Dec 01;102(15):1183-1192. [doi: [10.1136/heartjnl-2015-308966](https://doi.org/10.1136/heartjnl-2015-308966)] [Medline: [26936337](https://pubmed.ncbi.nlm.nih.gov/26936337/)]
43. NCES. Global digital population as of April 2019 (in millions). Statista: The Statistics Portal. URL: <https://www.statista.com/statistics/617136/digital-population-worldwide/> [accessed 2019-05-01]
44. NCES. Number of smartphone users worldwide from 2014 to 2020 (in billions). Statista: The Statistics Portal. URL: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> [accessed 2019-05-01]
45. NCES. Statista: The Statistics Portal. 2019. URL: <https://www.statista.com/topics/2263/mhealth/> [accessed 2019-05-01]
46. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of the theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *J Med Internet Res* 2010;12(1):e4 [FREE Full text] [doi: [10.2196/jmir.1376](https://doi.org/10.2196/jmir.1376)] [Medline: [20164043](https://pubmed.ncbi.nlm.nih.gov/20164043/)]
47. Maddison R, Rawstorn JC, Shariful Islam SM, Ball K, Tighe S, Gant N, et al. mHealth Interventions for Exercise and Risk Factor Modification in Cardiovascular Disease. *Exerc Sport Sci Rev* 2019 Apr;47(2):86-90 [FREE Full text] [doi: [10.1249/JES.0000000000000185](https://doi.org/10.1249/JES.0000000000000185)] [Medline: [30883472](https://pubmed.ncbi.nlm.nih.gov/30883472/)]
48. Hanlon P, Daines L, Campbell C, McKinsty B, Weller D, Pinnock H. Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer. *J Med Internet Res* 2017 May 17;19(5):e172 [FREE Full text] [doi: [10.2196/jmir.6688](https://doi.org/10.2196/jmir.6688)] [Medline: [28526671](https://pubmed.ncbi.nlm.nih.gov/28526671/)]
49. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text Message and Internet Support for Coronary Heart Disease Self-Management: Results From the Text4Heart Randomized Controlled Trial. *J Med Internet Res* 2015;17(10):e237 [FREE Full text] [doi: [10.2196/jmir.4944](https://doi.org/10.2196/jmir.4944)] [Medline: [26490012](https://pubmed.ncbi.nlm.nih.gov/26490012/)]
50. Dobson R, Whittaker R, Jiang Y, Maddison R, Shepherd M, McNamara C, et al. Effectiveness of text message based, diabetes self management support programme (SMS4BG): two arm, parallel randomised controlled trial. *BMJ* 2018 Dec 17;361:k1959. [doi: [10.1136/bmj.k1959](https://doi.org/10.1136/bmj.k1959)] [Medline: [29773539](https://pubmed.ncbi.nlm.nih.gov/29773539/)]
51. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of Lifestyle-Focused Text Messaging on Risk Factor Modification in Patients With Coronary Heart Disease: A Randomized Clinical Trial. *JAMA* 2015;314(12):1255-1263. [doi: [10.1001/jama.2015.10945](https://doi.org/10.1001/jama.2015.10945)] [Medline: [26393848](https://pubmed.ncbi.nlm.nih.gov/26393848/)]
52. Pfaeffli Dale L, Dobson R, Whittaker R, Maddison R. The effectiveness of mobile-health behaviour change interventions for cardiovascular disease self-management: A systematic review. *Eur J Prev Cardiol* 2016 May;23(8):801-817. [doi: [10.1177/2047487315613462](https://doi.org/10.1177/2047487315613462)] [Medline: [26490093](https://pubmed.ncbi.nlm.nih.gov/26490093/)]
53. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: An overview of systematic reviews. *PLoS One* 2017;12(3):e0173160 [FREE Full text] [doi: [10.1371/journal.pone.0173160](https://doi.org/10.1371/journal.pone.0173160)] [Medline: [28249025](https://pubmed.ncbi.nlm.nih.gov/28249025/)]
54. Maddison R, Rawstorn JC, Stewart RAH, Benatar J, Whittaker R, Rolleston A, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. *Heart* 2018;105(2):122-129. [doi: [10.1136/heartjnl-2018-313189](https://doi.org/10.1136/heartjnl-2018-313189)] [Medline: [30150328](https://pubmed.ncbi.nlm.nih.gov/30150328/)]
55. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. *Heart* 2014 Nov;100(22):1770-1779. [doi: [10.1136/heartjnl-2014-305783](https://doi.org/10.1136/heartjnl-2014-305783)] [Medline: [24973083](https://pubmed.ncbi.nlm.nih.gov/24973083/)]
56. Eysenbach G. The law of attrition. *J Med Internet Res* 2005;7(1):e11 [FREE Full text] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](https://pubmed.ncbi.nlm.nih.gov/15829473/)]
57. World Health Organisation. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. In: <https://apps.who.int/iris/bitstream/handle/10665/252183/9789241511766-eng.pdf>. Geneva, Switzerland: WHO Press; 2016:1-144.
58. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. *J Med Internet Res* 2013 Feb;15(2):e35 [FREE Full text] [doi: [10.2196/jmir.2439](https://doi.org/10.2196/jmir.2439)] [Medline: [23410671](https://pubmed.ncbi.nlm.nih.gov/23410671/)]
59. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. *J Med Internet Res* 2013;15(2):e37 [FREE Full text] [doi: [10.2196/jmir.2281](https://doi.org/10.2196/jmir.2281)] [Medline: [23425685](https://pubmed.ncbi.nlm.nih.gov/23425685/)]
60. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995;123(3):A12-A13. [Medline: [7582737](https://pubmed.ncbi.nlm.nih.gov/7582737/)]



61. Davies KS. Formulating the Evidence Based Practice Question: A Review of the Frameworks. *Evidence Based Library and Information Practice* 2011 Jun 24;6(2):75-80 [[FREE Full text](#)] [doi: [10.18438/B8WS5N](#)]
62. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009 Aug 18;151(4):264-269. [Medline: [19622511](#)]
63. Hong Q, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *EFI* 2018 Dec 18;34(4):285-291 [[FREE Full text](#)] [doi: [10.3233/EFI-180221](#)]
64. Eysenbach G, CONSORT EHEALTH GROUP. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *J Med Internet Res* 2011;13(4):e126 [[FREE Full text](#)] [doi: [10.2196/jmir.1923](#)] [Medline: [22209829](#)]
65. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *Int J Nurs Stud* 2013 May;50(5):587-592. [doi: [10.1016/j.ijnurstu.2012.09.010](#)] [Medline: [23159157](#)]
66. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. *J Med Internet Res* 2017 Jun 29;19(6):e232 [[FREE Full text](#)] [doi: [10.2196/jmir.7126](#)] [Medline: [28663162](#)]
67. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. *J Health Commun* 2012;17(1):11-21. [doi: [10.1080/10810730.2011.649103](#)] [Medline: [22548594](#)]
68. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. *J Med Internet Res* 2017 Dec 06;19(12):e402 [[FREE Full text](#)] [doi: [10.2196/jmir.8578](#)] [Medline: [29212630](#)]
69. Antypas K, Wangberg SC. E-Rehabilitation - an Internet and mobile phone based tailored intervention to enhance self-management of cardiovascular disease: study protocol for a randomized controlled trial. *BMC Cardiovasc Disord* 2012 Jul 09;12(50):1-7 [[FREE Full text](#)] [doi: [10.1186/1471-2261-12-50](#)] [Medline: [22776554](#)]
70. Antypas K, Wangberg SC. Combining users' needs with health behavior models in designing an internet- and mobile-based intervention for physical activity in cardiac rehabilitation. *JMIR Res Protoc* 2014;3(1):e4 [[FREE Full text](#)] [doi: [10.2196/resprot.2725](#)] [Medline: [24413185](#)]
71. Antypas K, Wangberg SC. An Internet- and mobile-based tailored intervention to enhance maintenance of physical activity after cardiac rehabilitation: short-term results of a randomized controlled trial. *J Med Internet Res* 2014 Mar 11;16(3):e77 [[FREE Full text](#)] [doi: [10.2196/jmir.3132](#)] [Medline: [24618349](#)]
72. Poppe L, Van der Mispel C, De Bourdeaudhuij I, Verloigne M, Shadid S, Crombez G. Users' thoughts and opinions about a self-regulation-based eHealth intervention targeting physical activity and the intake of fruit and vegetables: A qualitative study. *PLoS One* 2017;12(12):e0190020 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0190020](#)] [Medline: [29267396](#)]
73. Poppe L, Crombez G, De Bourdeaudhuij I, Van der Mispel C, Shadid S, Verloigne M. Experiences and Opinions of Adults with Type 2 Diabetes Regarding a Self-Regulation-Based eHealth Intervention Targeting Physical Activity and Sedentary Behaviour. *Int J Environ Res Public Health* 2018 May 10;15(5):954 [[FREE Full text](#)] [doi: [10.3390/ijerph15050954](#)] [Medline: [29748460](#)]
74. Voncken-Brewster V, Tange H, de Vries H, Nagykalai Z, Winkens B, van der Weijden T. A randomised controlled trial testing a web-based, computer-tailored self-management intervention for people with or at risk for chronic obstructive pulmonary disease: a study protocol. *BMC Public Health* 2013;13:557 [[FREE Full text](#)] [doi: [10.1186/1471-2458-13-557](#)] [Medline: [23742208](#)]
75. Voncken-Brewster V, Moser A, van der Weijden T, Nagykalai Z, de Vries H, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. *JMIR Res Protoc* 2013 Jan;2(1):e3 [[FREE Full text](#)] [doi: [10.2196/resprot.2246](#)] [Medline: [23612363](#)]
76. Voncken-Brewster V, Tange H, Moser A, Nagykalai Z, de Vries H, van der Weijden T. Integrating a tailored e-health self-management application for chronic obstructive pulmonary disease patients into primary care: a pilot study. *BMC Fam Pract* 2014;15:4 [[FREE Full text](#)] [doi: [10.1186/1471-2296-15-4](#)] [Medline: [24400676](#)]
77. Voncken-Brewster V, Tange H, de Vries H, Nagykalai Z, Winkens B, van der Weijden T. A randomized controlled trial evaluating the effectiveness of a web-based, computer-tailored self-management intervention for people with or at risk for COPD. *Int J Chron Obstruct Pulmon Dis* 2015;10:1061-1073 [[FREE Full text](#)] [doi: [10.2147/COPD.S81295](#)] [Medline: [26089656](#)]
78. Voncken-Brewster V, Amoureux M, de Vries H, Nagykalai Z, Winkens B, van der Weijden T, et al. The Impact of Participant Characteristics on Use and Satisfaction of a Web-Based Computer-Tailored Chronic Obstructive Pulmonary Disease Self-Management Intervention: A Process Evaluation. *JMIR Form Res* 2017 Aug 21;1(1):e1 [[FREE Full text](#)] [doi: [10.2196/formative.6585](#)] [Medline: [30684399](#)]
79. Claes J, Buys R, Woods C, Briggs A, Geue C, Aitken M, et al. PATHway I: design and rationale for the investigation of the feasibility, clinical effectiveness and cost-effectiveness of a technology-enabled cardiac rehabilitation platform. *BMJ Open* 2017 Jun 30;7(6):e016781. [doi: [10.1136/bmjopen-2017-016781](#)] [Medline: [28667228](#)]



80. Walsh DM, Moran K, Cornelissen V, Buys R, Cornelis N, Woods C. Electronic Health Physical Activity Behavior Change Intervention to Self-Manage Cardiovascular Disease: Qualitative Exploration of Patient and Health Professional Requirements. *J Med Internet Res* 2018 May 08;20(5):e163 [FREE Full text] [doi: [10.2196/jmir.9181](https://doi.org/10.2196/jmir.9181)] [Medline: [29739740](https://pubmed.ncbi.nlm.nih.gov/29739740/)]
81. Walsh DMJ, Moran K, Cornelissen V, Buys R, Claes J, Zampognaro P, et al. The development and codesign of the PATHway intervention: a theory-driven eHealth platform for the self-management of cardiovascular disease. *Transl Behav Med* 2019 Dec 01;9(1):76-98. [doi: [10.1093/tbm/iby017](https://doi.org/10.1093/tbm/iby017)] [Medline: [29554380](https://pubmed.ncbi.nlm.nih.gov/29554380/)]
82. Claes J, Cornelissen V, McDermott C, Moyna N, Pattyn N, Cornelis N, et al. Feasibility, Acceptability, and Clinical Effectiveness of a Technology-Enabled Cardiac Rehabilitation Platform (Physical Activity Toward Health-I): Randomized Controlled Trial. *J Med Internet Res* 2020 Feb 04;22(2):e14221 [FREE Full text] [doi: [10.2196/14221](https://doi.org/10.2196/14221)] [Medline: [32014842](https://pubmed.ncbi.nlm.nih.gov/32014842/)]
83. Yu CH, Parsons J, Mamdani M, Lebovic G, Shah BR, Bhattacharyya O, et al. Designing and evaluating a web-based self-management site for patients with type 2 diabetes--systematic website development and study protocol. *BMC Med Inform Decis Mak* 2012 Jun 24;12:57 [FREE Full text] [doi: [10.1186/1472-6947-12-57](https://doi.org/10.1186/1472-6947-12-57)] [Medline: [22726578](https://pubmed.ncbi.nlm.nih.gov/22726578/)]
84. Yu CH, Parsons JA, Hall S, Newton D, Jovicic A, Lottridge D, et al. User-centered design of a web-based self-management site for individuals with type 2 diabetes - providing a sense of control and community. *BMC Med Inform Decis Mak* 2014;14:60 [FREE Full text] [doi: [10.1186/1472-6947-14-60](https://doi.org/10.1186/1472-6947-14-60)] [Medline: [25056379](https://pubmed.ncbi.nlm.nih.gov/25056379/)]
85. Yu CH, Parsons JA, Mamdani M, Lebovic G, Hall S, Newton D, et al. A web-based intervention to support self-management of patients with type 2 diabetes mellitus: effect on self-efficacy, self-care and diabetes distress. *BMC Med Inform Decis Mak* 2014 Dec 14;14:117 [FREE Full text] [doi: [10.1186/s12911-014-0117-3](https://doi.org/10.1186/s12911-014-0117-3)] [Medline: [25495847](https://pubmed.ncbi.nlm.nih.gov/25495847/)]
86. Hofmann M, Dack C, Barker C, Murray E. The Impact of an Internet-Based Self-Management Intervention (HeLP-Diabetes) on the Psychological Well-Being of Adults with Type 2 Diabetes: A Mixed-Method Cohort Study. *J Diabetes Res* 2016;2016:1476384 [FREE Full text] [doi: [10.1155/2016/1476384](https://doi.org/10.1155/2016/1476384)] [Medline: [26682226](https://pubmed.ncbi.nlm.nih.gov/26682226/)]
87. Alkhaldi G, Modrow K, Hamilton F, Pal K, Ross J, Murray E. Promoting Engagement With a Digital Health Intervention (HeLP-Diabetes) Using Email and Text Message Prompts: Mixed-Methods Study. *Interact J Med Res* 2017 Aug 22;6(2):e14 [FREE Full text] [doi: [10.2196/ijmr.6952](https://doi.org/10.2196/ijmr.6952)] [Medline: [28829328](https://pubmed.ncbi.nlm.nih.gov/28829328/)]
88. Murray E, Sweeting M, Dack C, Pal K, Modrow K, Hudda M, et al. Web-based self-management support for people with type 2 diabetes (HeLP-Diabetes): randomised controlled trial in English primary care. *BMJ Open* 2017 Sep 27;7(9):e016009. [doi: [10.1136/bmjopen-2017-016009](https://doi.org/10.1136/bmjopen-2017-016009)] [Medline: [28954789](https://pubmed.ncbi.nlm.nih.gov/28954789/)]
89. Li J, Parrott S, Sweeting M, Farmer A, Ross J, Dack C, et al. Cost-Effectiveness of Facilitated Access to a Self-Management Website, Compared to Usual Care, for Patients With Type 2 Diabetes (HeLP-Diabetes): Randomized Controlled Trial. *J Med Internet Res* 2018 Jun 08;20(6):e201 [FREE Full text] [doi: [10.2196/jmir.9256](https://doi.org/10.2196/jmir.9256)] [Medline: [29884608](https://pubmed.ncbi.nlm.nih.gov/29884608/)]
90. Poduval S, Ahmed S, Marston L, Hamilton F, Murray E. Crossing the Digital Divide in Online Self-Management Support: Analysis of Usage Data From HeLP-Diabetes. *JMIR Diabetes* 2018 Dec 06;3(4):e10925 [FREE Full text] [doi: [10.2196/10925](https://doi.org/10.2196/10925)] [Medline: [30522988](https://pubmed.ncbi.nlm.nih.gov/30522988/)]
91. Dack C, Ross J, Stevenson F, Pal K, Gubert E, Michie S, et al. A digital self-management intervention for adults with type 2 diabetes: Combining theory, data and participatory design to develop HeLP-Diabetes. *Internet Interv* 2019 Sep;17:100241 [FREE Full text] [doi: [10.1016/j.invent.2019.100241](https://doi.org/10.1016/j.invent.2019.100241)] [Medline: [31372349](https://pubmed.ncbi.nlm.nih.gov/31372349/)]
92. Poppe L, De Bourdeaudhuij I, Verloigne M, Degroote L, Shadid S, Crombez G. A Self-Regulation-Based eHealth and mHealth Intervention for an Active Lifestyle in Adults With Type 2 Diabetes: Protocol for a Randomized Controlled Trial. *JMIR Res Protoc* 2019 Mar 22;8(3):e12413 [FREE Full text] [doi: [10.2196/12413](https://doi.org/10.2196/12413)] [Medline: [30901002](https://pubmed.ncbi.nlm.nih.gov/30901002/)]
93. Poppe L, De Bourdeaudhuij I, Verloigne M, Shadid S, Van Cauwenberg J, Compennolle S, et al. Efficacy of a Self-Regulation-Based Electronic and Mobile Health Intervention Targeting an Active Lifestyle in Adults Having Type 2 Diabetes and in Adults Aged 50 Years or Older: Two Randomized Controlled Trials. *J Med Internet Res* 2019 Aug 02;21(8):e13363 [FREE Full text] [doi: [10.2196/13363](https://doi.org/10.2196/13363)] [Medline: [31376274](https://pubmed.ncbi.nlm.nih.gov/31376274/)]
94. Sakakibara BM, Ross E, Arthur G, Brown-Ganzert L, Petrin S, Sedlak T, et al. Using Mobile-Health to Connect Women with Cardiovascular Disease and Improve Self-Management. *Telemed J E Health* 2017 Mar;23(3):233-239. [doi: [10.1089/tmj.2016.0133](https://doi.org/10.1089/tmj.2016.0133)] [Medline: [27623231](https://pubmed.ncbi.nlm.nih.gov/27623231/)]
95. Sakakibara BM, Chakrabarti S, Krahn A, Mackay MH, Sedlak T, Singer J, et al. Delivery of Peer Support Through a Self-Management mHealth Intervention (Healing Circles) in Patients With Cardiovascular Disease: Protocol for a Randomized Controlled Trial. *JMIR Res Protoc* 2019 Jan 11;8(1):e12322 [FREE Full text] [doi: [10.2196/12322](https://doi.org/10.2196/12322)] [Medline: [30635261](https://pubmed.ncbi.nlm.nih.gov/30635261/)]
96. Murray E, Dack C, Barnard M, Farmer A, Li J, Michie S, et al. HeLP-Diabetes: randomised controlled trial protocol. *BMC Health Serv Res* 2015 Dec 29;15:578 [FREE Full text] [doi: [10.1186/s12913-015-1246-9](https://doi.org/10.1186/s12913-015-1246-9)] [Medline: [26715038](https://pubmed.ncbi.nlm.nih.gov/26715038/)]
97. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013 Aug;46(1):81-95. [doi: [10.1007/s12160-013-9486-6](https://doi.org/10.1007/s12160-013-9486-6)] [Medline: [23512568](https://pubmed.ncbi.nlm.nih.gov/23512568/)]
98. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. *Diabetes Care* 2000 Jul;23(7):943-950. [doi: [10.2337/diacare.23.7.943](https://doi.org/10.2337/diacare.23.7.943)] [Medline: [10895844](https://pubmed.ncbi.nlm.nih.gov/10895844/)]
99. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. *Patient Educ Couns* 2007 May;66(2):192-201. [doi: [10.1016/j.pec.2006.12.002](https://doi.org/10.1016/j.pec.2006.12.002)] [Medline: [17320338](https://pubmed.ncbi.nlm.nih.gov/17320338/)]

100. van der Molen T, Willemse BWM, Schokker S, ten Hacken NH, Postma DS, Juniper EF. Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health Qual Life Outcomes* 2003 Apr 28;1(13):1-10 [[FREE Full text](#)] [doi: [10.1186/1477-7525-1-13](#)] [Medline: [12773199](#)]
101. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998 May 12;97(18):1837-1847. [doi: [10.1161/01.cir.97.18.1837](#)] [Medline: [9603539](#)]
102. Polonsky WH, Fisher L, Earles J, Dudl RJ, Lees J, Mullan J, et al. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. *Diabetes Care* 2005 Mar;28(3):626-631. [doi: [10.2337/diacare.28.3.626](#)] [Medline: [15735199](#)]
103. Simonsen J, Robertson T. Participatory Design: an introduction. In: Simonsen J, Robertson T, editors. *Routledge International Handbook of Participatory Design*. New York: Routledge; 2013:1-18.
104. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. *Soc Theory Health* 2013 Jun 19;11(3):256-270. [doi: [10.1057/sth.2013.10](#)]
105. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. *J Med Internet Res* 2015;17(1):e30 [[FREE Full text](#)] [doi: [10.2196/jmir.4055](#)] [Medline: [25639757](#)]
106. Bossen C, Dindler C, Garde J, Pipek V. Evaluation, sustainability and long-term effects of participatory design projects. In: *Proceedings of the 13th Participatory Design Conference: Short Papers, Industry Cases, Workshop Descriptions, Doctoral Consortium papers, and Keynote abstracts - Volume 2*. 2014 Presented at: Participatory Design Conference; October 2014; Windhoek Namibia p. 219-220 URL: <https://doi.org/10.1145/2662155.2662210> [doi: [10.1145/2662155.2662210](#)]
107. Andersen TO, Bansler JP, Kensing F, Moll J. From Prototype to Product: Making Participatory Design of mHealth Commercially Viable. *Stud Health Technol Inform* 2017;233:95-112. [doi: [10.3233/978-1-61499-740-5-95](#)] [Medline: [28125417](#)]
108. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of User Engagement with Mobile- and Web-Delivered Self-Care Interventions for Adults with T2DM: A Review of the Literature. *Curr Diab Rep* 2016 Dec;16(7):66 [[FREE Full text](#)] [doi: [10.1007/s11892-016-0755-1](#)] [Medline: [27255269](#)]
109. Chow CK, Ariyaratna N, Islam SMS, Thiagalingam A, Redfern J. mHealth in Cardiovascular Health Care. *Heart Lung Circ* 2016 Aug;25(8):802-807. [doi: [10.1016/j.hlc.2016.04.009](#)] [Medline: [27262389](#)]
110. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. *BMC Med Inform Decis Mak* 2016 Sep 15;16(1):120 [[FREE Full text](#)] [doi: [10.1186/s12911-016-0359-3](#)] [Medline: [27630020](#)]
111. Danaher BG, McKay HG, Seeley JR. The information architecture of behavior change websites. *J Med Internet Res* 2005;7(2):e12 [[FREE Full text](#)] [doi: [10.2196/jmir.7.2.e12](#)] [Medline: [15914459](#)]
112. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. *Ann Behav Med* 2009 Aug;38(1):18-27 [[FREE Full text](#)] [doi: [10.1007/s12160-009-9133-4](#)] [Medline: [19802647](#)]
113. Michie S, Johnston M. Theories and techniques of behaviour change: Developing a cumulative science of behaviour change. *Health Psychology Review* 2012 Mar;6(1):1-6 [[FREE Full text](#)] [doi: [10.1080/17437199.2012.654964](#)] [Medline: [26616119](#)]
114. Prochaska JJ, Prochaska JO. A Review of Multiple Health Behavior Change Interventions for Primary Prevention. *Am J Lifestyle Med* 2011 May;5(3):208-221 [[FREE Full text](#)] [doi: [10.1177/1559827610391883](#)] [Medline: [24358034](#)]
115. Drca N, Wolk A, Jensen-Ustad M, Larsson SC. Physical activity is associated with a reduced risk of atrial fibrillation in middle-aged and elderly women. *Heart* 2015 Oct;101(20):1627-1630. [doi: [10.1136/heartjnl-2014-307145](#)] [Medline: [26019224](#)]
116. Trost SG, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults' participation in physical activity: review and update. *Med Sci Sports Exerc* 2002 Dec;34(12):1996-2001. [doi: [10.1097/00005768-200212000-00020](#)] [Medline: [12471307](#)]
117. Brownson RC, Eyster AA, King AC, Brown DR, Shyu YL, Sallis JF. Patterns and correlates of physical activity among US women 40 years and older. *Am J Public Health* 2000 Feb;90(2):264-270. [doi: [10.2105/ajph.90.2.264](#)] [Medline: [10667189](#)]
118. Buczkowski K, Marcinowicz L, Czachowski S, Piszczek E. Motivations toward smoking cessation, reasons for relapse, and modes of quitting: results from a qualitative study among former and current smokers. *Patient Prefer Adherence* 2014;8:1353-1363 [[FREE Full text](#)] [doi: [10.2147/PPA.S67767](#)] [Medline: [25336926](#)]
119. Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. *Sports Med* 2014 Jan;44(1):81-121 [[FREE Full text](#)] [doi: [10.1007/s40279-013-0090-5](#)] [Medline: [24030837](#)]
120. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJF, Martin BW, Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not? *Lancet* 2012 Jul 21;380(9838):258-271. [doi: [10.1016/S0140-6736\(12\)60735-1](#)] [Medline: [22818938](#)]
121. Torres SJ, Nowson CA. Relationship between stress, eating behavior, and obesity. *Nutrition* 2007;23(11-12):887-894. [doi: [10.1016/j.nut.2007.08.008](#)] [Medline: [17869482](#)]
122. Murray SA, Kendall M, Boyd K, Sheikh A. Illness trajectories and palliative care. *BMJ* 2005 Apr 30;330(7498):1007-1011 [[FREE Full text](#)] [doi: [10.1136/bmj.330.7498.1007](#)] [Medline: [15860828](#)]

123. Phanareth K, Vingtoft S, Christensen AS, Nielsen JS, Svenstrup J, Berntsen GKR, et al. The Epital Care Model: A New Person-Centered Model of Technology-Enabled Integrated Care for People With Long Term Conditions. *JMIR Res Protoc* 2017 Jan 16;6(1):e6 [FREE Full text] [doi: [10.2196/resprot.6506](https://doi.org/10.2196/resprot.6506)] [Medline: [28093379](https://pubmed.ncbi.nlm.nih.gov/28093379/)]
124. Mol A, Moser I, Pols J. Care in Practice On Tinkering in Clinics, Homes and Farms. Verlag, Bielefeld: Transcript Verlag; 2010:1-326.
125. May CR, Eton DT, Boehmer K, Gallacher K, Hunt K, MacDonald S, et al. Rethinking the patient: using Burden of Treatment Theory to understand the changing dynamics of illness. *BMC Health Serv Res* 2014;14:281 [FREE Full text] [doi: [10.1186/1472-6963-14-281](https://doi.org/10.1186/1472-6963-14-281)] [Medline: [24969758](https://pubmed.ncbi.nlm.nih.gov/24969758/)]
126. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. *Am J Health Promot* 1997;12(1):38-48. [doi: [10.4278/0890-1171-12.1.38](https://doi.org/10.4278/0890-1171-12.1.38)] [Medline: [10170434](https://pubmed.ncbi.nlm.nih.gov/10170434/)]
127. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. *Am J Prev Med* 2016 Nov;51(5):833-842. [doi: [10.1016/j.amepre.2016.06.015](https://doi.org/10.1016/j.amepre.2016.06.015)] [Medline: [27745683](https://pubmed.ncbi.nlm.nih.gov/27745683/)]
128. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. *J Behav Med* 2017 Feb;40(1):85-98 [FREE Full text] [doi: [10.1007/s10865-016-9818-7](https://doi.org/10.1007/s10865-016-9818-7)] [Medline: [28058516](https://pubmed.ncbi.nlm.nih.gov/28058516/)]
129. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. *J Gen Intern Med* 2012 May;27(5):520-526 [FREE Full text] [doi: [10.1007/s11606-011-1931-2](https://doi.org/10.1007/s11606-011-1931-2)] [Medline: [22127797](https://pubmed.ncbi.nlm.nih.gov/22127797/)]
130. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. *Ann Behav Med* 2003 Aug;26(1):1-7. [doi: [10.1207/S15324796ABM2601\\_01](https://doi.org/10.1207/S15324796ABM2601_01)] [Medline: [12867348](https://pubmed.ncbi.nlm.nih.gov/12867348/)]
131. Triantafyllidis A, Filos D, Buys R, Claes J, Cornelissen V, Kouidi E, et al. Computerized decision support for beneficial home-based exercise rehabilitation in patients with cardiovascular disease. *Comput Methods Programs Biomed* 2018 Aug;162:1-10. [doi: [10.1016/j.cmpb.2018.04.030](https://doi.org/10.1016/j.cmpb.2018.04.030)] [Medline: [29903475](https://pubmed.ncbi.nlm.nih.gov/29903475/)]
132. Schwarzer R. Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. *Applied Psychology* 2008 Jan;57(1):1-29 [FREE Full text] [doi: [10.1111/j.1464-0597.2007.00325.x](https://doi.org/10.1111/j.1464-0597.2007.00325.x)]
133. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a Conceptual Model for eHealth: Qualitative Study with Key Informants. *J Med Internet Res* 2017 Oct 24;19(10):e324 [FREE Full text] [doi: [10.2196/jmir.8106](https://doi.org/10.2196/jmir.8106)] [Medline: [29066429](https://pubmed.ncbi.nlm.nih.gov/29066429/)]
134. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008 Sep 29;337:a1655 [FREE Full text] [doi: [10.1136/bmj.a1655](https://doi.org/10.1136/bmj.a1655)] [Medline: [18824488](https://pubmed.ncbi.nlm.nih.gov/18824488/)]

## Abbreviations

**CONSORT-EHEALTH:** Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

**COPD:** chronic obstructive pulmonary disease

**CVD:** cardiovascular disease

**HAPA:** Health Action Process Approach

**HbA<sub>1c</sub>:** glycated hemoglobin

**MET:** metabolic equivalents of task-minutes

**MMAT:** Mixed Methods Appraisal Tool

**MVPA:** moderate-to-vigorous physical activity

**NCD:** noncommunicable disease

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**SCT:** social cognitive theory

**TTM:** transtheoretical model

*Edited by G Eysenbach; submitted 23.10.19; peer-reviewed by J Wienert, N Anegondi, D Leightley; comments to author 30.12.19; revised version received 25.05.20; accepted 14.09.20; published 28.10.20.*

*Please cite as:*

Tighe SA, Ball K, Kensing F, Kayser L, Rawstorn JC, Maddison R  
 Toward a Digital Platform for the Self-Management of Noncommunicable Disease: Systematic Review of Platform-Like Interventions  
*J Med Internet Res* 2020;22(10):e16774  
 URL: <http://www.jmir.org/2020/10/e16774/>  
 doi:[10.2196/16774](https://doi.org/10.2196/16774)  
 PMID:[33112239](https://pubmed.ncbi.nlm.nih.gov/33112239/)

©Sarah A Tighe, Kylie Ball, Finn Kensing, Lars Kayser, Jonathan C Rawstorn, Ralph Maddison. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Review

# Acceptability and Effectiveness of NHS-Recommended e-Therapies for Depression, Anxiety, and Stress: Meta-Analysis

Melanie Simmonds-Buckley<sup>1\*</sup>, BSc, PhD; Matthew Russell Bennion<sup>1,2\*</sup>, BEng, MSc, PhD; Stephen Kellett<sup>1,3</sup>, BSc, MSc, ClinPsy; Abigail Millings<sup>1,4</sup>, BSc, PhD; Gillian E Hardy<sup>1</sup>, BA, MSc, PhD; Roger K Moore<sup>2</sup>, BA, MSc, PhD

<sup>1</sup>Department of Psychology, University of Sheffield, Sheffield, United Kingdom

<sup>2</sup>Department of Computer Science, The University of Sheffield, Sheffield, United Kingdom

<sup>3</sup>Sheffield Health and Social Care NHS Foundation Trust, Sheffield, United Kingdom

<sup>4</sup>Centre for Behavioural Science and Applied Psychology, Sheffield Hallam University, Sheffield, United Kingdom

\*these authors contributed equally

**Corresponding Author:**

Melanie Simmonds-Buckley, BSc, PhD

Department of Psychology

University of Sheffield

Cathedral Court, 1 Vicar Lane

S1 2LT

Sheffield, S1 2LT

United Kingdom

Phone: 44 01142226630

Email: [m.simmonds-buckley@sheffield.ac.uk](mailto:m.simmonds-buckley@sheffield.ac.uk)

## Abstract

**Background:** There is a disconnect between the ability to swiftly develop e-therapies for the treatment of depression, anxiety, and stress, and the scrupulous evaluation of their clinical utility. This creates a risk that the e-therapies routinely provided within publicly funded psychological health care have evaded appropriate rigorous evaluation in their development.

**Objective:** This study aims to conduct a meta-analytic review of the gold standard evidence of the acceptability and clinical effectiveness of e-therapies recommended for use in the National Health Service (NHS) in the United Kingdom.

**Methods:** Systematic searches identified appropriate randomized controlled trials (RCTs). Depression, anxiety, and stress outcomes at the end of treatment and follow-up were synthesized using a random-effects meta-analysis. The grading of recommendations assessment, development, and evaluation approach was used to assess the quality of each meta-analytic comparison. Moderators of treatment effect were examined using subgroup and meta-regression analysis. Dropout rates for e-therapies (as a proxy for acceptability) were compared against controls.

**Results:** A total of 24 studies evaluating 7 of 48 NHS-recommended e-therapies were qualitatively and quantitatively synthesized. Depression, anxiety, and stress outcomes for e-therapies were superior to controls (depression: standardized mean difference [SMD] 0.38, 95% CI 0.24 to 0.52, N=7075; anxiety and stress: SMD 0.43, 95% CI 0.24 to 0.63, n=4863), and these small effects were maintained at follow-up. Average dropout rates for e-therapies (31%, SD 17.35) were significantly higher than those of controls (17%, SD 13.31). Limited moderators of the treatment effect were found.

**Conclusions:** Many NHS-recommended e-therapies have not been through an RCT-style evaluation. The e-therapies that have been appropriately evaluated generate small but significant, durable, beneficial treatment effects.

**Trial Registration:** International Prospective Register of Systematic Reviews (PROSPERO) registration CRD42019130184; [https://www.crd.york.ac.uk/prospero/display\\_record.php?RecordID=130184](https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=130184)

(*J Med Internet Res* 2020;22(10):e17049) doi:[10.2196/17049](https://doi.org/10.2196/17049)

**KEYWORDS**

e-therapy; anxiety; depression; treatment effectiveness; National Health Service; meta-analysis; mobile phone



## Introduction

The potential contribution of digital technology in enabling access to evidenced-based psychological care for mental health problems is high on national and international research, policy, commissioning, and service management agendas [1]. In modern life, as digital tools (eg, mobile phones, tablets, laptops, and wearable devices) have become ubiquitous, psychological interventions delivered by such devices (ie, e-therapies) offer greater convenience and enable constant access to treatment compared with traditional face-to-face therapy with health professionals [2]. The increasing demand for primary care psychological services globally has provided the context within which e-therapies have been integrated into the offer of a suite of *low-intensity (LI)* psychological interventions [3], often delivered within stepped-care systems [4,5]. Although technological innovation in methods of treatment delivery usefully expands availability, it also creates the risk of commercial promotion and availability of ineffective or possibly harmful psychological interventions [6]. Therefore, commissioners, clinicians, and patients need access to reliable and contemporary guidance regarding the empirical status and clinical utility of e-therapies.

The potential organizational, therapeutic, and health economic benefits of e-therapies initially triggered a global wave of investment and interest [7]. In the United Kingdom, for example, the National Health Service (NHS) Commissioning Board launched the NHS Health Apps Library in March 2013 and NHS Mental Health Apps Library in March 2015. However, the libraries were removed in 2015 after questions were raised concerning e-therapy data security governance [8] and clinical effectiveness [9]. NHS England launched 2 new digital platforms in April 2017, a new beta of the NHS Digital Apps Library and a mobile health space, in an effort to close the gap between e-therapy development and thorough evaluation. Before the removal of the initial NHS App Libraries, a list of 48 NHS-recommended e-therapies was compiled for the National Institute for Health and Care Excellence (NICE) assessment of digitally enabled psychological therapies for use in Improving Access to Psychological Therapies (IAPT) services [10]. A recent quality assessment of the development process of NHS-recommended e-therapies strongly advocated developers to routinely adopt clinical trial methods to test acceptability and efficacy of e-therapies before wider dissemination [11]. NICE has also recently published an evidence standards framework for e-therapies providing guidance concerning efficacy and effectiveness standards [12].

This review aims to quantitatively synthesize the evidence base of e-therapies recommended for use in the NHS for depression, anxiety, and stress in adults to better inform the commissioning and use of e-therapies in clinical services. It was relevant to restrict this review to adults as the NHS-recommended e-therapies are intended for adults. Previously, an individual participant meta-analysis of the e-therapy clinical trial evidence base for depression showed that e-therapy was significantly more effective than controls [13], and there is clinical trial evidence for the efficacy of e-therapy as a treatment for anxiety [14]. This study had 3 aims. First, we sought to quantify the

effect of NHS-recommended e-therapies (ie, the 48 e-therapies identified by Bennion et al [10]), as no previous specific meta-analysis of the efficacy of NHS-recommended e-therapies has been attempted. As randomized controlled trials (RCTs) are viewed as the *gold standard evaluation* [15], we sought to only use RCT studies to increase the quality of the meta-analysis. Second because e-therapies are criticized for generating high dropout rates [16], we sought to compare dropout rates in contrast to controls to appraise acceptability. Finally, we sought to investigate the impact of potential moderating factors (eg, gender, age, severity, treatment approach, treatment duration, setting, focus problem, and risk of bias) on e-therapy outcomes via subgroup and meta-regression analyses.

## Methods

The review was registered on the International Prospective Register of Systematic Reviews (PROSPERO; CRD42019130184). The PRISMA (Preferred Reporting Guidelines for Systematic Reviews and Meta-Analyses) are used throughout [17].

### Study Selection

A 3-stage search strategy was developed to identify RCTs evaluating all of the e-therapies recommended by the NHS for the treatment of depression, anxiety, and stress. First, each of the 48 NHS-recommended e-therapies identified by Bennion et al [10] was used to determine those e-therapies to be included in the search strategy. The name of each e-therapy and its platform type (website or app) were combined to develop a series of search terms (eg, “Beating the Blues” AND “Website”) [18]. Electronic searches were conducted using PsycINFO, Web of Science, and PubMed databases to identify relevant e-therapy outcome studies published up until April 2019 (date of final search was April 11, 2019; see [Multimedia Appendix 1](#) for an example search strategy). Second, reference lists of identified studies and previous e-therapy reviews were also searched. Third, as many e-therapies are not developed under their commercial name, a survey was disseminated to the 48 app developers of the identified NHS-recommended e-therapies to identify additional gray literature not captured by the terms used in the database searches [11]. This process was to supplement the identification of all studies associated with any one e-therapy, even when the commercial name was not used in the reporting. A total of 36 out of 48 (75%) app developers responded to the survey, and the full process was reported by Bennion et al [11]. Titles and abstracts were screened initially (MB), with the full texts of identified studies then screened against inclusion and exclusion eligibility criteria (MB). Queries regarding study eligibility were resolved through discussion among reviewers (MB, SK, and AM).

### Eligibility Criteria

Studies were included if the web-based or smartphone app intervention used was one of the 48 NHS-recommended e-therapies [10] for depression, anxiety, and stress; therefore, all studies of other types of e-therapies and for other clinical conditions were excluded. Studies were eligible for inclusion if, and only if, they used an RCT design to examine the efficacy

of e-therapy with an adult population (ie, aged >18 years). To be included, the developer of the e-therapy had to be locatable via a Google search when entering the app name as the search term, and the app had to reference the targeted condition (ie, depression, anxiety, or stress) in its marketing literature or be based on a therapeutic tool known to benefit the targeted condition. Posttreatment outcomes were required to have been assessed using a validated measure of anxiety and/or depression symptoms. Comparators included any *control condition*, comprising a wait list or no treatment, placebo or attention-control activity, or treatment as usual (TAU). Only English language articles were included.

## Outcomes

The 2 main outcomes of interest were participant-reported outcomes of (1) depression and/or (2) anxiety and stress taken at posttreatment and at follow-up (where available, to assess the durability of e-therapy effectiveness). Where multiple measures of one outcome were used (ie, 2 measures of depression), the most frequently used measure across the included studies was prioritized. Therefore, each study only contributed one effect size per outcome. Dropout (as a proxy for acceptability) was classified as the percentage of e-therapy and comparator condition noncompleters, as determined by the definition applied in the original study.

## Data Extraction

A priority data extraction tool was designed for the purpose of the review. MB extracted data from the original studies and then reviewers (SK and AM) independently verified the findings. Data were coded according to the following criteria: (1) *study information*—sample size, trial design, context, comparator type, study length, analytic approach (intention to treat [ITT] or completers), and trial quality; (2) *participant characteristics*—mean age, percentage of males, population sample, presenting problem, and diagnostic information or relevant inclusion criteria; (3) *outcome characteristics*—outcome measure and, if applicable, length of follow-up; and (4) *intervention features*—e-therapy program, regularity of instructed use, duration, intervention component details of the comparator condition, and self-help typology. The self-help typology for each e-therapy was coded based on the framework by Newman et al [19]: minimal contact therapy, predominantly self-help, predominantly therapist-administered treatment, or self-administered therapy. This was selected to provide an assessment of the level and extent of therapist support within the e-therapies. Outcome data on depression, anxiety, and stress symptoms and dropout rates were extracted at treatment completion and follow-up (ie, at 6 months or the closest assessment point available).

## Study and Evidence Quality

The Cochrane risk of bias tool [20] was used to assess the methodological quality of the original studies using the Cochrane Review Manager (RevMan) program [21]. All included studies were assessed on 7 elements: (1) randomization, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of outcome assessment, (5) data attrition, (6) selective outcome reporting, and (7) other threats to validity.

Elements were rated as having low risk, unclear, or high risk of bias. One rater assessed all the included studies, with all studies double rated by 2 other raters (rater 1 assessed 63% [15/24] and rater 2 assessed 37% [9/24]). Cohen kappa coefficient ( $k$ ) was used to assess the interrater agreement on risk of bias overall scores between the primary rater and 2 second raters [22], and these were interpreted using the Landis and Koch [23] categories: <0 as indicating no agreement, 0 to 0.20 as slight, 0.21 to 0.40 as fair, 0.41 to 0.60 as moderate, 0.61 to 0.80 as substantial, and 0.81 to 1 as almost perfect agreement. There was substantial agreement between the primary rater and rater 1 ( $k=.63$ ) and moderate agreement between the primary rater and rater 2 ( $k=.54$ ). Any differences in rating were discussed by the raters to reach a consensus on the overall risk of bias rating for each included study. The grading of recommendations assessment, development, and evaluation (GRADE) approach was used to rate the quality of the evidence included in each meta-analysis conducted [24]. The quality of evidence was assessed on 5 domains: (1) risk of bias in the individual included studies, (2) publication bias, (3) inconsistency, (4) imprecision, and (5) indirectness of treatment estimate effects. The meta-analysis was graded by 2 reviewers (SK and MS) and a consensus agreed (rated as high, moderate, low, or very low quality).

## Effect Sizes

Standardized mean differences (SMDs) were used to assess differences in outcome between e-therapy and the comparator conditions at posttreatment and follow-up. SMDs were computed by calculating Cohen  $d$  (mean outcome score of the comparator condition subtracted from the mean outcome score of the e-therapy and dividing by the pooled standard deviation). Where available, effect sizes were computed using ITT outcome data. To account for potential biases in studies with small sample sizes, SMDs were converted to Hedges  $g$  using the  $J$  adjustment [25]. Effect sizes were calculated so that a beneficial effect of e-therapy was represented by a positive SMD and vice versa. Interpretations of effect size magnitude were classified as 0.20 to 0.49=small, 0.50 to 0.79 = medium, and >0.80=large [26]. When studies had multiple treatment arms delivering e-therapies that could be considered comparable (ie, the same e-therapy with different component combinations, such as reminders and telephone support), the data were collapsed into a single group using Cochrane guidelines [20]. When studies had multiple treatment arms that could not be collapsed (ie, three-arm trial comparing 2 different types of recommended e-therapy to a control), the treatment arms were included independently. The sample size of the shared comparator condition was split evenly across independent treatment arm comparisons to avoid participant data being included twice.

## Data Synthesis

Meta-Essentials workbooks were used to synthesize e-therapy treatment effects in a random-effects meta-analysis to account for the extent of expected study heterogeneity [27]. Individual study effect sizes were weighted using the inverse of the variance to produce overall pooled treatment effect estimates and 95% CIs. The threshold for statistical significance was set at an  $\alpha$ value of .05. The  $I^2$  statistic was employed as an indicator

of the percentage of between-study heterogeneity, whereas the  $Q$  statistic provided a test of the statistical significance of the presence of study variation. Thresholds of heterogeneity were interpreted as <40% may not be relevant, 30% to 60% representing moderate heterogeneity, 50% to 90% representing substantial heterogeneity, and 75% to 100% representing considerable heterogeneity [28]. As recommended by Cochrane, the magnitude and direction of effect sizes were used to interpret the implications of  $I^2$  percentages. The overall pooled effect sizes of e-therapy were translated into *numbers needed to treat* (NNTs) [29]. NNT is an approximation of how many patients would need treatment with e-therapy to generate an additional outcome of benefit when compared with another intervention (ie, the comparator condition). A Mann-Whitney U test was used to assess for differences in dropout rates between e-therapy and controls.

### Moderator and Sensitivity Analyses

Preplanned random-effects moderator analyses were performed using the Meta-Essentials workbooks to evaluate between-study variation in treatment effects in posttreatment comparisons with a minimum of 10 studies [20]. Moderators were selected based on methodological, clinical, and intervention features that were likely to vary between studies. Meta-regressions were applied to 5 continuous variables: mean age, mean number of sessions completed, percentage of males, baseline symptom severity (standardized Z scores), and risk of bias (number of items meeting criteria for low risk of bias: 0-7). Subgroup analyses were applied to 6 categorical variables: 4 of them were specified a priori (control type, e-therapy type, self-help typology, and recruitment setting) and 2 were conducted post hoc (focus problem and analysis method). Owing to multiple testing, the  $\alpha$  threshold for significance of the meta-regression beta-coefficients and the between-subgroup differences was lowered to  $P<.01$ . A series of sensitivity analyses were

performed to assess the impact of outliers on the pooled effect sizes (with extreme outliers removed) and to further explore treatment effect durability (comparisons of follow-up effects separately at short-term [1-2 months], medium-term [6 months], and long-term [>8 months] follow-up).

### Publication Bias

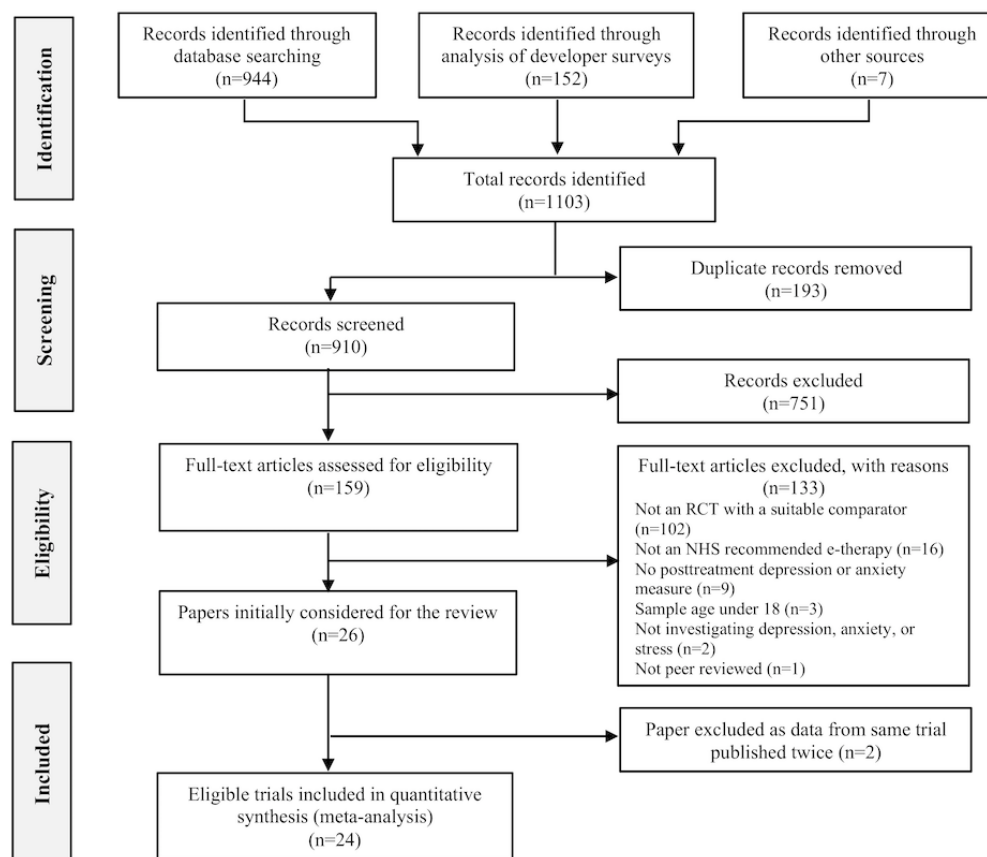
Several methods were employed to assess for the presence of publication bias in the posttreatment comparisons that had a sufficient number of studies ( $k>10$ ). Visual inspection of the asymmetry of a funnel plot (SE plotted against effect sizes) gave an indication of the extent of potential publication bias, whereas the accompanying Trim and Fill imputation [30] accounted for any reporting bias to provide an adjusted treatment estimate. Finally, additional statistical testing of asymmetrical study distribution was undertaken using Egger regression [31].

## Results

### Study Selection

The electronic searches returned a total of 944 records. This was combined with the 152 records collected by surveying app developers and 7 records from a manual reference list and review searches, giving a combined total of 1103 records (Figure 1). Duplicates were removed, leaving a total of 910 records to be screened. After excluding records that did not meet the inclusion criteria based on abstracts, 159 full-text articles were retrieved and assessed. Overall, 26 trials were considered eligible, and 2 were excluded because they contained duplicate data from another trial. Thus, a total of 24 studies that tested the efficacy of 7 of the 48 NHS-recommended e-therapies (Beating the Blues, FearFighter, MoodGYM, IESO, Headspace, Silver Cloud, and Work Guru) in an RCT design were included in the meta-analysis. Details of the included studies can be found in Multimedia Appendix 2 [32-55].

**Figure 1.** PRISMA (Preferred Reporting Guidelines for Systematic Reviews and Meta-Analyses) flowchart of study selection. NHS: National Health Service; RCT: randomized controlled trial.



The risk of bias ratings are presented in Table 1. Of the 24 included studies, quality ranged between 1 and 7 quality items meeting low risk of bias criteria (maximum of 7). The overall study quality was moderate to good, with 13 studies meeting low risk of bias criteria on at least five items. A lack of or

unclear blinding of participants and personnel or outcome assessment and incomplete outcome data were the most common reasons for risk of bias. For the most poorly rated item across studies, only 3 trials demonstrated suitable blinding of participants and personnel.

**Table 1.** Risk of bias assessment of the included studies.

Study	Risk of bias items						
	1 <sup>a</sup>	2 <sup>b</sup>	3 <sup>c</sup>	4 <sup>d</sup>	5 <sup>e</sup>	6 <sup>f</sup>	7 <sup>g</sup>
Proudfoot et al (2003) [32]	+ <sup>h</sup>	+	— <sup>h</sup>	? <sup>h</sup>	+	+	+
Grime (2004) [33]	+	+	?	?	?	?	+
Proudfoot et al (2004) [34]	+	+	—	?	+	+	+
Marks et al (2004) [35]	+	+	—	+	?	+	?
Schneider et al (2005) [36]	+	+	—	+	+	+	+
Mackinnon et al (2008) [37]	?	?	—	?	+	+	+
Kessler et al (2009) [38]	+	+	+	+	+	+	+
Ellis et al (2011) [39]	?	?	—	?	?	?	+
Farrer et al (2011) [40]	+	+	?	?	+	?	+
Høifødt et al (2013) [41]	+	+	—	?	+	+	+
Lintvedt et al (2013) [42]	+	+	—	?	+	+	+
Powell et al (2013) [43]	+	+	—	?	+	+	+
Sethi (2013) [44]	+	+	—	—	+	+	+
Howells et al (2016) [45]	+	+	+	?	—	?	+
Phillips et al (2014) [46]	+	+	+	+	+	+	+
Twomey et al (2014) [47]	+	+	—	—	—	+	?
Gilbody et al (2015) [48]	+	+	—	—	+	+	+
Richards et al (2015) [49]	+	+	?	?	+	+	+
Richards et al (2016) [50]	+	+	?	?	+	+	+
Carolan et al (2017) [51]	+	+	—	?	+	+	+
Flett et al (2018) [52]	+	?	—	?	+	+	+
Forand et al (2018) [53]	+	?	—	—	+	+	+
Bostock et al (2019) [54]	+	?	—	?	+	?	+
Löbner et al (2019) [55]	+	+	—	?	+	+	?

<sup>a</sup>Random sequence generation (selection bias).<sup>b</sup>Allocation concealment (selection bias).<sup>c</sup>Blinding of participants and personnel (performance bias).<sup>d</sup>Blinding of outcome assessment (performance bias).<sup>e</sup>Incomplete outcome data (attrition bias).<sup>f</sup>Selective outcome reporting.<sup>g</sup>Other potential threats to validity.<sup>h</sup>+ = low risk; — = high risk; ? = unclear risk.

## Study Characteristics

Out of the 48 NHS e-therapies identified by Bennion et al [10], a total of 7 (15%) were based on RCT evidence of efficacy, which comprised 6 web-based e-therapies and 1 smartphone-based e-therapy (Table 2). MoodGYM was the e-therapy with the greatest degree of evaluation ( $k=11$  studies), with 2 of the e-therapies having a single RCT evaluation (ie, Ieso and WorkGuru). All 6 web-based e-therapies had both clinical and academic personnel adding expertise during technological development, but the smartphone-based e-therapy had no clinical or academic personnel being involved in its technological development phase [11]. A summary of e-therapy

version numbers used in each study and whether a CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [56] was provided (for studies published post-2011 after the checklist was developed) is reported in Multimedia Appendix 3 [32-55]. Reporting of version numbers was generally inconsistent, meaning establishing whether the e-therapies had been updated between studies was difficult. Beating the Blues had been updated between studies, with version 1.0 used in the early studies (2003-2004) [32,34] and version 2.5 used in the most recent study (2018) [53]. Updates to MoodGYM were unable to be established because of inconsistent reporting of version numbers, but there was an



indication that the studies between 2011 and 2018 used version III [41,42,55]. It appeared that Headspace was updated from version 1.0 or above in 2014 to a version equal to or above 2.0 in studies from 2019. Studies of FearFighter, SilverCloud, IESO, and WorkGuru either did not refer to version numbers or were only evaluated in 1 RCT, so updates could not be conclusively determined.

All but one of the therapies were based on the cognitive behavioral theory (CBT) [11]. E-therapy treatments lasted between 10 and 70 days (mean 44.52, SD 16.11), comprising between 3 and 45 sessions (mean 8.37, SD 7.98) lasting 10 to 60 min each (mean 48.21, SD 15.26). The majority of e-therapies were administered weekly ( $k=19$ ), whereas 3 of the trials required daily e-therapy usage (2 trials did not report the instructed frequency of usage). Self-help typology was characterized as self-administered therapy ( $k=7$  studies), predominantly self-help ( $k=11$  studies), minimal contact therapy ( $k=5$  studies), and predominantly therapist-delivered treatment ( $k=1$  study).

The control conditions employed in the studies were waitlist or no treatment ( $k=13$ ), TAU ( $k=5$ ), and placebo or attention-control tasks ( $k=9$ ; note:  $k=3$  studies had multiple control conditions). TAU comprised usual general practitioner (GP) care, allowing access to any treatment prescribed or referred to by a GP. Placebo or attention-control conditions included depression information websites (eg, Bluepages;  $k=2$ ), online peer support forums (eg, MoodGarden;  $k=1$ ), tracking or structured weekly phone calls ( $k=2$ ), neutral tasks or note-taking organization apps (eg, Catch notes software or

Evernote;  $k=2$ ), or online self-relaxation (without exposure, ie, a sham treatment; eg, managing anxiety or de-STRESS;  $k=2$ ). In  $k=12$  trials, clinical participants were recruited from primary care ( $k=7$ ), psychiatric outpatients ( $k=2$ ), a university counseling center ( $k=1$ ), public sector employees ( $k=1$ ), and a telephone counseling service ( $k=1$ ). In the remaining 12 trials, community participants were recruited from university students ( $k=3$ ), occupational health attendees ( $k=3$ ), the internet ( $k=2$ ), electoral role ( $k=1$ ), youth center ( $k=1$ ), charity users ( $k=1$ ), and *treatment-seeking adults* ( $k=1$ ). Mean ages across the samples ranged from 20 to 45 years (mean 35.71, SD 7.76).

E-therapies were delivered for symptoms of depression ( $k=10$ ), anxiety or panic and phobia ( $k=3$ ), stress ( $k=2$ ), or a combination of anxiety and depression symptoms ( $k=6$ ). Three of the trials did not require participants to have any symptoms or indicators of poor mental health. The Beck Depression Inventory (I or II) was the most commonly used depression outcome measure ( $k=7$ ), followed by the Centre for Epidemiologic Studies Depression Scale (CES-d;  $k=6$ ). The most commonly employed anxiety outcome measures were the Generalized Anxiety Disorder-7 ( $k=4$ ) and the Depression Anxiety Stress Scales—anxiety subscale ( $k=4$ ). Follow-up assessments were conducted in 18 trials ( $k=2$  had insufficient data to be included in the follow-up analysis). The duration of follow-up ranged between 1 and 20 months (mean 5 months). Dropout rates ranged from 0% to 64%. The average e-therapy dropout rate was 31% (SD 17.35), and the average dropout rate for controls was 17% (SD 13.31). Therefore, significantly more participants dropped out during e-therapies compared with controls ( $U=181.000$ ;  $Z=-3.026$ ;  $P=.002$ ).

**Table 2.** Types of e-therapies used in included studies.

E-therapy	Number of trials <sup>a</sup>	Delivery platform	Clinical involvement	Academic involvement	Psychological theory or clinical approach used	Evidence of updates between studies
Beating the Blues	5	Web-based	Y <sup>b</sup>	Y	CBT <sup>c</sup>	Yes
Fear Fighter	2	Web-based	Y	Y	CBT	Could not be determined
Headspace	3	Phone-based	N <sup>d</sup>	N	Mindfulness	Yes
IESO	1	Web-based	Y	Y	CBT	N/A <sup>e</sup>
MoodGYM	11	Web-based	Y	Y	CBT	Could not be determined
SilverCloud Health	2	Web-based	Y	Y	CBT	N/A
WorkGuru	1	Web-based	Y	Y	CBT, mindfulness, and PP <sup>f</sup>	N/A

<sup>a</sup>A total of 2 e-therapies were evaluated in one trial; therefore, the total number of trials exceeded the overall number of included studies.

<sup>b</sup>Y: yes.

<sup>c</sup>CBT: cognitive behavioral therapy.

<sup>d</sup>N: no.

<sup>e</sup>N/A: not applicable, as e-therapy content was not assessed in multiple studies.

<sup>f</sup>PP: positive psychology.

## Meta-Analysis of E-Therapy Versus Controls

Meta-analytic comparisons were performed to aggregate the effect of e-therapy vs controls on (1) depression and (2) anxiety and stress symptoms at posttreatment and follow-up. GRADE assessments are reported for each comparison, indicating the quality of evidence. All comparisons were based on RCT

evidence so they started as high-quality evidence. Across the meta-analyses, limited issues were found in terms of study limitations or publication bias, but some limitations were found for heterogeneity, treatment comparisons, and imprecision. As a result, the level of evidence was downgraded for all comparisons, with the majority demonstrating moderate quality. Comparisons were downgraded one level specifically due to

significant and considerable  $I^2$  statistic indicating marked heterogeneity in the original studies, variability in primary outcome measure, differing control groups, and varied effects based on lower and upper bounds of confidence intervals. One comparison was downgraded 2 levels to low-quality evidence because of additional limitations created by the small number of studies restricting subsequent moderator analyses and variability in follow-up time.

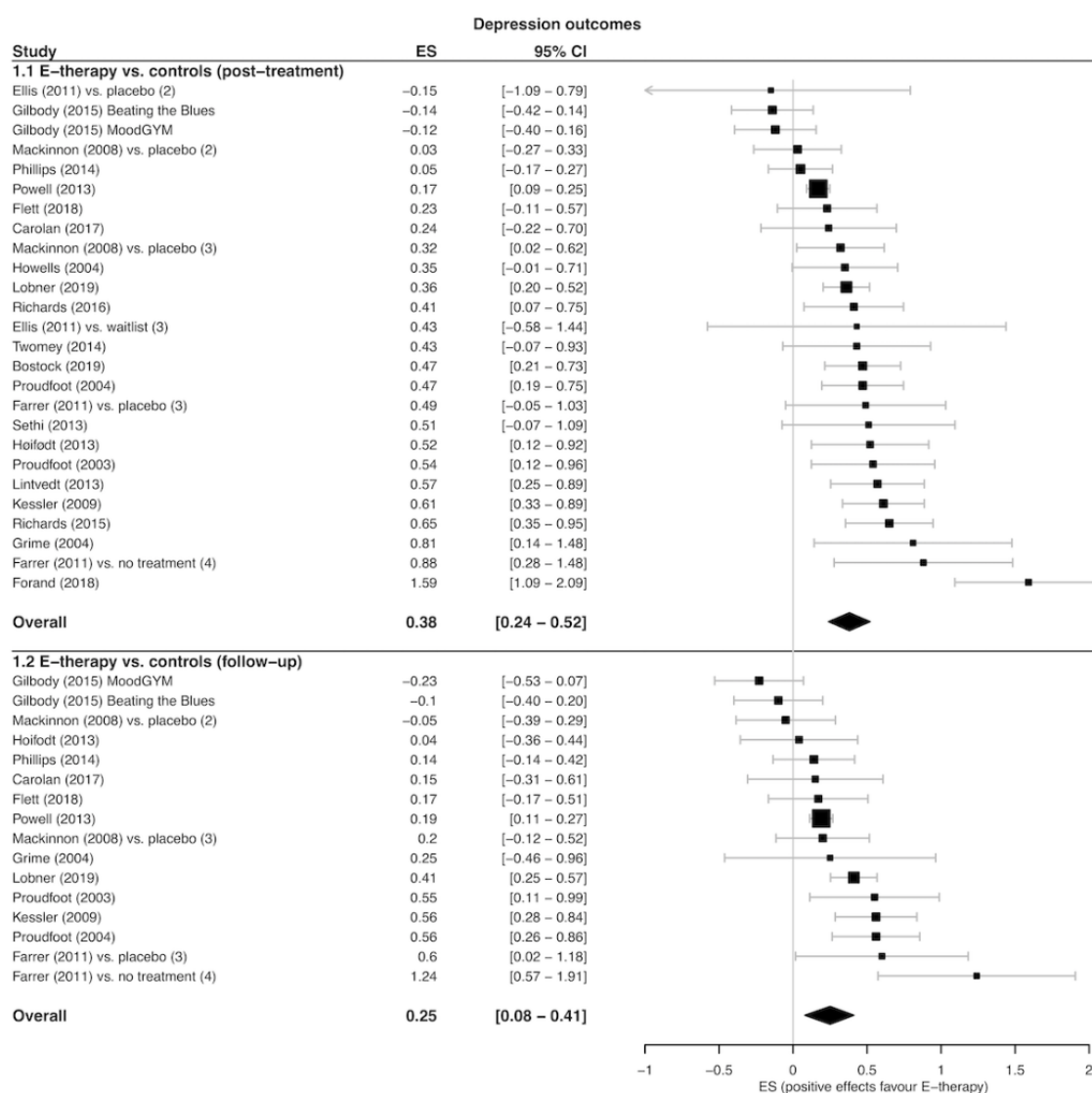
### Effect of E-Therapy on Depression Outcomes

#### Posttreatment and Follow-Up Comparisons

Overall, 26 treatment arm comparisons (extracted from 22 studies) totaling 7075 participants evaluated posttreatment e-therapy depression outcomes in comparison with a control condition (e-therapy,  $n=3545$ ; control,  $n=3530$ ). The pooled SMD presented in Figure 2 signified a small, significant treatment effect in favor of greater depression reductions

following e-therapy (SMD 0.38; 95% CI 0.24 to 0.52;  $Z=5.78$ ;  $P<.001$ ; GRADE=moderate). The NNT was 4.72, indicating that for every 5 patients who received e-therapy, there was one additional beneficial depression outcome compared with if they had received a control condition. Between-study variation was significant, indicating substantial heterogeneity between studies ( $I^2=73\%$ ; 95% CI 60% to 82%;  $Q=92.30$ ;  $P<.001$ ). Furthermore, 16 follow-up treatment arm comparisons (extracted from 13 studies) provided follow-up data on depression outcomes for e-therapies versus control conditions for 5709 participants (e-therapy,  $n=2850$ ; control,  $n=2859$ ). There was a small significant pooled SMD in favor of depression outcomes at follow-up compared with controls (Figure 2; SMD 0.25; 95% CI 0.08 to 0.41;  $Z=3.23$ ;  $P=.001$ ; NNT=7.12; GRADE=moderate). The between-study variation was significant, indicating moderate-to-substantial heterogeneity ( $I^2=69\%$ ; 95% CI 48% to 81%;  $Q=48.11$ ;  $P<.001$ ).

**Figure 2.** Forest plot of post-treatment and follow-up depression outcome effect sizes (ES) for e-therapy versus controls.



### Moderator and Sensitivity Analyses

The significant heterogeneity between studies at posttreatment and follow-up was investigated using meta-regression (Table 3) and subgroup moderator analyses (Table 4). Meta-regression analyses found that variations in e-therapy treatment effects were not explained by gender, age, number of sessions, or study quality at posttreatment or follow-up. Although initial depression severity was not significantly associated with effect size at posttreatment, higher levels of depression severity were associated with larger beneficial effects of e-therapy at follow-up. Subgroup analyses showed that variation in posttreatment effect size was associated with the type of control

condition (although the effect fell short of significance after accounting for multiple testing). A moderate effect was observed in favor of e-therapy vs wait list controls, whereas the effects for e-therapy compared with placebo conditions and TAU were small. At follow-up, e-therapy effect sizes did not significantly differ according to the control type with e-therapy, showing a small significant beneficial effect compared with placebo and TAU controls and a small nonsignificant effect compared with wait list. Posttreatment and follow-up effects were not significantly affected by the e-therapy type, self-help typology, recruitment setting, focus problem, or analysis method. Substantial significant heterogeneity was evident in approximately half of the subgroups.

**Table 3.** Meta-regression analyses of effect e-therapy vs controls on depression and anxiety outcomes (posttreatment and follow-up).

Time point and outcome, variable	<i>k</i> <sup>a</sup>	B coefficient	95% CI	SE	<i>P</i> value <sup>b</sup>	<i>R</i> <sup>2</sup> (%) <sup>c</sup>
<b>Posttreatment</b>						
<b>Depression</b>						
Initial severity	26	0.07	−0.06 to 0.21	0.06	.26	4.15
Percentage of males	26	−0.01	−0.02 to 0.00	0.01	.09	8.30
Mean age (years)	26	0.00	−0.02 to 0.01	0.01	.58	0.95
Mean number of sessions completed	17	0.02	0.00 to 0.05	0.01	.08	10.23
Risk of bias	26	−0.01	−0.11 to 0.08	0.05	.77	0.28
<b>Follow-up<sup>c</sup></b>						
<b>Depression</b>						
Initial severity	16	0.25	0.12 to 0.39	0.06	<.001	53.17
Percentage of males	16	−0.01	−0.03 to 0.01	0.01	.13	11.64
Mean age (years)	16	0.01	−0.01 to 0.04	0.01	.38	3.88
Mean number of sessions completed	11	0.01	−0.06 to 0.08	0.03	.78	0.44
Risk of bias	16	0.02	−0.11 to 0.14	0.06	.78	0.40
<b>Posttreatment</b>						
<b>Anxiety<sup>d</sup></b>						
Initial severity	17	0.12	−0.07 to 0.31	0.09	.17	8.84
Percentage of males	17	−0.01	−0.03 to 0.01	0.01	.24	5.85
Mean age (years)	17	−0.01	−0.03 to 0.01	0.01	.43	3.03
Mean number of sessions completed	11	0.02	0.00 to 0.05	0.01	.07	23.93
Risk of bias	17	−0.01	−0.14 to 0.12	0.06	.85	0.18

<sup>a</sup>*k*: number of comparisons.

<sup>b</sup>Alpha threshold Bonferroni adjusted to  $P < .01$  for multiple testing.

<sup>c</sup>Insufficient number of comparisons and limited between-study heterogeneity to warrant moderator analyses of anxiety outcomes at follow-up.

<sup>d</sup>*R*<sup>2</sup>: percentage of variance explained by the moderator.

**Table 4.** Subgroup analysis of effect e-therapy versus controls on depression outcomes (posttreatment and follow-up).

Time point and variable, Subgroup	<i>k</i> <sup>a</sup>	SMD <sup>b</sup> (Hedges <i>g</i> ) <sup>c</sup>	95% CI	<i>I</i> <sup>2</sup> (%) <sup>d</sup>	<i>P</i> value (between subgroups) <sup>e</sup>	<i>R</i> <sup>2</sup> (%) <sup>f</sup>	NNT <sup>g</sup>
<b>Posttreatment</b>							
<b>Control type</b>							
Wait list	12	0.54 <sup>h</sup>	0.34 to 0.75	79 <sup>h</sup>	.02	8.00	3.36
TAU <sup>i</sup>	7	0.32 <sup>h</sup>	0.06 to 0.58	79 <sup>h</sup>	— <sup>j</sup>	—	5.58
Placebo	7	0.20 <sup>h</sup>	0.06 to 0.34	2	—	—	8.89
<b>E-therapy type</b>							
MoodGYM	14	0.29 <sup>h</sup>	0.15 to 0.43	57 <sup>h</sup>	.30	3.94	6.15
Beating the Blues	5	0.55 <sup>h</sup>	0.00 to 1.10	89 <sup>h</sup>	—	—	3.30
Headspace	3	0.36 <sup>h</sup>	0.22 to 0.49	0	—	—	4.97
Other	4	0.50 <sup>h</sup>	0.32 to 0.68	2	—	—	3.61
<b>Self-help typology</b>							
Self-administered	8	0.30 <sup>h</sup>	0.15 to 0.45	65 <sup>h</sup>	.08	5.87	5.95
Predominantly self-help	14	0.39 <sup>h</sup>	0.16 to 0.62	76 <sup>h</sup>	—	—	4.60
Minimal contact	3	0.53 <sup>h</sup>	0.39 to 0.67	0	—	—	3.42
Predominantly therapist delivered	1 <sup>k</sup>	0.61	—	—	—	—	2.95
<b>Setting</b>							
Clinical	12	0.39 <sup>h</sup>	0.22 to 0.57	68 <sup>h</sup>	.91	0.01	4.60
Community	14	0.38 <sup>h</sup>	0.18 to 0.58	76 <sup>h</sup>	—	—	4.72
<b>Focus problem</b>							
Depression	12	0.39 <sup>h</sup>	0.13 to 0.64	84 <sup>h</sup>	.74	0.79	4.60
Anxiety or stress	3	0.38 <sup>h</sup>	0.25 to 0.52	0	—	—	4.72
Both	7	0.47 <sup>h</sup>	0.29 to 0.65	0	—	—	3.84
<b>Analysis method</b>							
ITT <sup>l</sup>	9	0.39 <sup>h</sup>	0.24 to 0.54	76 <sup>h</sup>	.50	0.49	4.60
Completers	3	0.33 <sup>h</sup>	0.21 to 0.44	0	—	—	5.42
<b>Follow-up</b>							
<b>Control type</b>							
Wait list	4	0.29	−0.15 to 0.73	71 <sup>h</sup>	.75	1.19	6.15
TAU	7	0.29 <sup>h</sup>	0.03 to 0.54	79 <sup>h</sup>	—	—	6.15

Time point and variable, Subgroup	<i>k</i> <sup>a</sup>	SMD <sup>b</sup> (Hedges <i>g</i> ) <sup>c</sup>	95% CI	<i>I</i> <sup>2</sup> (%) <sup>d</sup>	<i>P</i> value (between subgroups) <sup>e</sup>	<i>R</i> <sup>2</sup> (%) <sup>f</sup>	NNT <sup>g</sup>
Placebo	5	0.18 <sup>h</sup>	0.00 to 0.36	0	—	—	9.87
<b>E-therapy type</b>							
MoodGYM	9	0.21	−0.01 to 0.43	73 <sup>h</sup>	.79	0.96	8.47
Beating the Blues	4	0.31	−0.03 to 0.64	73 <sup>h</sup>	—	—	5.76
Other	3	0.32 <sup>h</sup>	0.05 to 0.59	51	—	—	5.58
<b>Self-help typology</b>							
Self-administered	4	0.16	−0.10 to 0.41	80 <sup>h</sup>	.46	1.29	11.10
Predominantly self-help	10	0.29 <sup>h</sup>	0.07 to 0.51	65 <sup>h</sup>	—	—	6.15
Minimal contact	1 <sup>k</sup>	0.04	—	—	—	—	44.32
Predominantly therapist delivered	1 <sup>k</sup>	0.56	—	—	—	—	3.25
<b>Setting</b>							
Clinical	10	0.33 <sup>h</sup>	0.09 to 0.57	77 <sup>h</sup>	.13	4.68	5.42
Community	6	0.14 <sup>h</sup>	0.07 to 0.21	0	—	—	12.68
<b>Focus problem</b>							
Depression	10	0.22	−0.01 to 0.46	77 <sup>h</sup>	.07	7.42	8.08
Anxiety or stress	1 <sup>k</sup>	0.15	—	—	—	—	11.83
Both	3	0.49 <sup>h</sup>	0.32 to 0.66	0	—	—	3.69
<b>Analysis method</b>							
ITT	3	0.27 <sup>h</sup>	0.09 to 0.45	71 <sup>h</sup>	—	—	6.60
Completers	1 <sup>k</sup>	0.17	—	—	—	—	10.45

<sup>a</sup>*k*: number of comparisons.

<sup>b</sup>SMD: standardized mean difference.

<sup>c</sup>Positive effect size indicates in favor of e-therapy.

<sup>d</sup>Significance of associated *Q* statistic.

<sup>e</sup>Alpha threshold Bonferroni adjusted to *P* < .01 for multiple testing.

<sup>f</sup>*R*<sup>2</sup>: percentage of variance explained by moderator.

<sup>g</sup>NNT: number needed to treat.

<sup>h</sup>Significant at *P* < .05.

<sup>i</sup>TAU: treatment as usual.

<sup>j</sup>One between-groups *P* value and *R*<sup>2</sup> value are provided for each subgroup comparison, reported on the row of the first subgroup category.

<sup>k</sup>Where there is only one comparison within a subgroup, 95% confidence intervals and *I*<sup>2</sup> values are not reported.

<sup>l</sup>ITT: intention to treat.

Sensitivity analyses explored the impact of the extreme outliers and length of follow-up on the pooled depression effect sizes. Although the removal of outlier effects resulted in a slight reduction in the effect of e-therapy on depression from 0.38 to 0.34 at posttreatment and from 0.25 to 0.22 at follow-up,

outcomes still indicated small, significant benefits of e-therapy compared with controls. E-therapy demonstrated a small, beneficial effect compared with controls at short-term and medium-term follow-up, which diminished at long-term



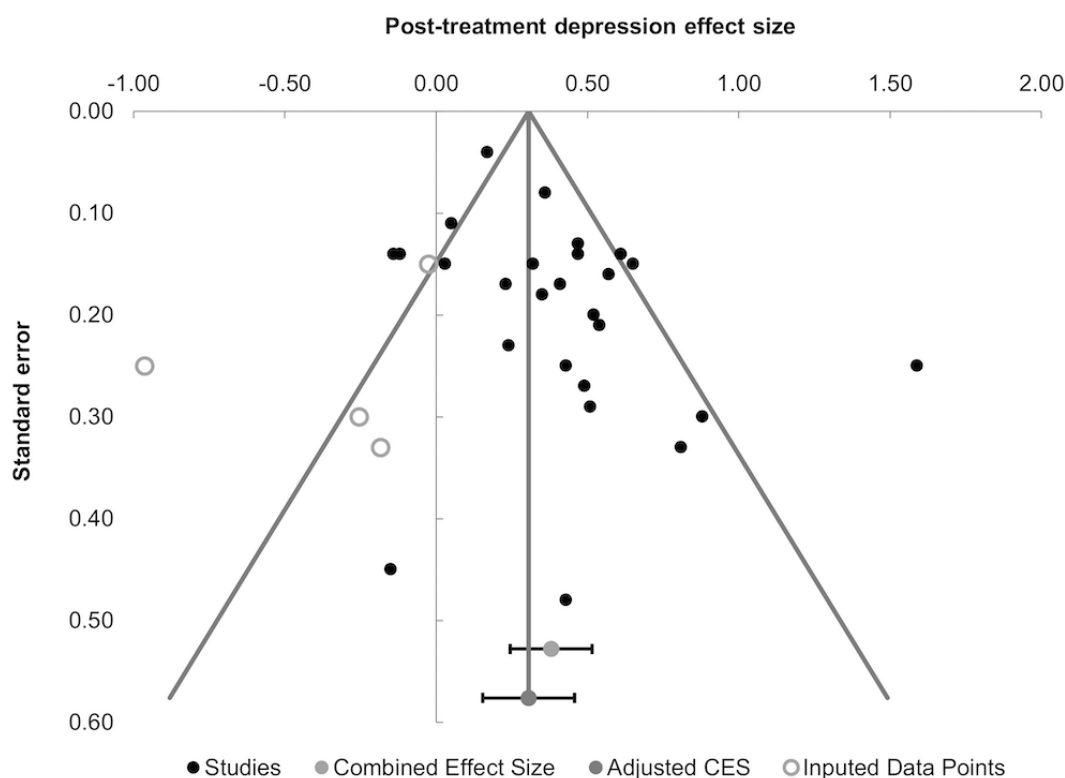
follow-up. The full sensitivity analysis results are reported in [Multimedia Appendix 4](#).

### Assessment of Publication Bias

Visual inspection of the posttreatment funnel plot ([Figure 3](#)) suggested that there was some asymmetry in the distribution of studies, indicating that the smaller included studies were more likely to report larger effects for e-therapy interventions. Trim and fill imputed missing data to represent 4 smaller studies with effects more in favor of controls, producing a slightly reduced

adjusted effect size in favor of e-therapy (SMD 0.31; 95% CI 0.15 to 0.46). Statistical testing of publication bias using Egger's regression did not detect significant asymmetry in the study distribution for posttreatment outcomes ( $B=-0.15$ ;  $t_{25}=1.49$ ;  $P=.15$ ). Assessment of study distribution for follow-up depression outcomes also did not detect a significant influence of publication bias ( $B=0.31$ ;  $t_{15}=1.34$ ;  $P=.20$ ). Taken together, the multiple assessments of publication bias suggest a minimal-to-small influence of bias on the overall e-therapy treatment effect for depression outcomes.

**Figure 3.** Funnel plot for distribution of studies reporting e-therapy versus controls post-treatment depression outcomes.

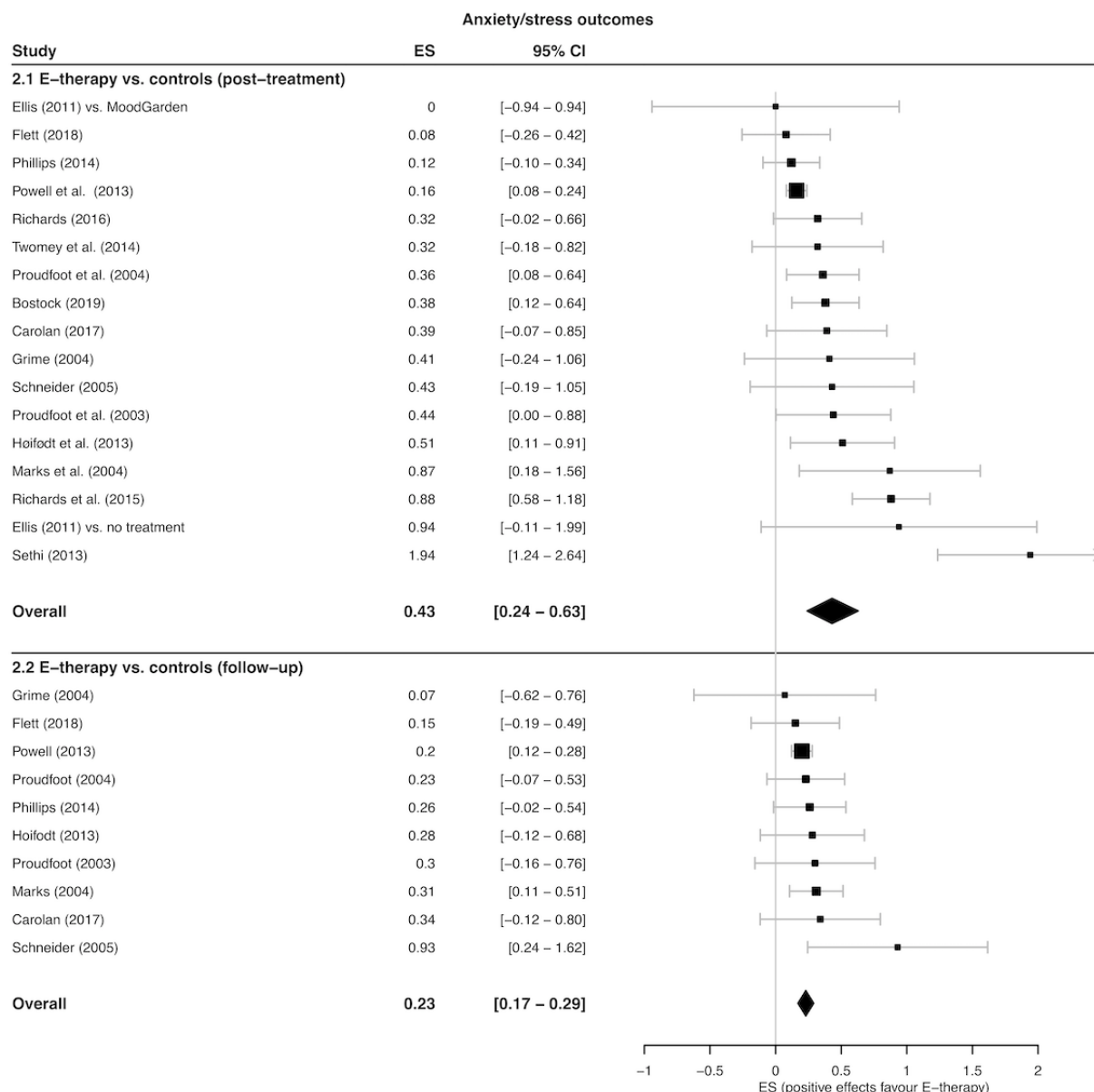


### Effect of E-Therapy on Anxiety and Stress Outcomes

#### Posttreatment and Follow-Up Comparisons

Overall, 17 treatment arm comparisons (extracted from 16 studies) totaling 4863 participants evaluated posttreatment e-therapy anxiety and stress outcomes alongside a control condition (e-therapy,  $n=2443$ ; control,  $n=2420$ ). The pooled SMD presented in [Figure 4](#) signified a small-to-moderate, significant treatment effect in favor of greater anxiety reductions following e-therapy (SMD=0.43; 95% CI 0.24 to 0.63;  $Z=4.63$ ;  $P<.001$ ; GRADE=moderate). The NNT was 4.18, indicating that for approximately every 4 patients who received e-therapy,

there was one additional beneficial anxiety and stress outcome compared with if they had received a control condition. The between-study variation was significant, indicating substantial heterogeneity ( $I^2=73\%$  [95% CI 56% to 83%];  $Q=59.13$ ;  $P<.001$ ). Furthermore, 10 studies provided follow-up data on anxiety and stress outcomes for e-therapies vs control conditions for 3983 participants (e-therapy,  $n=2000$ ; control,  $n=1983$ ). At follow-up, there was a small, significant pooled SMD in favor of e-therapy compared with controls ([Figure 4](#); SMD=0.23; 95% CI 0.17 to 0.29;  $Z=8.30$ ;  $P<.001$ ; NNT=7.74; GRADE=low). The between-study variation was minimal and not significant ( $I^2=0\%$  [95% CI 0% to 46%];  $Q=6.31$ ;  $P=.71$ ).

**Figure 4.** Forest plot of post-treatment and follow-up stress/anxiety outcome effect sizes (ES) for e-therapy versus controls.

### Moderator and Sensitivity Analyses

The significant heterogeneity between studies at posttreatment was investigated with meta-regression (Table 3) and subgroup moderator analyses. Minimal heterogeneity and an insufficient number of studies ( $k < 10$ ) negated the need for moderator analysis of follow-up effects. Meta-regression analyses found variations in e-therapy posttreatment anxiety and stress effects were not explained by initial severity, gender, age, number of sessions, or study quality. Subgroup analyses showed that posttreatment effect sizes for anxiety and stress symptoms did not significantly differ for different control conditions. However, e-therapy vs wait list produced a moderate, significant effect compared with the small effects observed for TAU and placebo controls (placebo effect not significant). Posttreatment effects were not significantly affected by the e-therapy type, recruitment setting, focus problem, or analysis method. Self-help typology indicated larger effects were observed for therapies with greater

therapist involvement ( $P = .02$ ); however, the effect did not remain significant when applying a Bonferroni correction. Substantial significant heterogeneity was evident in about a quarter of the subgroups.

Sensitivity analyses explored the impact of extreme outliers and length of follow-up on the pooled anxiety and stress effect sizes. Although the removal of outlier effects resulted in a slight reduction in the e-therapy treatment effect on anxiety from 0.43 to 0.37 at posttreatment and from 0.23 to 0.22 at follow-up, the outcomes still indicated small, significant benefits of e-therapy compared with controls. E-therapy demonstrated a small, beneficial effect compared with controls at both short-term and medium-term follow-up (insufficient studies of long-term follow-up were available). The full sensitivity analysis results are reported in Multimedia Appendix 4.

The significant heterogeneity between studies at posttreatment was investigated with subgroup moderator analyses (Table 5)

**Table 5.** Subgroup analysis of effect e-therapy versus controls on anxiety and stress outcomes (posttreatment).

Time point <sup>a</sup> and variable, Subgroup	<i>k</i> <sup>b</sup>	SMD <sup>c</sup> (Hedges <i>g</i> ) <sup>d</sup>	95% CI	<i>I</i> <sup>2</sup> (%) <sup>e</sup>	<i>P</i> value (between subgroups) <sup>f</sup>	<i>R</i> <sup>2</sup> (%) <sup>g</sup>	NNT <sup>h</sup>
<b>Posttreatment</b>							
<b>Control type</b>							
Wait list	9	0.55 <sup>i</sup>	0.24 to 0.86	84 <sup>i</sup>	.41	2.99	3.04
TAU <sup>j</sup>	3	0.40 <sup>i</sup>	0.35 to 0.45	0	— <sup>k</sup>	—	4.49
Placebo	5	0.26	−0.02 to 0.55	28	—	—	6.86
<b>E-therapy type</b>							
MoodGYM	7	0.44 <sup>i</sup>	0.01 to 0.86	80 <sup>i</sup>	.86	0.50	4.09
Beating the Blues	3	0.40 <sup>i</sup>	0.35 to 0.45	0	—	—	4.49
Other	7	0.46 <sup>i</sup>	0.24 to 0.68	61 <sup>i</sup>	—	—	3.92
<b>Self-help typology</b>							
Self-administered	4	0.23 <sup>i</sup>	0.09 to 0.36	8	.02	13.38	7.74
Predominantly self-help	8	0.47 <sup>i</sup>	0.11 to 0.83	74 <sup>i</sup>	—	—	3.84
Minimal contact	5	0.60 <sup>i</sup>	0.36 to 0.83	45	—	—	3.04
Predominantly therapist delivered	0 <sup>l</sup>	—	—	—	—	—	—
<b>Setting</b>							
Clinical	8	0.44 <sup>i</sup>	0.33 to 0.54	0	.99	0.00	4.09
Community	9	0.44 <sup>i</sup>	0.10 to 0.78	84 <sup>i</sup>	—	—	4.09
<b>Focus problem</b>							
Depression	3	0.49 <sup>i</sup>	0.05 to 0.93	88 <sup>i</sup>	.85	0.82	3.69
Anxiety or stress	5	0.44 <sup>i</sup>	0.27 to 0.62	0	—	—	4.09
Anxiety or depression	7	0.58 <sup>i</sup>	0.14 to 1.02	70 <sup>i</sup>	—	—	3.14
<b>Analysis method</b>							
ITT <sup>m</sup>	7	0.47 <sup>i</sup>	0.27 to 0.68	75 <sup>i</sup>	.06	5.76	3.84
Completers	2	0.18	−0.05 to 0.42	0	—	—	9.87

<sup>a</sup>Insufficient number of comparisons and limited between-study heterogeneity to warrant moderator analyses of anxiety outcomes at follow-up.

<sup>b</sup>*k*: number of comparisons.

<sup>c</sup>SMD: standardized mean difference.

<sup>d</sup>Positive effect size indicates in favor of e-therapy.

<sup>e</sup>Significance of associated *Q* statistic.

<sup>f</sup>Alpha threshold Bonferroni adjusted to  $P < .01$  for multiple testing.

<sup>g</sup>*R*<sup>2</sup>: percentage of variance explained by moderator.

<sup>h</sup>NNT: number needed to treat.

<sup>i</sup>Significant at  $P < .05$ .

<sup>j</sup>TAU: treatment as usual.

<sup>k</sup>One between-groups *P* value and *R*<sup>2</sup> value are provided for each subgroup comparison, reported on the row of the first subgroup category.

<sup>l</sup>Where there are no comparisons within a subgroup, SMD, 95% confidence intervals and *I*<sup>2</sup> values are not reported.

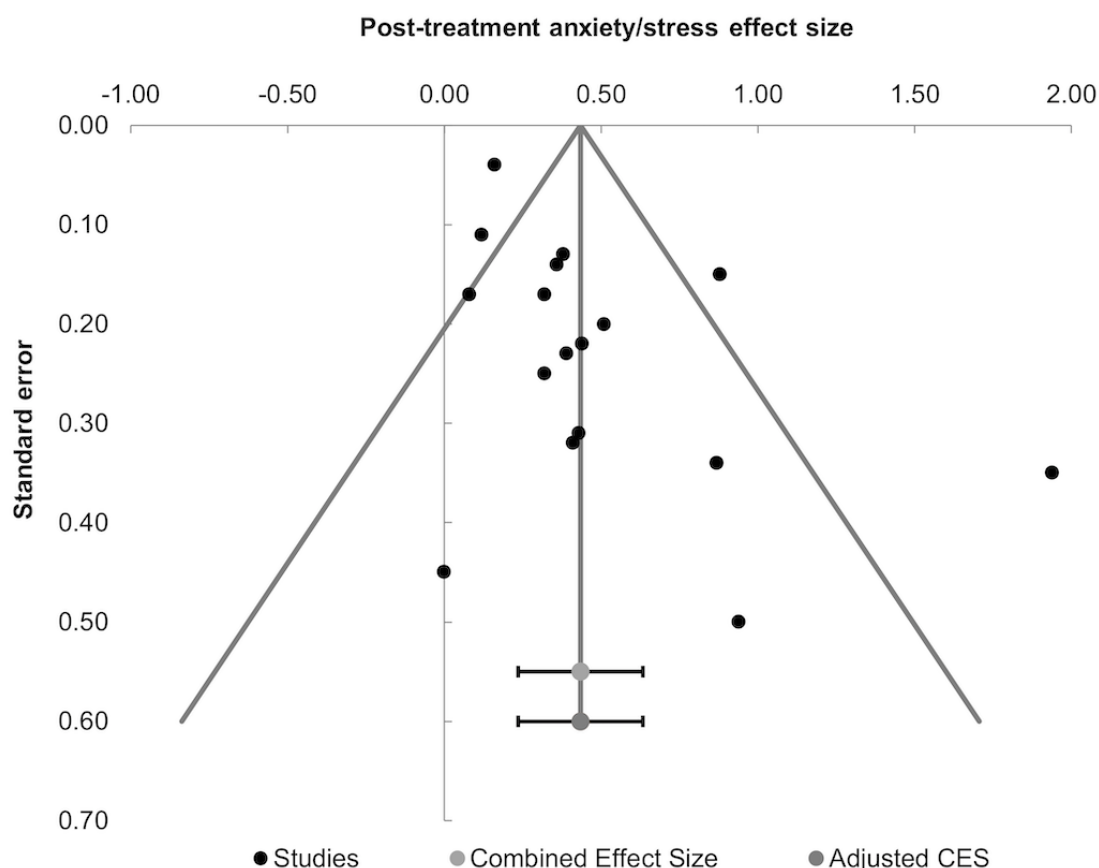
<sup>m</sup>ITT: intention to treat.

### Assessment of Publication Bias

Visual inspection of the funnel plot in Figure 5 suggested that there was some asymmetry in the distribution of studies reporting posttreatment anxiety and stress outcomes. However, the trim and fill imputation did not impute any missing data in relation to smaller studies in favor of controls or minimal differences between groups producing an adjusted effect size identical to the initial pooled SMD. The Egger regression failed

to detect sufficient asymmetry in the study distribution of posttreatment anxiety and stress outcomes ( $B=-0.35$ ;  $t_{16}=1.82$ ;  $P=.09$ ). Taken together, the multiple assessments of publication bias imply a minimal-to-small influence of reporting bias on the overall e-therapy treatment effect for anxiety and stress outcomes. There were insufficient studies ( $k<10$ ) to enable accurate assessment of publication bias on comparisons of follow-up anxiety and stress outcomes.

**Figure 5.** Funnel plot for distribution of studies reporting e-therapy versus controls post-treatment anxiety/stress outcomes.



## Discussion

### Principal Findings

This study has been the first attempt to assess the breadth and quality of the evidence base for NHS-recommended e-therapies and to quantify the efficacy of this health technology through a meta-analysis of the clinical trial evidence base. Only 15% (7/48) of the NHS-recommended e-therapies had eligible RCT studies underpinning their clinical evaluation. Of the 7 e-therapies with RCT evidence, 2 contributed a single RCT study to the meta-analysis, and there was poor and variable reporting of version numbers across studies. These findings are at odds with the philosophy of evidenced-based practice,

whereby clinical guidelines are underpinned by gold standard evidence of efficacy. Overall, however, the available good quality evidence shows that the e-therapies tested do benefit adult participants in better managing anxiety, stress and depression compared with controls, and this appears to be a durable effect in the short to medium term. The magnitude of the e-therapy treatment effects found here mirrors the effect sizes seen in the overall LI intervention evidence base ( $g=0.2-0.5$ ) [5]. The NNT analysis suggests that for every 5 patients treated with an e-therapy, one has a good outcome. The acceptability and efficacy of the e-therapies without RCT evidence (ie, 85%, 41/48) of those actually recommended for use in the NHS) remains open to question. It would be premature to clinically champion any single e-therapy as being the most

effective at this point in time. MoodGYM has been exposed to most evaluation and scrutiny, but it was unclear whether differing versions were being tested.

The acceptability of e-therapies can be called into question because of the higher dropout rates compared with controls reported here. Criticisms of LI psychological interventions, and e-therapies in particular, have been previously made concerning their high dropout rates being an index for poor patient acceptability, because of the low therapist contact and time approach [13,16,57,58]. Dropout rates may also have been influenced by multiple (unmeasured) factors such as the poor face validity of the CBT theoretical approach [59], low readiness to change, poor attitudes to the delivery of eHealth [60], and the usability or characteristics of the web or app design itself [61,62]. Ongoing issues with poor acceptability will remain an obstacle in the commissioning and delivery of e-therapies as frontline LI psychological interventions. Clearly, the clinical utility of any e-therapies needs to be considered in a matrix of cost, safety, acceptability, feasibility, and efficacy evidence [63].

Comparison of study characteristics highlighted noteworthy commonalities and differences across and between e-therapies. First, 5 of the 7 e-therapies evaluated were based on CBT (one other was based on CBT alongside other approaches). This mirrors that LI interventions as a whole tend to be based and focused on variants of CBT [64]. Recent innovations in e-therapies have included acceptance and commitment therapy [65], interpersonal psychotherapy [66], mindfulness [67], and psychodynamic psychotherapy [68]. Second, 6 of the 7 e-therapies were web based, so the clinical utility of smartphone-based app delivery of NHS-recommended e-therapies has not been appropriately empirically evaluated.

Variations in e-therapy treatment effects were explored with moderator analyses, as a previous individual participant meta-analysis of e-therapies for depression found few significant moderators [13]. Significantly larger e-therapy effects were apparent when compared with wait list controls (for posttreatment depression outcomes), for patients with greater baseline severity (for follow-up depression outcomes), and when there was a greater amount of therapist input (for end of treatment anxiety and stress outcomes). However, the effects of control type and amount of therapist input did not remain significant after accounting for multiple testing, so caution should be taken with any conclusions. Larger wait list comparison effects are commonly observed in psychotherapy trials and when taken in isolation can lead to overestimated treatment effects [69]. E-therapy effects shrunk as the activeness of comparators increased. In this review, baseline severity was only a significant moderator at follow-up. Greater e-therapy benefits for higher baseline depression severity have previously been shown to predict better outcomes for internet-based CBT [70]. The trend for e-therapies with a greater amount of therapist input generating better outcomes has been widely reported [71-73]. It is worth noting that e-therapy typologies in this meta-analysis emphasized some therapist contact, but that contact time was still relatively brief because of the LI approach. Furthermore, 75% (18/24 studies of 4 different apps) had less than 30 min of real-time person-to-person support. The efficacy

of LI interventions appears to be better enabled when supported by even brief interpersonal contact [72,73].

## Limitations

This review has several limitations, which also highlight how the e-therapy evidence base could be further developed. First, although the included studies were restricted to high-quality RCT evidence, the GRADE approach highlighted issues with inconsistency across results, treatment comparisons, and some imprecision resulting in meta-analytic comparisons of moderate-to-low quality. Second, there are limitations concerning the generalizability of the findings. This review was limited to the treatment of depression, anxiety, and stress with e-therapies and so cannot comment on applicability to other clinical presentations. Services in the United Kingdom use the NICE guidelines to organize the delivery of treatments for anxiety and depression via stepped-care principles. Therefore, the generalizability of results from this meta-analysis is less applicable for different approaches to mental health delivery, for example, via stratified care [74]. The inclusion of only those e-therapies recommended by the NHS excluded those e-therapies very similar in technical format and content.

Third, there were some methodological weaknesses that may have introduced bias, and the conclusions should be treated with caution. The lack of formal screening and selection of articles by a second reviewer is a major limitation that may have led to bias in terms of which studies were selected for inclusion and therefore influenced the results. Similarly, the quality ratings of the studies were made by raters that were not independent from the meta-analysis, and levels of agreement were not optimal [75]. In addition, restrictions in the search strategy may have missed eligible studies or excluded studies evaluating an NHS e-therapy for other clinical presentations or outcomes [76]. Given that eHealth is a rapidly expanding area that makes reviews outdated relatively quickly, the duration since the final searches were conducted (April 2019) means there will undoubtedly be additional relevant e-therapy trials now available. Since the final searches, trials of 3 NHS e-therapies (all with existing trial evidence) have been published; an RCT of SilverCloud used in IAPT [77], evaluations of MoodGYM [78], and Headspace in student samples [79,80].

Finally, synthesis and analysis were restricted by the data from the available studies. The number of trials conducted was small, and thus restricted the power and range of possible moderator analyses. The original studies had the common methodological flaws of limited diagnostic assessments of participants, inconsistent reporting of e-therapy version numbers, overuse of self-reported measures rather than independent assessment, lack of reporting of adverse event rates [63], lack of measures of e-therapy adherence, and lack of true long-term follow-up. The frequent use of passive controls risked inflating treatment effect sizes in meta-analyses [81], and there were insufficient active comparators to establish efficacy of e-therapies vs other therapies. There was no standard definition of dropout or treatment completion across the studies, and therefore, we were forced to adopt the definition used by each study. It is acknowledged that dropout is a limited proxy for acceptability



[82] and that wider indices of acceptability also include understanding barriers to e-therapy engagement.

### Research and Service Implications

Finding studies relating to a specific e-therapy by searching for its name in academic databases proved difficult. This was because before commercialization, many e-therapy platforms were known by their initial project name and not their eventual product name. A solution to this problem would be to ensure that e-therapy developers and researchers register their software on a public database with a unique identifier to be referenced in any subsequent publications. Trials of e-therapies should also be reported according to the CONSORT-EHEALTH checklist [56], and the e-therapy version should be indicated using semantic versioning to clarify whether the e-therapy program being evaluated has been updated (ie, reporting the major, minor, and patch version [eg, version 2.1.1]).

Several e-therapies included in this review were developed to be available without clinical support or guidance (eg, MoodGYM and Headspace). Given that e-therapies outperform controls (with moderate effects compared with wait list), e-therapies may offer particular promise as a waitlist intervention. Although unguided e-therapy may be beneficial to patients waiting for face-to-face psychological interventions, the trend observed in this review and findings from previous studies imply that some clinician involvement is important for ensuring good outcomes if an e-therapy is the sole intervention [72,73]. The manner in which e-therapies can be effectively blended with face-to-face psychological therapies is currently poorly understood and demands more research. Studies also need to be conducted on the utility of e-therapies as wait list interventions.

Given the recent availability of differing theoretical approaches, patient choice for e-therapy can now be offered and researched. Treatment completion rates need to be consistently reported, and trials adopt the ITT approach to reduce biasing treatment effects. Consistent reporting of safety issues (eg, via untoward

incident rates) is needed for e-therapies. Health economic evaluations that are embedded in clinical trials need to be increased. A dropout meta-analysis (with independent study quality ratings of all studies using the latest version of the Cochrane risk of bias tool) of this evidence base is now also indicated to better index e-therapy acceptability issues [83]. Little is known about why patients' drop out of e-therapies, and qualitative investigations would be useful here. Treatment adherence (ie, how much time is spent and how many modules of eHealth are completed by participants) needs to be more consistently reported. The role of moderating factors of treatment outcome in e-therapies needs to be better researched, particularly the role of variables such as blended vs pure e-therapy approaches, time spent on the app, and theoretical approach. E-therapies potentially still play an important role in clinical services, regardless of the organizational system used to coordinate delivery of care [84], particularly when the approach has been well evaluated.

### Conclusions

In this meta-analysis of *gold standard* clinical trials, e-therapies have been found to be efficacious as LI psychological interventions that produce small beneficial effects for adults with depression, anxiety, and stress compared with controls. However, only a relatively small proportion of NHS-recommended e-therapies had been subjected to such *gold standard* evaluation. Although these conclusions should be considered in light of the methodological limitations, the targeted nature of this review to NHS-recommended e-therapies still has relevance to the global field of e-therapies. This is particularly through highlighting the need to consistently integrate high quality and controlled evaluation into the technological development of e-therapies. This is to ensure eventual safe and evidence-based e-therapy practice in routine clinical services. Technological development and scrupulous evaluation of e-therapies need to be conducted in parallel and considered in equipoise.

### Acknowledgments

This study was supported by a PhD studentship awarded by the University of Sheffield to MB and by the Economic and Social Research Council grant number ES/L001365/1.

### Conflicts of Interest

AM was an employee of Ultrasis PLC (no longer trading), the original distributor of Beating the Blues, from September 2010 to December 2012.

#### Multimedia Appendix 1

Example search strategy.

[PDF File (Adobe PDF File), 83 KB - [jmir\\_v22i10e17049\\_app1.pdf](#)]

#### Multimedia Appendix 2

Characteristics of included studies.

[PDF File (Adobe PDF File), 154 KB - [jmir\\_v22i10e17049\\_app2.pdf](#)]

#### Multimedia Appendix 3

Summary of e-therapy version numbers.

[[PDF File \(Adobe PDF File\), 154 KB - jmir\\_v22i10e17049\\_app3.pdf](#)]

Multimedia Appendix 4

Sensitivity analyses.

[[PDF File \(Adobe PDF File\), 33 KB - jmir\\_v22i10e17049\\_app4.pdf](#)]

## References

1. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. *World Psychiatry* 2019 Feb;18(1):97-98 [[FREE Full text](#)] [doi: [10.1002/wps.20592](#)] [Medline: [30600619](#)]
2. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. *Cochrane Database Syst Rev* 2015 Mar 5(3):CD011565. [doi: [10.1002/14651858.CD011565](#)] [Medline: [25742186](#)]
3. Orman J, O'Dea B. e-Therapy in primary care mental health. *Aust J Gen Pract* 2018 Apr;47(4):168-172 [[FREE Full text](#)] [doi: [10.31128/AJGP-11-17-4410](#)] [Medline: [29621855](#)]
4. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. *Br J Psychiatry* 2005 Jan;186:11-17. [doi: [10.1192/bjp.186.1.11](#)] [Medline: [15630118](#)]
5. Delgadillo J. Guided self-help in a brave new world. *Br J Psychiatry* 2018 Feb;212(2):65-66. [doi: [10.1192/bjp.2017.17](#)] [Medline: [29436327](#)]
6. Hill C, Martin JL, Thomson S, Scott-Ram N, Penfold H, Creswell C. Navigating the challenges of digital health innovation: considerations and solutions in developing online and smartphone-application-based interventions for mental health disorders. *Br J Psychiatry* 2017 Aug;211(2):65-69. [doi: [10.1192/bjp.bp.115.180372](#)] [Medline: [28522435](#)]
7. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. *J Anxiety Disord* 2018 Apr;55:70-78 [[FREE Full text](#)] [doi: [10.1016/j.janxdis.2018.01.001](#)] [Medline: [29422409](#)]
8. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. *BMC Med* 2015 Sep 7;13:214 [[FREE Full text](#)] [doi: [10.1186/s12916-015-0444-y](#)] [Medline: [26404673](#)]
9. Leigh S, Flatt S. App-based psychological interventions: friend or foe? *Evid Based Ment Health* 2015 Nov;18(4):97-99. [doi: [10.1136/eb-2015-102203](#)] [Medline: [26459466](#)]
10. Bennion MR, Hardy G, Moore RK, Millings A. E-therapies in England for stress, anxiety or depression: what is being used in the NHS? a survey of mental health services. *BMJ Open* 2017 Jan 23;7(1):e014844. [doi: [10.1136/bmjopen-2016-014844](#)] [Medline: [28115336](#)]
11. Bennion M, Hardy G, Moore R, Kellett S, Millings A. e-Therapies in England for stress, anxiety or depression: how are apps developed? a survey of NHS e-therapy developers. *BMJ Health Care Inform* 2019 Jun;26(1):e100027. [doi: [10.1136/bmjhci-2019-100027](#)] [Medline: [31171556](#)]
12. Cowie MR. National institute for health and care excellence. *Eur Heart J* 2015 Jan 21;36(4):195. [Medline: [25756143](#)]
13. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. *J Am Med Assoc Psychiatry* 2017 Apr 1;74(4):351-359. [doi: [10.1001/jamapsychiatry.2017.0044](#)] [Medline: [28241179](#)]
14. Bell CJ, Colhoun HC, Carter FA, Frampton CM. Effectiveness of computerised cognitive behaviour therapy for anxiety disorders in secondary care. *Aust N Z J Psychiatry* 2012 Jul;46(7):630-640. [doi: [10.1177/0004867412437345](#)] [Medline: [22327097](#)]
15. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the gold standard: lessons from the history of RCTs. *N Engl J Med* 2016 Jun 2;374(22):2175-2181. [doi: [10.1056/NEJMms1604593](#)] [Medline: [27248626](#)]
16. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. *Psychol Med* 2009 May;39(5):705-712. [doi: [10.1017/S0033291708004224](#)] [Medline: [18812006](#)]
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009 Aug 18;151(4):264-9, W64. [doi: [10.7326/0003-4819-151-4-200908180-00135](#)] [Medline: [19622511](#)]
18. Bennion M. An Investigation Into E-therapies for Older Adults. White Rose eTheses Online. 2019. URL: <http://etheses.whiterose.ac.uk/23347/> [accessed 2020-09-28]
19. Newman M, Szkodny L, Llera S, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? *Clin Psychol Rev* 2011 Feb;31(1):89-103. [doi: [10.1016/j.cpr.2010.09.008](#)] [Medline: [21130939](#)]
20. Cochrane Handbook for Systematic Reviews of Interventions. New York, USA: Cochrane; 2011.

21. RevMan. Cochrane Training. 2014. URL: <https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman> [accessed 2020-09-28]
22. Cohen J. A Coefficient of Agreement for Nominal Scales. Sage Journals: Your Gateway to World-class Research Journals. 1960. URL: <https://doi.org/10.1177/001316446002000104> [accessed 2018-03-23]
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977 Mar;33(1):159-174. [Medline: [843571](#)]
24. Dijkers M. Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development. *KT Update* Aug 2013;1(5).
25. Hedges LV. *Statistical Methods Meta-analysis*. Orlando, FL: Academic Press; 1985.
26. Cohen J. A power primer. *Psychol Bull* 1992 Jul;112(1):155-159. [doi: [10.1037/0033-2909.112.1.155](https://doi.org/10.1037/0033-2909.112.1.155)] [Medline: [19565683](#)]
27. Suurmond R, van Rhee H, Hak T. Introduction, comparison, and validation of meta-essentials: a free and simple tool for meta-analysis. *Res Synth Methods* 2017 Dec;8(4):537-553 [FREE Full text] [doi: [10.1002/jrsm.1260](https://doi.org/10.1002/jrsm.1260)] [Medline: [28801932](#)]
28. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Br Med J* 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: [10.1136/bmj.327.7414.557](https://doi.org/10.1136/bmj.327.7414.557)] [Medline: [12958120](#)]
29. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry* 2006 Jun 1;59(11):990-996. [doi: [10.1016/j.biopsych.2005.09.014](https://doi.org/10.1016/j.biopsych.2005.09.014)] [Medline: [16368078](#)]
30. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000 Jun;56(2):455-463. [doi: [10.1111/j.0006-341x.2000.00455.x](https://doi.org/10.1111/j.0006-341x.2000.00455.x)] [Medline: [10877304](#)]
31. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: [10.1136/bmj.315.7109.629](https://doi.org/10.1136/bmj.315.7109.629)] [Medline: [9310563](#)]
32. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. *Psychol Med* 2003 Feb;33(2):217-227. [doi: [10.1017/s0033291702007225](https://doi.org/10.1017/s0033291702007225)] [Medline: [12622301](#)]
33. Grime PR. Computerized cognitive behavioural therapy at work: a randomized controlled trial in employees with recent stress-related absenteeism. *Occup Med (Lond)* 2004 Aug;54(5):353-359. [doi: [10.1093/occmed/kqh077](https://doi.org/10.1093/occmed/kqh077)] [Medline: [15289593](#)]
34. Proudfoot J, Ryden C, Everitt B, Shapiro DA, Goldberg D, Mann A, et al. Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. *Br J Psychiatry* 2004 Jul;185:46-54. [doi: [10.1192/bjp.185.1.46](https://doi.org/10.1192/bjp.185.1.46)] [Medline: [15231555](#)]
35. Marks IM, Kenwright M, McDonough M, Whittaker M, Mataix-Cols D. Saving clinicians' time by delegating routine aspects of therapy to a computer: a randomized controlled trial in phobia/panic disorder. *Psychol Med* 2004 Jan;34(1):9-17. [doi: [10.1017/s003329170300878x](https://doi.org/10.1017/s003329170300878x)] [Medline: [14971623](#)]
36. Schneider AJ, Mataix-Cols D, Marks IM, Bachofen M. Internet-guided self-help with or without exposure therapy for phobic and panic disorders. *Psychother Psychosom* 2005;74(3):154-164. [doi: [10.1159/000084000](https://doi.org/10.1159/000084000)] [Medline: [15832066](#)]
37. Mackinnon A, Griffiths KM, Christensen H. Comparative randomised trial of online cognitive-behavioural therapy and an information website for depression: 12-month outcomes. *Br J Psychiatry* 2008 Feb;192(2):130-134. [doi: [10.1192/bjp.bp.106.032078](https://doi.org/10.1192/bjp.bp.106.032078)] [Medline: [18245031](#)]
38. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered Internet psychotherapy for depression in primary care: a randomised controlled trial. *Lancet* 2009 Aug 22;374(9690):628-634. [doi: [10.1016/S0140-6736\(09\)61257-5](https://doi.org/10.1016/S0140-6736(09)61257-5)] [Medline: [19700005](#)]
39. Ellis L, Campbell A, Sethi S, O'Dea B. Comparative randomized trial of an online cognitive-behavioral therapy program and an online support group for depression and anxiety. *J Cybertherapy Rehabil* 2011;4(4):467.
40. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. *PLoS One* 2011;6(11):e28099 [FREE Full text] [doi: [10.1371/journal.pone.0028099](https://doi.org/10.1371/journal.pone.0028099)] [Medline: [22140514](#)]
41. Høifødt RS, Lillevoll KR, Griffiths KM, Wilsaard T, Eisemann M, Waterloo K, et al. The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: randomized controlled trial. *J Med Internet Res* 2013 Aug 5;15(8):e153 [FREE Full text] [doi: [10.2196/jmir.2714](https://doi.org/10.2196/jmir.2714)] [Medline: [23916965](#)]
42. Lintvedt OK, Griffiths KM, Sørensen K, Østvik AR, Wang CE, Eisemann M, et al. Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: a randomized controlled trial. *Clin Psychol Psychother* 2013;20(1):10-27. [doi: [10.1002/cpp.770](https://doi.org/10.1002/cpp.770)] [Medline: [21887811](#)]
43. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. *J Med Internet Res* 2012 Dec 31;15(1):e2 [FREE Full text] [doi: [10.2196/jmir.2240](https://doi.org/10.2196/jmir.2240)] [Medline: [23302475](#)]
44. Sethi S. Treating Youth Depression and Anxiety: A Randomised Controlled Trial Examining the Efficacy of Computerised versus Face-to-face Cognitive Behaviour Therapy. *Australian Psychologist* 2013 Feb 14;48(4):249-257. [doi: [10.1111/ap.12006](https://doi.org/10.1111/ap.12006)]
45. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in Happiness: A Randomised Controlled Trial of a Smartphone-Based Mindfulness Intervention to Enhance Wellbeing. *J Happiness Stud* 2014 Oct 29;17(1):163-185. [doi: [10.1007/s10902-014-9589-1](https://doi.org/10.1007/s10902-014-9589-1)]

46. Phillips R, Schneider J, Molosankwe I, Leese M, Foroushani PS, Grime P, et al. Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: effectiveness and costs of a workplace intervention. *Psychol Med* 2014 Mar;44(4):741-752 [FREE Full text] [doi: [10.1017/S0033291713001323](https://doi.org/10.1017/S0033291713001323)] [Medline: [23795621](https://pubmed.ncbi.nlm.nih.gov/23795621/)]
47. Twomey C, O'Reilly G, Byrne M, Bury M, White A, Kissane S, et al. A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions. *Br J Clin Psychol* 2014 Nov;53(4):433-450. [doi: [10.1111/bjc.12055](https://doi.org/10.1111/bjc.12055)] [Medline: [24831119](https://pubmed.ncbi.nlm.nih.gov/24831119/)]
48. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. *Br Med J* 2015 Nov 11;351:h5627. [doi: [10.1136/bmj.h5627](https://doi.org/10.1136/bmj.h5627)] [Medline: [26559241](https://pubmed.ncbi.nlm.nih.gov/26559241/)]
49. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression. *Behav Res Ther* 2015 Dec;75:20-31. [doi: [10.1016/j.brat.2015.10.005](https://doi.org/10.1016/j.brat.2015.10.005)] [Medline: [26523885](https://pubmed.ncbi.nlm.nih.gov/26523885/)]
50. Richards D, Timulak L, Rashleigh C, McLoughlin O, Colla A, Joyce C, et al. Effectiveness of an internet-delivered intervention for generalized anxiety disorder in routine care: a randomised controlled trial in a student population. *Internet Interv* 2016 Nov;6:80-88 [FREE Full text] [doi: [10.1016/j.invent.2016.10.003](https://doi.org/10.1016/j.invent.2016.10.003)] [Medline: [30135817](https://pubmed.ncbi.nlm.nih.gov/30135817/)]
51. Carolan S, Harris PR, Greenwood K, Cavanagh K. Increasing engagement with an occupational digital stress management program through the use of an online facilitated discussion group: results of a pilot randomised controlled trial. *Internet Interv* 2017 Dec;10:1-11 [FREE Full text] [doi: [10.1016/j.invent.2017.08.001](https://doi.org/10.1016/j.invent.2017.08.001)] [Medline: [30135747](https://pubmed.ncbi.nlm.nih.gov/30135747/)]
52. Flett JAM, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile Mindfulness Meditation: a Randomised Controlled Trial of the Effect of Two Popular Apps on Mental Health. *Mindfulness* 2018 Oct 31;10(5):863-876. [doi: [10.1007/s12671-018-1050-9](https://doi.org/10.1007/s12671-018-1050-9)]
53. Forand NR, Barnett JG, Strunk DR, Hindiyeh MU, Feinberg JE, Keefe JR. Efficacy of guided iCBT for depression and mediation of change by cognitive skill acquisition. *Behav Ther* 2018 Mar;49(2):295-307 [FREE Full text] [doi: [10.1016/j.beth.2017.04.004](https://doi.org/10.1016/j.beth.2017.04.004)] [Medline: [29530267](https://pubmed.ncbi.nlm.nih.gov/29530267/)]
54. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. *J Occup Health Psychol* 2019 Feb;24(1):127-138 [FREE Full text] [doi: [10.1037/ocp0000118](https://doi.org/10.1037/ocp0000118)] [Medline: [29723001](https://pubmed.ncbi.nlm.nih.gov/29723001/)]
55. Löbner M, Pabst A, Stein J, Dorow M, Matschinger H, Lupp M, et al. Computerized cognitive behavior therapy for patients with mild to moderately severe depression in primary care: a pragmatic cluster randomized controlled trial (@ktiv). *J Affect Disord* 2018 Oct 1;238:317-326. [doi: [10.1016/j.jad.2018.06.008](https://doi.org/10.1016/j.jad.2018.06.008)] [Medline: [29902736](https://pubmed.ncbi.nlm.nih.gov/29902736/)]
56. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. *J Med Internet Res* 2011 Dec 31;13(4):e126 [FREE Full text] [doi: [10.2196/jmir.1923](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]
57. Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CA. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. *J Med Internet Res* 2010 Dec 16;12(4):e68 [FREE Full text] [doi: [10.2196/jmir.1642](https://doi.org/10.2196/jmir.1642)] [Medline: [21163776](https://pubmed.ncbi.nlm.nih.gov/21163776/)]
58. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. *Br J Clin Psychol* 2010 Nov;49(Pt 4):455-471. [doi: [10.1348/014466509X472138](https://doi.org/10.1348/014466509X472138)] [Medline: [19799804](https://pubmed.ncbi.nlm.nih.gov/19799804/)]
59. Hind D, O'Cathain A, Cooper CL, Parry GD, Isaac CL, Rose A, et al. The acceptability of computerised cognitive behavioural therapy for the treatment of depression in people with chronic physical disease: a qualitative study of people with multiple sclerosis. *Psychol Health* 2010 Jul;25(6):699-712. [doi: [10.1080/08870440902842739](https://doi.org/10.1080/08870440902842739)] [Medline: [20204955](https://pubmed.ncbi.nlm.nih.gov/20204955/)]
60. Currie M, Philip LJ, Roberts A. Attitudes towards the use and acceptance of eHealth technologies: a case study of older adults living with chronic pain and implications for rural healthcare. *BMC Health Serv Res* 2015 Apr 16;15:162 [FREE Full text] [doi: [10.1186/s12913-015-0825-0](https://doi.org/10.1186/s12913-015-0825-0)] [Medline: [25888988](https://pubmed.ncbi.nlm.nih.gov/25888988/)]
61. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. Usability, acceptability, and effectiveness of web-based conversational agents to facilitate problem solving in older adults: controlled study. *J Med Internet Res* 2020 May 27;22(5):e16794 [FREE Full text] [doi: [10.2196/16794](https://doi.org/10.2196/16794)] [Medline: [32384055](https://pubmed.ncbi.nlm.nih.gov/32384055/)]
62. Wozney L, Huguet A, Bennett K, Radomski AD, Hartling L, Dyson M, et al. How do ehealth programs for adolescents with depression work: a realist review of persuasive system design components in internet-based psychological therapies. *J Med Internet Res* 2017 Aug 9;19(8):e266 [FREE Full text] [doi: [10.2196/jmir.7573](https://doi.org/10.2196/jmir.7573)] [Medline: [28793983](https://pubmed.ncbi.nlm.nih.gov/28793983/)]
63. Smart A. A multi-dimensional model of clinical utility. *Int J Qual Health Care* 2006 Oct;18(5):377-382. [doi: [10.1093/intqhc/mzl034](https://doi.org/10.1093/intqhc/mzl034)] [Medline: [16951425](https://pubmed.ncbi.nlm.nih.gov/16951425/)]
64. The Improving Access to Psychological Therapies Manual. NHS England. 2020. URL: <https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-manual/> [accessed 2020-09-28]
65. Buhrman M, Skoglund A, Husell J, Bergström K, Gordh T, Hursti T, et al. Guided internet-delivered acceptance and commitment therapy for chronic pain patients: a randomized controlled trial. *Behav Res Ther* 2013 Jun;51(6):307-315. [doi: [10.1016/j.brat.2013.02.010](https://doi.org/10.1016/j.brat.2013.02.010)] [Medline: [23548250](https://pubmed.ncbi.nlm.nih.gov/23548250/)]



66. Dagö J, Asplund RP, Bsenko HA, Hjerling S, Holmberg A, Westh S, et al. Cognitive behavior therapy versus interpersonal psychotherapy for social anxiety disorder delivered via smartphone and computer: a randomized controlled trial. *J Anxiety Disord* 2014 May;28(4):410-417. [doi: [10.1016/j.janxdis.2014.02.003](https://doi.org/10.1016/j.janxdis.2014.02.003)] [Medline: [24731441](https://pubmed.ncbi.nlm.nih.gov/24731441/)]
67. Boettcher J, Aström V, Pålsson D, Schenström O, Andersson G, Carlbring P. Internet-based mindfulness treatment for anxiety disorders: a randomized controlled trial. *Behav Ther* 2014 Mar;45(2):241-253 [FREE Full text] [doi: [10.1016/j.beth.2013.11.003](https://doi.org/10.1016/j.beth.2013.11.003)] [Medline: [24491199](https://pubmed.ncbi.nlm.nih.gov/24491199/)]
68. Andersson G, Paxling B, Roch-Norlund P, Östman G, Norgren A, Almlöv J, et al. Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: a randomized controlled trial. *Psychother Psychosom* 2012;81(6):344-355 [FREE Full text] [doi: [10.1159/000339371](https://doi.org/10.1159/000339371)] [Medline: [22964540](https://pubmed.ncbi.nlm.nih.gov/22964540/)]
69. Cuijpers P, van Straten A, Bohlmeijer E, Hollon S, Andersson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol Med* 2010 Feb;40(2):211-223. [doi: [10.1017/S0033291709006114](https://doi.org/10.1017/S0033291709006114)] [Medline: [19490745](https://pubmed.ncbi.nlm.nih.gov/19490745/)]
70. Spek V, Nyklíček I, Cuijpers P, Pop V. Predictors of outcome of group and internet-based cognitive behavior therapy. *J Affect Disord* 2008 Jan;105(1-3):137-145. [doi: [10.1016/j.jad.2007.05.001](https://doi.org/10.1016/j.jad.2007.05.001)] [Medline: [17543392](https://pubmed.ncbi.nlm.nih.gov/17543392/)]
71. Spek V, Nyklíček I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. *Psychol Med* 2007 Dec;37(12):1797-1806. [doi: [10.1017/S0033291707000542](https://doi.org/10.1017/S0033291707000542)] [Medline: [17466110](https://pubmed.ncbi.nlm.nih.gov/17466110/)]
72. Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? *Expert Rev Pharmacoecon Outcomes Res* 2007 Jun;7(3):291-297. [doi: [10.1586/14737167.7.3.291](https://doi.org/10.1586/14737167.7.3.291)] [Medline: [20528315](https://pubmed.ncbi.nlm.nih.gov/20528315/)]
73. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. *Cogn Behav Ther* 2009;38(2):66-82. [doi: [10.1080/16506070802694776](https://doi.org/10.1080/16506070802694776)] [Medline: [20183688](https://pubmed.ncbi.nlm.nih.gov/20183688/)]
74. Richards DA, Bower P, Pagel C, Weaver A, Utley M, Cape J, et al. Delivering stepped care: an analysis of implementation in routine practice. *Implement Sci* 2012 Jan 16;7:3 [FREE Full text] [doi: [10.1186/1748-5908-7-3](https://doi.org/10.1186/1748-5908-7-3)] [Medline: [22248385](https://pubmed.ncbi.nlm.nih.gov/22248385/)]
75. Russo MW. How to review a meta-analysis. *Gastroenterol Hepatol (N Y)* 2007 Aug;3(8):637-642 [FREE Full text] [Medline: [21960873](https://pubmed.ncbi.nlm.nih.gov/21960873/)]
76. Bush NE, Smolenski DJ, Denneson LM, Williams HB, Thomas EK, Dobscha SK. A virtual hope box: randomized controlled trial of a smartphone app for emotional regulation and coping with distress. *Psychiatr Serv* 2017 Apr 1;68(4):330-336. [doi: [10.1176/appi.ps.201600283](https://doi.org/10.1176/appi.ps.201600283)] [Medline: [27842473](https://pubmed.ncbi.nlm.nih.gov/27842473/)]
77. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. *NPJ Digit Med* 2020;3:85 [FREE Full text] [doi: [10.1038/s41746-020-0293-8](https://doi.org/10.1038/s41746-020-0293-8)] [Medline: [32566763](https://pubmed.ncbi.nlm.nih.gov/32566763/)]
78. Howell AN, Rheingold AA, Uhde TW, Guille C. Web-based CBT for the prevention of anxiety symptoms among medical and health science graduate students. *Cogn Behav Ther* 2019 Sep;48(5):385-405 [FREE Full text] [doi: [10.1080/16506073.2018.1533575](https://doi.org/10.1080/16506073.2018.1533575)] [Medline: [30526364](https://pubmed.ncbi.nlm.nih.gov/30526364/)]
79. Flett JA, Conner TS, Riordan BC, Patterson T, Hayne H. App-based mindfulness meditation for psychological distress and adjustment to college in incoming university students: a pragmatic, randomised, waitlist-controlled trial. *Psychol Health* 2020 Sep;35(9):1049-1074. [doi: [10.1080/08870446.2019.1711089](https://doi.org/10.1080/08870446.2019.1711089)] [Medline: [32046499](https://pubmed.ncbi.nlm.nih.gov/32046499/)]
80. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. *J Med Internet Res Mhealth Uhealth* 2019 Jun 25;7(6):e14273. [doi: [10.2196/14273](https://doi.org/10.2196/14273)] [Medline: [31237569](https://pubmed.ncbi.nlm.nih.gov/31237569/)]
81. Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a placebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr Scand* 2014 Sep;130(3):181-192. [doi: [10.1111/acps.12275](https://doi.org/10.1111/acps.12275)] [Medline: [24697518](https://pubmed.ncbi.nlm.nih.gov/24697518/)]
82. Simon N, McGillivray L, Roberts NP, Barawi K, Lewis CE, Bisson JI. Acceptability of internet-based cognitive behavioural therapy (i-CBT) for post-traumatic stress disorder (PTSD): a systematic review. *Eur J Psychotraumatol* 2019;10(1):1646092 [FREE Full text] [doi: [10.1080/20008198.2019.1646092](https://doi.org/10.1080/20008198.2019.1646092)] [Medline: [31497259](https://pubmed.ncbi.nlm.nih.gov/31497259/)]
83. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *Br Med J* 2019 Aug 28;366:l4898. [doi: [10.1136/bmj.l4898](https://doi.org/10.1136/bmj.l4898)] [Medline: [31462531](https://pubmed.ncbi.nlm.nih.gov/31462531/)]
84. Firth N, Barkham M, Kellert S. The clinical effectiveness of stepped care systems for depression in working age adults: a systematic review. *J Affect Disord* 2015 Jan 01;170:119-130. [doi: [10.1016/j.jad.2014.08.030](https://doi.org/10.1016/j.jad.2014.08.030)] [Medline: [25240141](https://pubmed.ncbi.nlm.nih.gov/25240141/)]

## Abbreviations

**GP:** general practitioner

**IAPT:** Improving Access to Psychological Therapies

**ITT:** intention to treat

**LI:** low intensity



**NHS:** National Health Service

**NICE:** National Institute for Health and Care Excellence

**NNT:** number needed to treat

**RCT:** randomized controlled trial

**SMD:** standardized mean difference

**TAU:** treatment as usual

*Edited by G Eysenbach; submitted 13.11.19; peer-reviewed by L Donkin, P Romero-Sanchiz, R Nogueira-Arjona, C Christ, K Kaipainen; comments to author 23.03.20; revised version received 18.05.20; accepted 24.07.20; published 28.10.20.*

*Please cite as:*

*Simmonds-Buckley M, Bennion MR, Kellett S, Millings A, Hardy GE, Moore RK*

*Acceptability and Effectiveness of NHS-Recommended e-Therapies for Depression, Anxiety, and Stress: Meta-Analysis*

*J Med Internet Res 2020;22(10):e17049*

URL: <http://www.jmir.org/2020/10/e17049/>

doi: [10.2196/17049](https://doi.org/10.2196/17049)

PMID: [33112238](https://pubmed.ncbi.nlm.nih.gov/33112238/)

©Melanie Simmonds-Buckley, Matthew Russell Bennion, Stephen Kellett, Abigail Millings, Gillian E Hardy, Roger K Moore. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Disaster eHealth: Scoping Review

Samaneh Madanian<sup>1\*</sup>, PhD; Tony Norris<sup>1\*</sup>, PhD; Dave Parry<sup>1\*</sup>, PhD

Auckland University of Technology (AUT), Auckland, New Zealand

\* all authors contributed equally

**Corresponding Author:**

Samaneh Madanian, PhD

Auckland University of Technology

Private Bag 92006

Auckland, 1142

New Zealand

Phone: 64 99219999 ext 6539

Email: [sam.madanian@aut.ac.nz](mailto:sam.madanian@aut.ac.nz)

## Abstract

**Background:** Although both disaster management and disaster medicine have been used for decades, their efficiency and effectiveness have been far from perfect. One reason could be the lack of systematic utilization of modern technologies, such as eHealth, in their operations. To address this issue, researchers' efforts have led to the emergence of the disaster eHealth (DEH) field. DEH's main objective is to systematically integrate eHealth technologies for health care purposes within the disaster management cycle (DMC).

**Objective:** This study aims to identify, map, and define the scope of DEH as a new area of research at the intersection of disaster management, emergency medicine, and eHealth.

**Methods:** An extensive scoping review using published materials was carried out in the areas of disaster management, disaster medicine, and eHealth to identify the scope of DEH. This review procedure was iterative and conducted in multiple scientific databases in 2 rounds, one using controlled indexed terms and the other using similar uncontrolled terms. In both rounds, the publications ranged from 1990 to 2016, and all the appropriate research studies discovered were considered, regardless of their research design, methodology, and quality. Information extracted from both rounds was thematically analyzed to define the DEH scope, and the results were evaluated by the field experts through a Delphi method.

**Results:** In both rounds of the research, searching for eHealth applications within DMC yielded 404 relevant studies that showed eHealth applications in different disaster types and disaster phases. These applications varied with respect to the eHealth technology types, functions, services, and stakeholders. The results led to the identification of the scope of DEH, including eHealth technologies and their applications, services, and future developments that are applicable to disasters as well as to related stakeholders. Reference to the elements of the DEH scope indicates what, when, and how current eHealth technologies can be used in the DMC.

**Conclusions:** Comprehensive data gathering from multiple databases offered a grounded method to define the DEH scope. This scope comprises concepts related to DEH and the boundaries that define it. The scope identifies the eHealth technologies relevant to DEH and the functions and services that can be provided by these technologies. In addition, the scope tells us which groups can use the provided services and functions and in which disaster types or phases. DEH approaches could potentially improve the response to health care demands before, during, and after disasters. DEH takes advantage of eHealth technologies to facilitate DMC tasks and activities, enhance their efficiency and effectiveness, and enhance health care delivery and provide more quality health care services to the wider population regardless of their geographical location or even disaster types and phases.

(*J Med Internet Res* 2020;22(10):e18310) doi:[10.2196/18310](https://doi.org/10.2196/18310)

## KEYWORDS

medical informatics; disaster planning; disaster medicine; medical informatics applications; disasters

## Introduction

Disasters are destructive events that threaten public health and the environment and disrupt and/or impede normal operations.

They also impose considerable pressure on health care systems. The source of disasters can be natural or the result of human actions (eg, fires and terrorist attacks) [1]. Disease epidemics can also be seen as disasters, albeit over a longer time scale

rather than a point event, killing thousands of people or making them homeless, destroying health infrastructure, and disrupting public and commercial services. Data from the Centre for Research on the Epidemiology of Disasters (CRED) [2] show that the severity and complexity (damage to life and property) of disasters have grown exponentially over recent decades. In 2018, 331 natural disasters were identified; these caused 14,854 deaths, affected a further 81,143,283 and cost US \$130,655,327,000.

Disaster management and disaster medicine are complementary disciplines that can significantly reduce the harmful effects of disasters. Disaster management conveniently encompasses 4 phases: mitigation, preparedness, response, and recovery [3]. Addressing issues arising in these phases requires good collaboration among different governmental and nongovernmental organizations and disaster medicine groups. However, both disaster management and disaster medicine operations are frequently far from perfect and have a long list of failures [4]. In some cases, disaster medicine groups have poor communication with disaster management organizations; in this regard [5], claims *emergency management and the health sectors are natural allies that have, seemingly, only recently begun to recognize each other*. According to the reviewed literature, some major contributions to this issue are as follows:

1. Disaster management and disaster medicine have different roots, development, and priorities [6]. Therefore, coordination and communication between them in disasters are often missing, which leads to delayed, substandard, improper, or sometimes no care.
2. Although both areas emerged to work side by side, they sometimes fail to share their tools and personnel and have not collaborated smoothly in preparing for and responding to mass emergencies.
3. Neither disaster medicine nor disaster management routinely uses information or modern eHealth technologies [6].

Therefore, there is a pressing need for efficient disaster management and emergency medicine to mitigate human pain and suffering and the overall impact of disasters [1]. These issues are the reasons that renewed interest in this field has appeared and necessitates more research.

Despite the growth of information technology capacities and services (eg, new communication technologies, ubiquitous computing, the internet, and advanced smart devices [7]) in mainstream health care, their applications in disaster management and disaster medicine fields are currently limited. However, they offer great potential for improving efficiency and effectiveness in disaster situations, especially when eHealth technologies are integrated and applied systematically throughout the disaster management cycle (DMC) [7] to improve disaster health planning before, response, during, and recovery after disasters. To achieve this goal, the domain of disaster eHealth (DEH) has been proposed in the study by Norris et al on 2015 [6] at the intersection of disaster management, disaster medicine, and eHealth fields (Figure 1). The definitions of these terms are provided in Textbox 1.

DEH is an emerging field that was introduced earlier in the study by Norris et al [6]. In that research, besides the definition of DEH, its vision was defined. However, as DEH has not yet been generally acknowledged, profiling and defining the field and its scope and assessing the role of eHealth technologies in each phase of disasters are important. Therefore, the aim of this study is to define the scope of DEH rather than providing detailed analysis and reviewing the literature related to the field of DEH.

DEH can be seen as a model telling us what, when, and how current eHealth technologies can be used in the DMC. These technologies include not only those used in established eHealth practices but also those recently made available by the rapid development in mobile and sensor technologies.

**Figure 1.** Disaster eHealth and its components.



**Textbox 1.** Disaster eHealth and the definition of its components.

- Disaster eHealth: the application of information and eHealth technologies in a disaster situation to restore and maintain the health of individuals to their predisaster levels
- Disaster management: the coordination and integration of all activities necessary to build, sustain, and improve the capabilities to prepare for, respond to, recover from, or mitigate against threatened or actual disasters or emergencies, regardless of cause [8]
- Disaster medicine: a system of study and medical practice associated primarily with the disciplines of emergency medicine and public health [9]
- eHealth: the cost-effective and secure use of information and communications technology in support of health and health-related fields, including health care services, health surveillance, health literature, and health education, knowledge, and research [10]

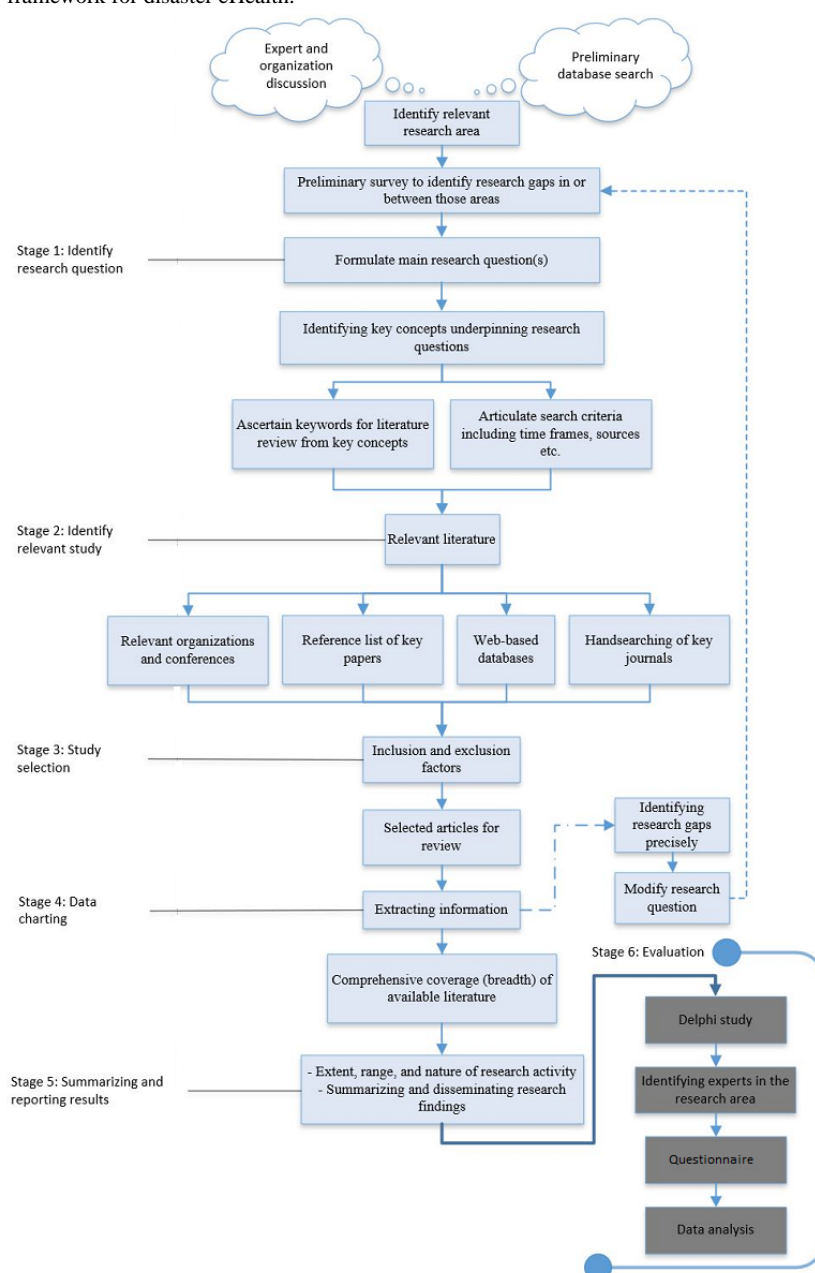
## Methods

### Study Design

To undertake this research, a rigorous scoping study was conducted based on the framework of Arksey and O'Malley

[11]. Our adapted framework included an additional stage to evaluate the research results, as suggested by Levac et al [12]. The adapted framework is presented in Figure 2.

**Figure 2.** The scoping study framework for disaster eHealth.



This research is limited to publications from 1990 to 2016. This scoping study was undertaken in 2 complementary rounds: uncontrolled and controlled search. The uncontrolled search was commenced in multiple databases using free-text terms rather than indexed terms. This approach uses a search engine to identify documents of interest based on terms occurring in the papers' titles, abstracts, or main bodies. This allowed us to extensively and fully explore the area and extract a broad range of articles to define the scope of DEH. However, to improve

the accuracy of free-text search and to decrease the potential searching bias or missing data in the search, a controlled search was employed [13]. In this method, keywords were selected from an existing list of index terms on which articles are indexed. The method is used in particular databases, such as MeSH in PubMed and Controlled Indexing in IEEE Xplore. A sample list of searched terms and queries in both research procedures is provided in Table 1.

**Table 1.** A sample of search terms and queries.

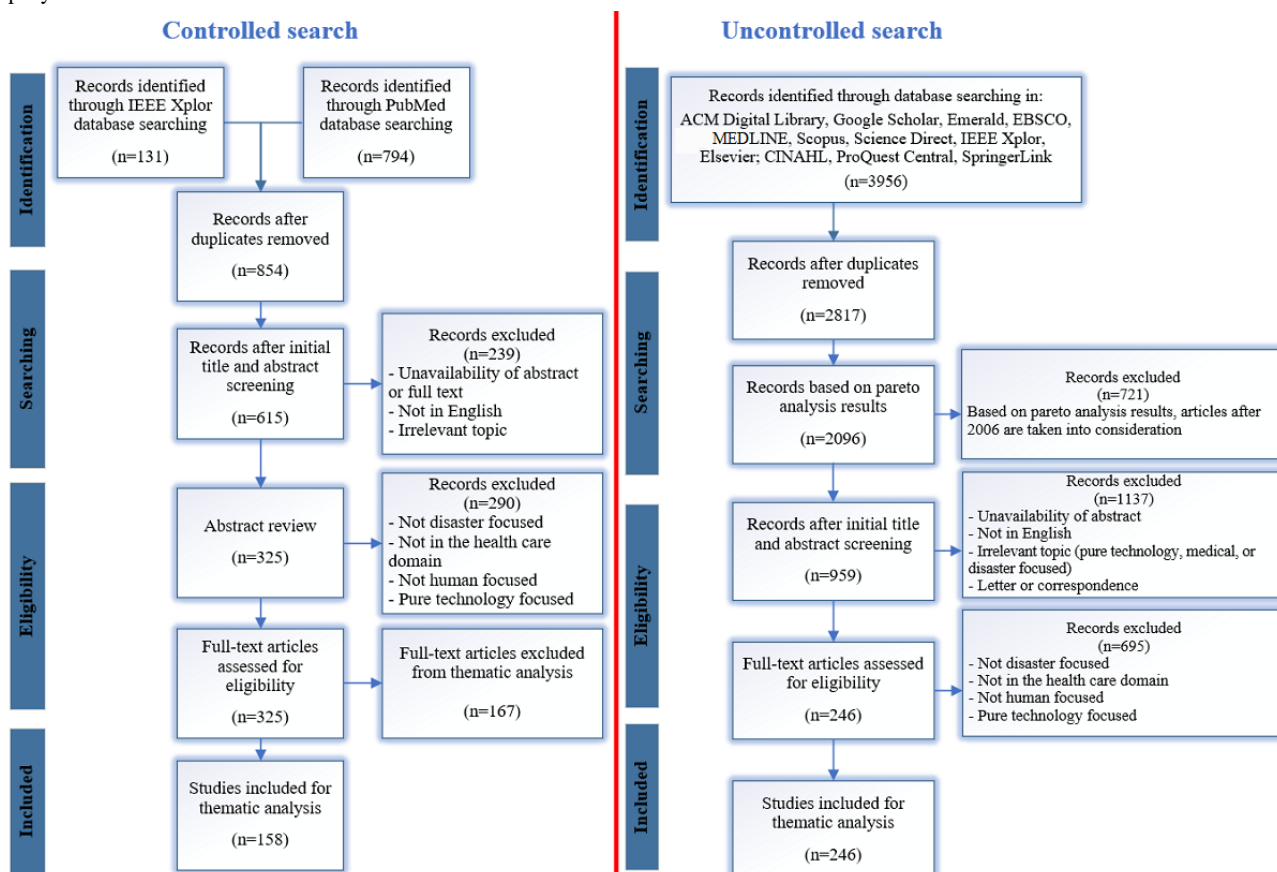
Searching terms	Uncontrolled search	Controlled search
Search terms	<ul style="list-style-type: none"> <li>disaster management</li> <li>disaster</li> <li>disaster medicine</li> <li>ehealth</li> <li>e*health</li> </ul>	<ul style="list-style-type: none"> <li>disaster</li> <li>medical informatics</li> <li>disaster medicine</li> <li>emergency management</li> </ul>
Search queries	<ul style="list-style-type: none"> <li>(ehealth OR e-health OR e*health) AND disaster</li> <li>(ehealth OR e-health OR e*health) AND "disaster medicine"</li> </ul>	<ul style="list-style-type: none"> <li>("Disasters"[Mesh]) AND "Medical Informatics"[Mesh])</li> <li>"INSPEC Controlled Terms": emergency management AND "INSPEC Controlled Terms":disaster</li> </ul>

## Inclusion and Exclusion Criteria

Both search procedures were iterative and captured relevant articles regardless of their research design, methodology, and quality (recommended by Valaitis et al [14]). The selected articles were in the English language and explicitly referred to

eHealth technologies and their applications in disaster management and/or disaster medicine. The examined databases, the overall flow of study selection, and inclusion and exclusion criteria for the uncontrolled and controlled search are depicted in Figure 3.

**Figure 3.** Controlled and uncontrolled searching steps. CINAHL: Cumulative Index to Nursing and Allied Health Literature; EBSCO: Elton B Stephens Company.



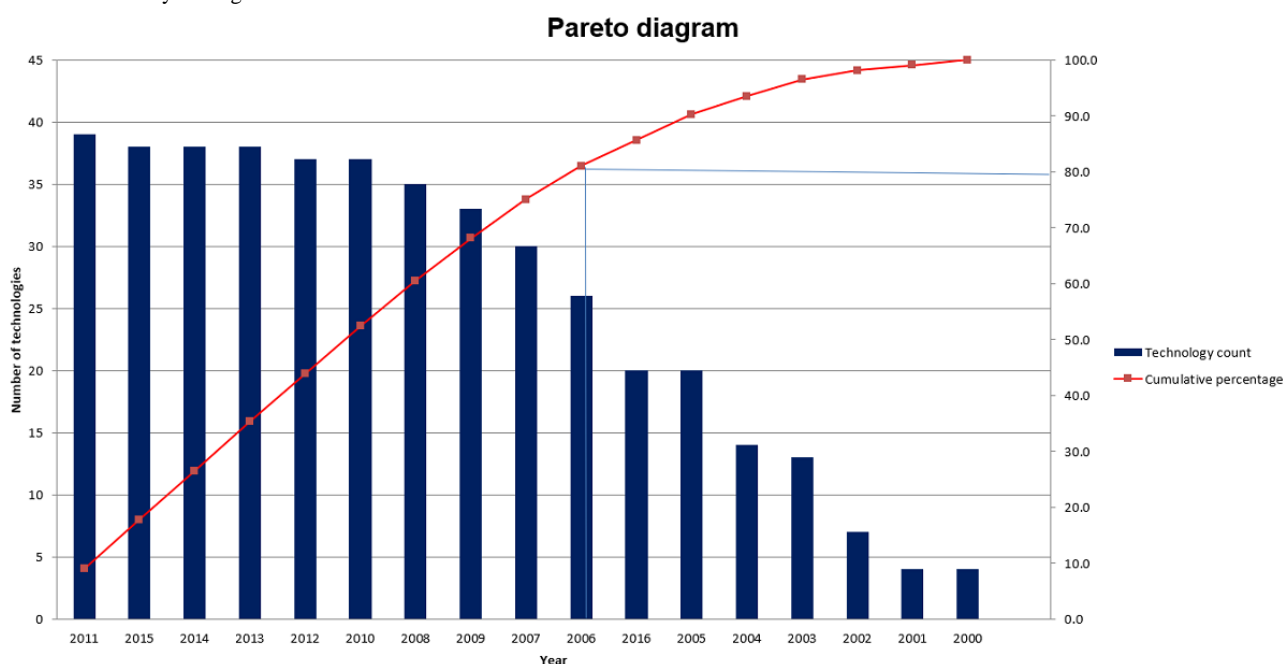


In the uncontrolled search to facilitate the preliminary eligibility examination phase, Pareto analysis was used to exclude a larger number of articles in a shorter time without affecting the quality of the results. Pareto analysis is a well-established statistical technique in the business and management field known as 80/20 rule, that is, 20% of the major tasks and activities can generate 80% of the benefit of doing the entire job [15]. This method helped the researcher, instead of dealing with all the studies, to focus on those that had the most significant impact on the

research output. Pareto analysis was performed based on the number of technology types used in DMC for health care purposes in each year, called *technology saturation*. The rationale for technology selection in Pareto analysis is rooted in the rapid technological revolution or advancement in the field of computation.

By following the Pareto formula [15,16], only articles after 2006 were included for review. The Pareto analysis diagram is depicted in Figure 4.

**Figure 4.** Pareto analysis diagram in uncontrolled search.



## Data Analysis Method

After selecting the studies for the in-depth review, in controlled and uncontrolled rounds, their full text was added to EndNote (Web of Science Group) [17] and NVivo (QSR International) [18] databases for qualitative analysis, summarizing, and drawing conclusions. Thematic analysis [19] and content analysis [20] were used to interpret different aspects of each study and identify the main themes. The analysis mainly intended to identify the important features related to this research and categorizing the findings into different thematic groups. These themes were inductively derived from the data for which we first familiarized ourselves with the data sample by reading a number of articles. Each created theme consisted of information that concentrated on or covered a particular aspect related to eHealth in disaster management or disaster medicine. The themes included potential technologies and their functions, attributes, stakeholders, and other related concepts.

On each theme, a conventional content analysis was performed to interpret the findings of the theme. The analyses identify the opinions and general trends in eHealth adoptions and applications within disaster management and disaster medicine fields. Finally, the Delphi method was conducted in 2 rounds in which the fields 'experts evaluated the initial DEH scope. The results of this evaluation are reflected in the reported DEH scope in this paper.

## Results

### Disaster Types and Phases in DEH

Identification of disaster type is necessary to select appropriate approaches to DEH for particular cases. The research results highlighted a high diversity in the literature with regard to disaster types. A comprehensive list of disaster types was extracted based on the CRED database [2].

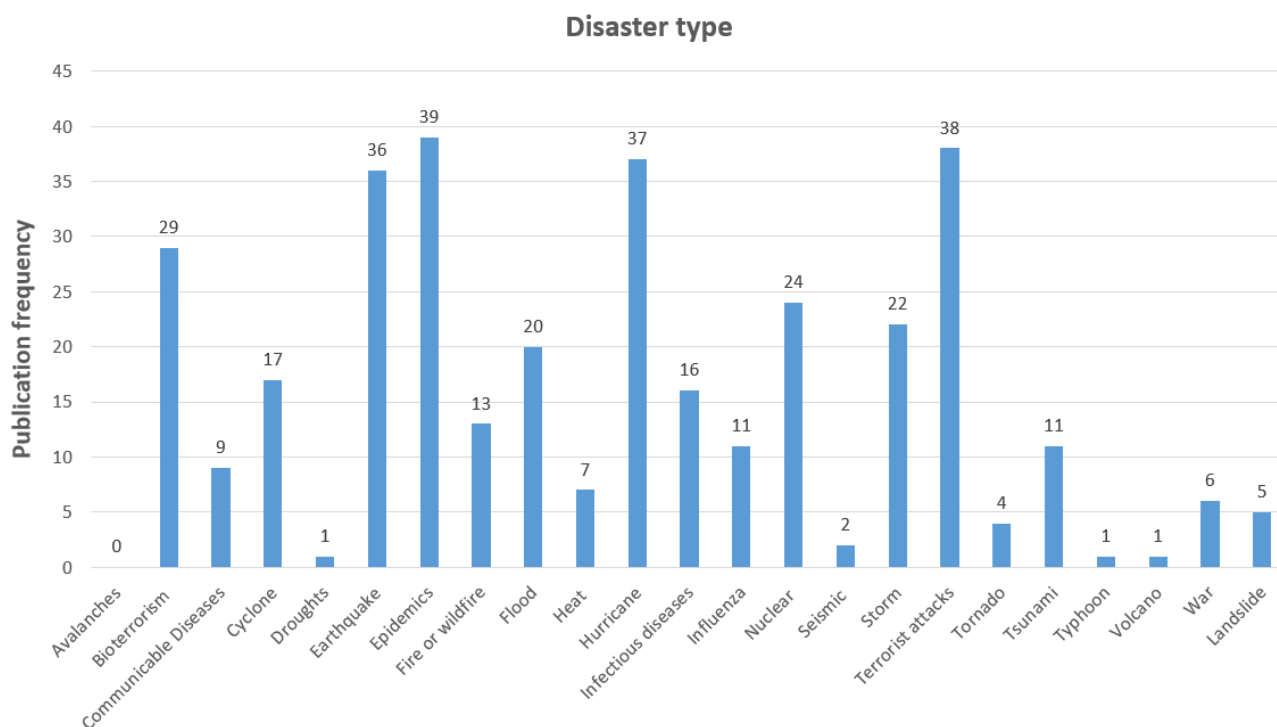
Comparison of this disaster's classifications with the research findings reveals that DEH can cover almost all types of disasters. As in the research we searched for eHealth applications within DMC, this, in turn, may be interpreted as eHealth can be used to support disaster management and disaster medicine activities across a wide range of disaster types, regardless of their sources. The detailed findings of the identified disaster types are presented in Figure 5.

On the basis of the frequency of disaster types found in the scoping results, eHealth technologies' usage is distributed across a broad range of disaster types. Nevertheless, the use of eHealth in epidemics (39/349, 11.1%), terrorist incidents (38/349, 10.8%), hurricanes (37/349, 10.6%), and earthquakes (36/349, 10.3%) is discussed more than in other disaster contexts in the literature. This may mean that these areas are likely to be more researched either because of researcher interest or the frequency of their occurrence.

To identify the disaster phases on which eHealth technologies can be used, within the DEH scope, it was referred to as four-phase DMC: mitigation, preparedness, response, and recovery [3]. The examination of the literature suggests that eHealth technologies can be used in all disaster phases, as

according to the literature in all phases eHealth technologies were used. However, they were mostly reported as being used or useful in the preparedness (255/513, 49.7%) and response (150/513, 29.2%) phases in contrast with mitigation (49/513, 9.5%) and recovery phases (59/513, 11.5%).

**Figure 5.** Identified disaster types in disaster eHealth scoping.



## Technologies Within DEH

The scoping analysis indicates an extensive variability in the list of identified technologies from different domains, most commonly related to information system and telecommunication, but extended to areas such as artificial intelligence and robotics. To reduce this complexity, technologies were demonstrated in a hierarchical representation mapped to an existing hierarchical taxonomy of eHealth technologies. Among the consulted databases, PubMed, CINAHL (Cumulative Index to Nursing and Allied Health Literature), and EBSCO (Elton B. Stephens Company) Health databases have a taxonomy of eHealth from which PubMed was chosen because of its comprehensiveness, quality, and equality of depth and breadth of the field. In CINAHL, database subject headings for eHealth are slightly different and at a higher level. In the EBSCO Health database, only *internet* was the most suitable thesaurus term for eHealth, as the rest are mostly related to health care subjects.

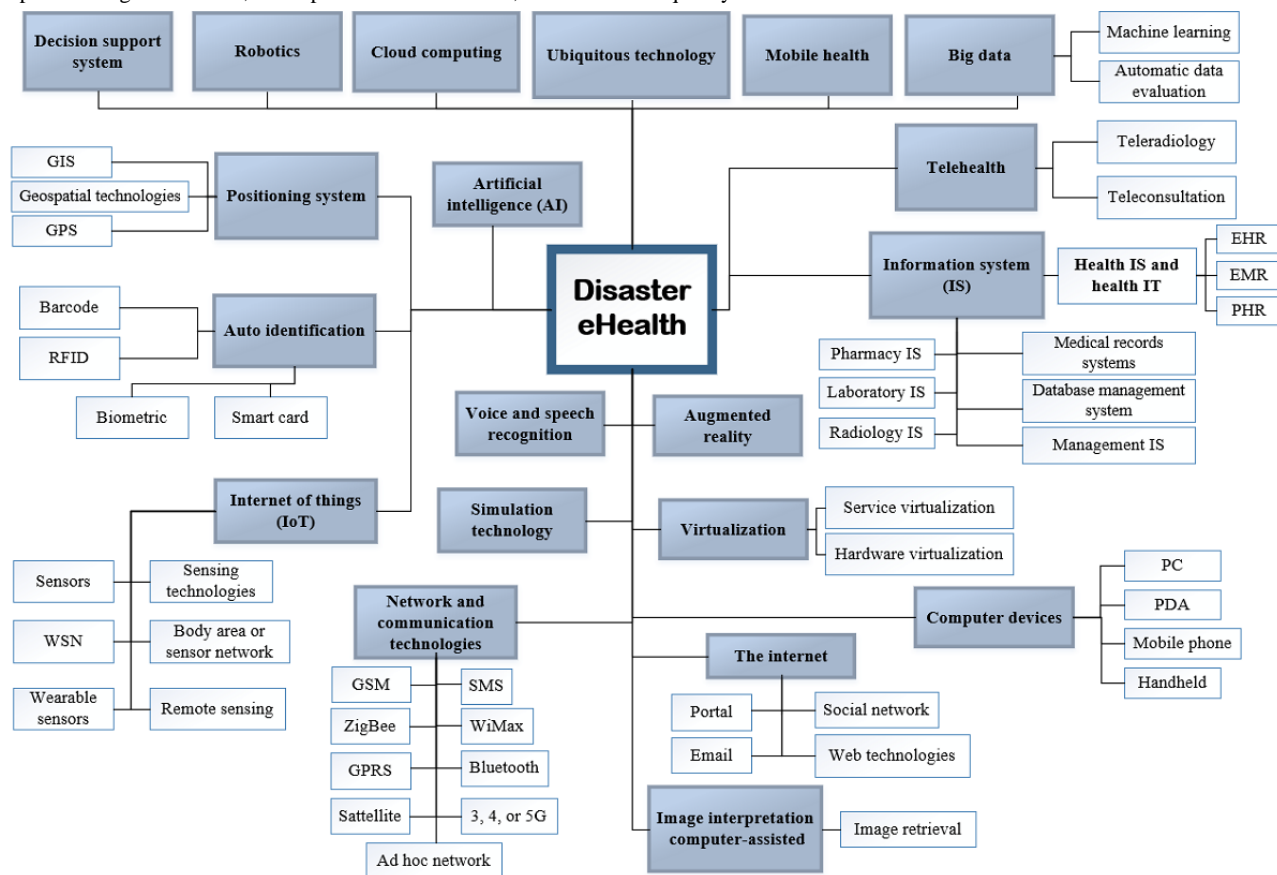
PubMed places health care computer and digital technologies in the category of *information science* with *medical informatics* as a subcategory. On the basis of the PubMed taxonomies, the identified eHealth technologies related to DEH were mapped to these top-level classifications and then further classified into

low-level technologies for a number of categories (Figure 6). In this diagram, the main technologies are shaded. This diagram may help to identify eHealth technologies that can be used in DMC to facilitate its activities.

DEH embraces a wide range of technologies to support health care activities in different disaster types and phases. A number of these technologies are specifically designed for health care environments such as *electronic health record* (EHR), *teleradiology*, and *radiology information system*. In contrast, there are some technologies that are not designed for health care environments. However, based on their positive outcomes in other areas, the health care sector has started using them for the same purposes or other clinical purposes. Technologies such as *auto identification*, *decision support system [DSS]*, or *big data* are among these technologies.

There is a vast range of eHealth technologies in the DEH scope. Among the identified eHealth technologies, our results indicate *information system* and its subtechnologies present prevailing functions and applications in DMC. Furthermore, technologies range from well-established fields, such as *DSSs*, *telehealth*, and *information system*, to the new emergent fields, such as *Internet of Things* (IoT), *augmented realities*, and *big data*.

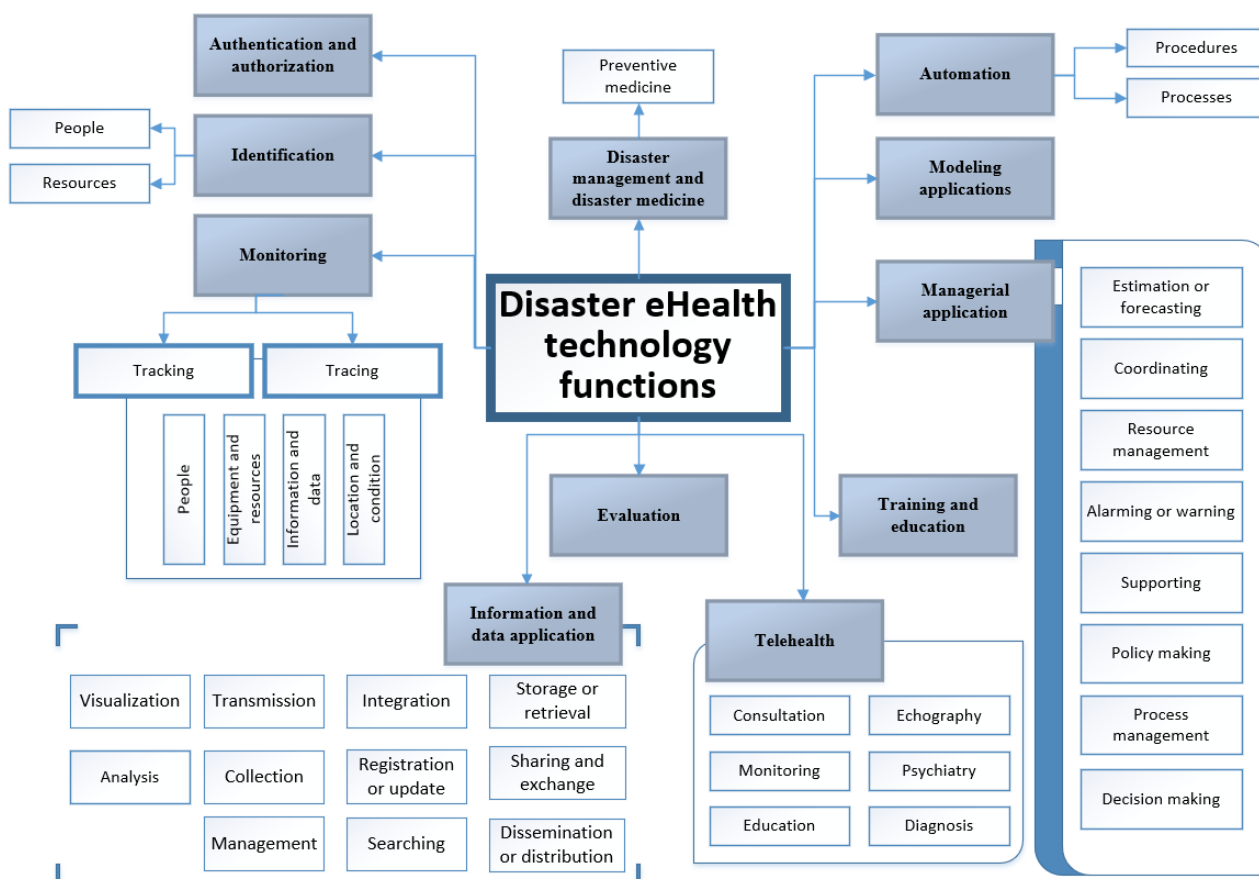
**Figure 6.** Technologies within disaster eHealth scope. EHR: electronic health record; EMR: electronic medical record; GIS: geographic information system; GPRS: General Packet Radio Service; GSM: Global System for Mobile Communications; IS: information system; IT: information technology; PDA: personal digital assistant; PHR: personal health record; RFID: radio-frequency identification.



### Technology Functions Within DEH

The identified eHealth technologies could serve specific functions within the DMC and DEH scope. Our thematic

analysis identified these functions and highlighted the major ones and classified them based on their features and applications (Figure 7). These functions demonstrate how DEH can support disaster management and medicine people in their tasks.

**Figure 7.** Technology functions within Disaster eHealth.

### Attributes of Solutions for DEH

The identified technologies and their applications support different attributes that could be important in disaster management or disaster medicine activities. On the basis of the

thematic and content analysis results, [Textbox 2](#) presents the list of the most important attributes found for eHealth technologies in DEH. These attributes could potentially facilitate disaster preparation, response, or recovery activities.

**Textbox 2.** Attributes of eHealth technologies in Disaster eHealth.

## Feature name

- Accessibility
- Accountability
- Accuracy
- Availability
- Awareness
- Collaboration
- Completeness
- Computerization
- Confidentiality
- Consistency
- Continuity
- Control
- Cooperation
- Coordination
- Effectiveness
- Efficiency
- Immediacy
- Integration
- Interoperability
- Localization
- Optimization
- Protection
- Quality
- Readiness
- Real time
- Recognition
- Relevancy
- Reliability
- Responsibility
- Robustness
- Safety
- Scalability
- Security
- Sustainability
- Telemetry
- Usability
- Web base

**DEH Purposes**

On the basis of [Figures 6](#) and [7](#), the overall purposes of the identified technologies and the functions within the DEH scope can be divided into 2 main clusters: *clinical* and *nonclinical*.

Nonclinical purposes can be further divided into administrative, education, training, and research. These DEH purposes can be defined as follows:



1. Clinical purposes: All the tasks and activities, the objectives of which are rooted in providing or expanding disaster health care services for the population.
  2. Administrative purposes: All the tasks and activities, the objectives of which directly or indirectly facilitate providing or expanding disaster health care services for the wider population and can cover health care administrative procedures from admitting patients to discharging them or making patient information transfer possible.
  3. Education and training purposes: Covers activities where the main purpose is to train and prepare citizens or special groups such as responders for different disaster phases.
  4. Research purposes: The cases where their aims are directly or indirectly related to investigation and research. They intend to improve quality, cost-effectiveness, and equity of access to health care services within the DMC.
- On the basis of these definitions, a number of example activities for each group of disaster phases are shown in [Table 2](#).

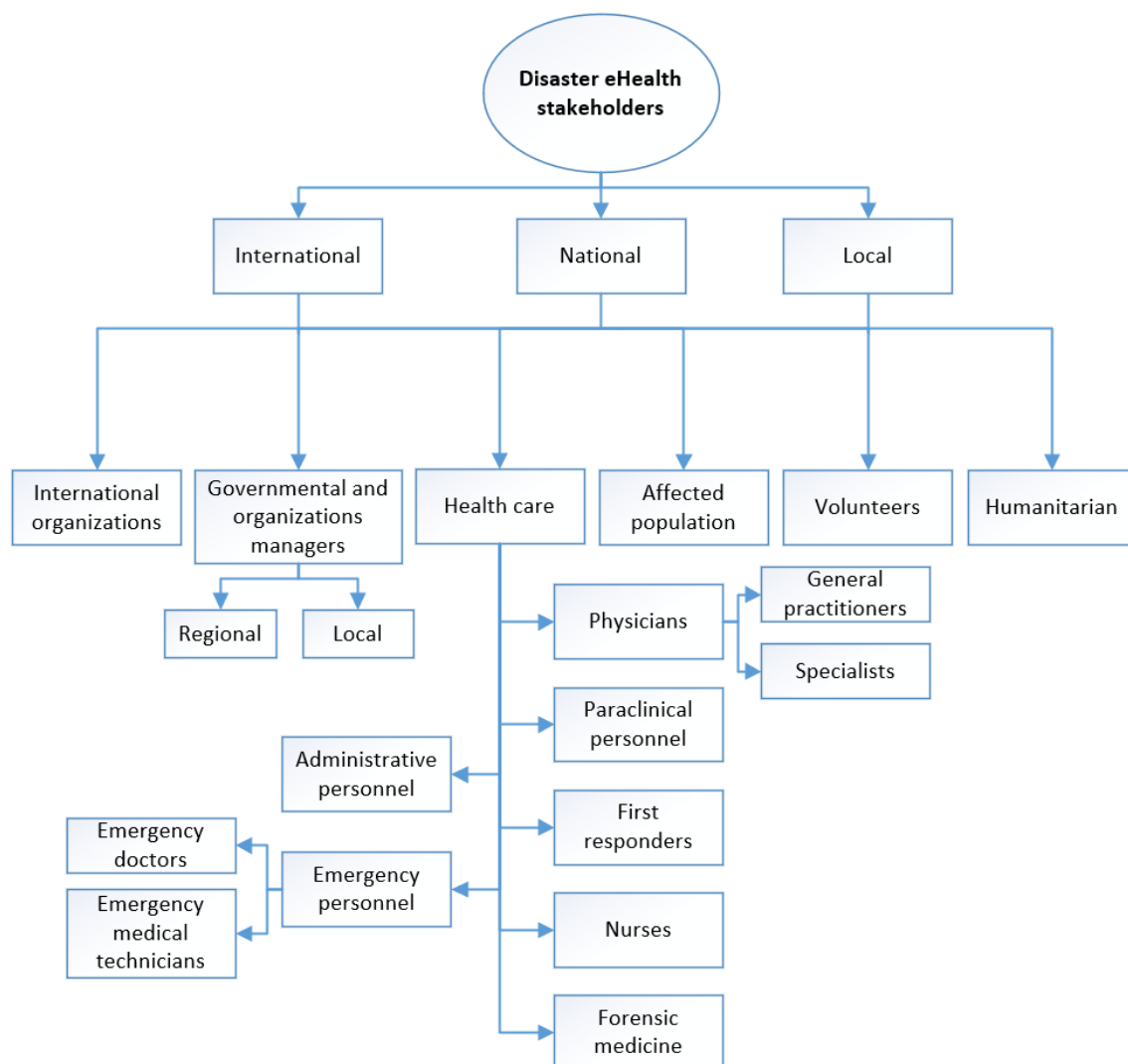
**Table 2.** Disaster eHealth activities examples with regard to disaster eHealth purposes.

Purpose and disaster phase	Sample of activity
<b>Clinical</b>	
Mitigation	Medical planning
Preparedness	Transferring and sharing of medical information
Response	Remote triaging of injured patients before arriving at hospitals
Recovery	Helping injured patients to recover at home
<b>Administrative</b>	
Mitigation	Preparing humanitarian aid program
Preparedness	Disseminating predisaster warnings
Response	Providing automation to help responders on documentation during disaster response
Recovery	Identifying and locating missing children
<b>Education and training</b>	
Mitigation	Educating people on the foundational elements of preparedness
Preparedness	Training the practitioner on disaster skills
Response	Providing continuous medical education for practitioners without enough experience
Recovery	Providing tele-education supports and services
<b>Research</b>	
Mitigation	Studying the previous disasters and research on the probability of their occurrence and their consequences
Preparedness	Carrying research on required competencies for the response time
Response	Not applicable
Recovery	Conducting studies to identify the long-term impacts of disasters on people's health

## DEH Stakeholders

The review revealed a wide range of DEH stakeholders, targets, or users of the specified technologies and functions at local, national, and international levels. The *international* level refers to the international organizations that work in any areas that directly or indirectly can have a role in any disaster phase regardless of their nature, and their branches and their supports cover almost all countries; examples of which include the World Health Organization, Red Cross, and Red Crescent. In contrast, the *national*-level organizations exist, and their rules are applied within the boundaries of each country and could vary from

country to country. At the bottom of the hierarchy, *local* organizations are located in the country at the subnational level; they work under national or governmental organizations and follow their rules. These organizations are responsible for the needs and demands of specific regions or areas; they are not supposed to make any rules, are just executors of the governmental rules, and need to report to the national organizations. Police and fire bureaus and the regional health care environments are considered at this level. On the basis of their level, we can set out the categories and subcategories in terms of different parties ([Figure 8](#)).

**Figure 8.** Disaster eHealth stakeholders.

### Services Within DEH

On the basis of the different parties and their needs in each phase of a disaster and the supported technologies and functions, DEH covers a diverse spectrum of services and purposes within the DMC. These services, from a high-level view, can be categorized into 3 levels:

1. **Operational-level services:** The function of this group of services is to support or control operations against rules and standards and encompass day-by-day decisions. Although diversity in these services can create islands of automation, they make operations more efficient. A large number of services in DMC and DEH can be categorized under this level, such as rapid victim identification, victim tracking, damage assessment, and critical resource distribution.
2. **Tactical-level services:** The purpose of this range of services is to support management and provide interconnection among different parties or organizations through diverse information management tools. These services take care of medium-term planning and are used in creating procedures. Under this level of service are categorized DSSs as well as

capacity assessment databases, health information exchange, and appropriate allocation of resources.

3. **Strategical-level services:** The purpose of these level services is to support the system as a whole, and their output is mostly policies and overall structural decisions, either among an organization's functions or among other organizations. For this group of services, information planning tools are used, and integrated infrastructure and global compatibility are essential. Above all, few services are available for this stage not only in DEH but also in conventional disaster management and medicine. These services cover long-term, complex, and nonroutine planning in DMC and DEH, such as planning for vulnerable population needs and safety, long-term care, or cloud-based coordination. To be useful, strategy must translate into tactics and delivery so that services defined at this level will have related examples at both the tactical and operational levels.

According to this classification and explanation, we can place these services for each disaster phase within the DEH scope. Some examples are presented in [Table 3](#).

**Table 3.** Disaster eHealth service level examples.

Service level and disaster phase	Sample of service
<b>Strategic</b>	
Mitigation	<ul style="list-style-type: none"> <li>Enhancing patient education and empowerment</li> <li>Analyze out-of-hospital emergency medical services</li> </ul>
Preparedness	<ul style="list-style-type: none"> <li>Preparing back-up communications systems</li> <li>Education and training of health workers</li> </ul>
Response	<ul style="list-style-type: none"> <li>Monitor, aggregate, and analyze social media data</li> <li>Cloud-based coordination</li> </ul>
Recovery	<ul style="list-style-type: none"> <li>Long-term care</li> <li>Injured people information</li> </ul>
<b>Tactical</b>	
Mitigation	<ul style="list-style-type: none"> <li>Information integration</li> <li>Medical record sharing within and across institutions</li> </ul>
Preparedness	<ul style="list-style-type: none"> <li>Resource database</li> <li>Decision support systems for bioterrorism preparedness</li> </ul>
Response	<ul style="list-style-type: none"> <li>Dynamic information collection</li> <li>Coordinating the distribution of the available medical resources</li> </ul>
Recovery	<ul style="list-style-type: none"> <li>Supply chain management</li> <li>Integrate the delivery of care after disasters</li> </ul>
<b>Operational</b>	
Mitigation	<ul style="list-style-type: none"> <li>Organize medical resources</li> <li>Analyze daily operations in emergency departments</li> </ul>
Preparedness	<ul style="list-style-type: none"> <li>Remote sensing</li> <li>Response plans that rely on local hospitals</li> </ul>
Response	<ul style="list-style-type: none"> <li>Situational awareness</li> <li>Electronic triage tag</li> </ul>
Recovery	<ul style="list-style-type: none"> <li>Psychiatrist video conferencing</li> <li>Evaluation and identification of psychological problems</li> </ul>

## Discussion

### Principal Findings

eHealth can facilitate health care data exchange and dissemination, improving communication, support, and education among communities, health care professionals, and their patients [21]. These points also apply to DMC activities that can be facilitated and supported by eHealth technologies. Nevertheless, there is a paucity of studies in this area, and eHealth technologies have not been systematically integrated into disaster management nor in disaster medicine. Integrating and utilizing eHealth technologies throughout the DMC on a systematic rather than ad-hoc basis could enhance the efficiency and effectiveness of both disciplines. This integration can improve the performance of health care and enhance the quality of its delivered services before, during, and after disasters. This differentiates DEH from its constituent fields that only deal with health care in normal medicine (such as eHealth) or have

a limited technology usage (such as disaster management and disaster medicine). DEH also covers the whole range of DMC activities and addresses all disaster phases; it accommodates all disaster types and is not a disaster-specific model.

In this study, the initial scope of the DEH is defined. DEH tries to maximize health care engagement in and integration into the DMC because an effective and successful response is almost unachievable without appropriate levels of different sectors' readiness, including health care. By integrating health care into the predisaster phases, health care can be shifted from a reactive to a proactive system when disasters occur. In this regard, the variation of eHealth technologies within the DEH scope offers a broad range of functions and applications to facilitate health care management and delivery. For example, as education and communication play a vital role, telehealth and social networks could be useful in raising public awareness or providing remote and special clinical education for physicians. The importance of this application in educating the general population [22] or providing patient care and education by physicians [23] is

acknowledged in recent studies to reduce the health consequences of the most recent disaster, COVID-19. In addition, by using IoT technology, tracking and sharing medical resources or data may be carried out automatically, which, in turn, could result in enhancing data accuracy.

The technologies themselves can also be integrated to optimize the outcome. For example, the collected data through IoT could be aggregated by cloud computing and then analyzed with big data analytics tools to support strategic disaster management planning or possibly scenario prediction. Such a framework for technology integration was proposed by Madanian and Parry [24] and recently been applied to prevent the spread of COVID-19, as suggested by Adly et al [25]. In addition, as assessing a variety of national and local medical information is a disaster mitigation requirement [26], it could be facilitated by using big data and cloud computing. Some of these technologies could offer integration within and across disaster phases. For example, continuous clinical monitoring of people with chronic diseases is an available service through the IoT and mobile health in normal medicine. If the service becomes common for the postdisaster phase, it could support continuity of care with minimal care plan disruption. This service on its own or by integrating with EHR may improve follow-up treatment and allow better care regardless of the geographical location of physicians and patients or providing better health care when there is a lack of specialists or experts in disaster-stricken areas. This area has attracted an exponential interest after the COVID-19 pandemic and has been recognized by different researchers as a viable solution to support health care professionals and reduce pressure on health care systems [27,28].

More recently, the IoT, big data, and cloud computing have attracted an exponential interest in automatic data gathering, integration, and analysis for data sharing and decision support applications. The usage of specific eHealth technologies, such as *EHR* and *telehealth*, are currently limited in disaster settings. However, recently, they have attracted significant interest in responding to COVID-19, but most applications are in developed countries, and most low-resource countries are still suffering [29,30]. This reflected on the broader advanced technological challenges and availability context in disaster situations. On the basis of our analysis, issues related to the availability of technology in developing countries and their security challenges and network requirements are some of the reasons that affect their usage. In contrast, the use of some technologies, such as positioning technologies and social networks, has rapidly increased because of their wide accessibility and availability.

In our research, we explicitly identified DEH stakeholders as technology users or targets who can benefit from DEH. These groups could be involved in DMC for a variety of purposes and in different positions. From a broader perspective, we can have the following groups:

- DEH seeks to enhance clinical and nonclinical personnel's disaster-related awareness, education, elements, standards, and procedures, mainly in disaster mitigation and preparedness phases. This could possibly result in better response and in meeting wider health care demands, as

health care teams are familiar with the very concept of disaster.

- For disaster management and disaster medicine people, DEH may facilitate technology adoption in their fields, one of the consequences of which, possibly, is rapid communication and data sharing among involved parties in DMC. This results in enhancing access to precise information in a timely manner, which, in turn, may lead to improving the quality of decisions while decreasing the decision-making time.
- DEH may also appeal to the general population who may be affected by different types of disasters. DEH, based on its defined goal, raises disaster awareness among all people, especially communities in disaster-prone areas. Therefore, population empowerment can be enhanced, and the population can access and use information, become familiar with disaster consequences, and be prepared for them. This will increase preparedness against disasters and if any disasters strike their areas, they are able to take care of their basic health care requirements until disaster responders arrive. Then, as disaster responders have proper training and are equipped with different types of eHealth technologies, they are able to transfer timely and accurate information from disaster sites to top authorities so that they can make appropriate decisions. This, in turn, provides better health care services to disaster casualties.

## Conclusions

Preparedness and planning to reduce the harmful effects of disasters is becoming one of the highest priorities of governments. These activities are features of disaster management and disaster medicine; disciplines that despite their long standing still generate many debates about their effectiveness and capabilities in responding properly to health care demands in major disasters [31].

This research, along with other studies such as by Sieben et al [32] and Norris et al [6], is seeking ways to develop systematic principles for coordinating disaster management and disaster medicine more effectively by incorporating eHealth technologies. These technologies may improve the response to health care demands before, during, and after disasters. They can be used to improve overall disaster management, facilitate response when disasters occur, enhance support after disasters, and keep records for better future preparedness. Although some eHealth tools are employed in disaster settings [31], their applications are not generally systematic and routine; rather, they are more ad hoc. The most recent example is the COVID-19 pandemic that has increased the interest in eHealth technology adaption to have a more effective response while reducing the strain on health care. However, the issues with the utilization of these technologies are still their ad hoc usage and lack of system integration with health care mainstream activities. Moreover, there are different examples of using eHealth technologies to respond and recover from COVID-19 but with no overarching guidelines of when and how to apply them, and in some cases, training is required to promote success.

The DEH domain has been introduced mainly to facilitate addressing the current challenges within disaster management

and disaster medicine that hinder their operations and created many debates regarding their efficiency and effectiveness. DEH emergence contributes to the design of a systematic model for eHealth technologies that are currently used in nondisaster circumstances but have the potential to be used in disaster situations along with those technologies that were previously used in DMC and had a significant impact on DMC operations.

DEH takes advantage of eHealth technologies to facilitate DMC tasks and activities, enhance their efficiency and effectiveness, and improve health care delivery to a wider population regardless of disaster types and phases.

In this research, we extensively reviewed the academic literature to define the DEH scope. We have built our scope mostly based

on available international hierarchies to make easier embedding DEH into disaster management, disaster medicine, and eHealth fields. Some of the international hierarchies that we referred to are the disaster types offered by CRED and PubMed Medical Informatics taxonomies. However, this work is mostly limited to academic and scientific publications, and gray literature is not extensively reviewed.

eHealth technologies are developing rapidly, and the COVID-19 pandemic has revealed some of the eHealth potentials in practice on addressing health care issues. Therefore, we would consider and continue to work on the DEH model and add the most recent applications, such as contact tracing, into the model in the near future.

## Conflicts of Interest

None declared.

## References

- Baldini G, Hess E, Oliveri F, Seuschek H, Braun M. Secure RFID for Humanitarian Logistics. In: Turcu C, editor. Designing and Deploying RFID Applications. New York, USA: IntechOpen; 2011.
- Criteria and Definition. Center for Research on the Epidemiology of Disasters. 2009. URL: <http://www.emdat.be/criteria-and-definition> [accessed 2019-11-18]
- Baldini G, Braun M, Hess E, Oliveri F, Seuschek H. The Use of Secure RFID to Support the Resolution of Emergency Crises. In: 43rd Annual 2009 International Carnahan Conference on Security Technology. 2009 Presented at: International Carnahan Conference on Security Technology; October 5-8, 2009; Zurich, Switzerland. [doi: [10.1109/ccst.2009.5335517](https://doi.org/10.1109/ccst.2009.5335517)]
- Levett J. Disastrous events and political failures. *Prehosp Disaster Med* 2015 Jun;30(3):227-228. [doi: [10.1017/S1049023X15004689](https://doi.org/10.1017/S1049023X15004689)] [Medline: [25882131](https://pubmed.ncbi.nlm.nih.gov/25882131/)]
- Bissell R. Public healthmedicine in emergency management. In: McEntire D, editor. Disciplines, Disasters, and Emergency Management. Springfield, IL: Charles C Thomas; 2007:213-223.
- Norris A, Martinez S, Labaka L, Madanian S, Gonzalez J, Parry D. Disaster E-health: A New Paradigm for Collaborative Healthcare in Disasters. In: International Conference on Information Systems for Crisis Response and Management. 2015 Presented at: ISCRAM'15; May 24-27, 2015; Krystiansand, Norway URL: <https://dblp.org/db/conf/iscram/iscram2015.html>
- Chronaki C, Kontoyiannis V, Charalambous E, Vrouchos G, Mamantopoulos A, Vourvahakis D. Satellite-enabled eHealth applications in disaster management-experience from a readiness exercise. *Comput Cardiol* 2008;14:1005-1008. [doi: [10.1109/cic.2008.4749214](https://doi.org/10.1109/cic.2008.4749214)]
- Blanchard B. Guide to Emergency Management and Related Terms, Definitions, Concepts, Acronyms, Organizations, Programs, Guidance, Executive Orders and Legislation 2008. Emergency Management Institute. URL: <https://training.fema.gov/hiedu/docs/terms> [accessed 2020-07-10]
- Hogan D, Burstein J. Basic perspectives on disaster. In: Hogan D, Burstein J, editors. Disaster Medicine. Philadelphia, PA: Lippincott Williams & Wilkins; 2007:1-11.
- Resolutions and Deliberations on Ehealth. World Health Organization. 2005. URL: <https://www.who.int/ehealth/about/en/> [accessed 2020-10-01]
- Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005 Feb;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
- Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implementation Science* 2010 Sep 20;5:69 [FREE Full text] [doi: [10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69)] [Medline: [20854677](https://pubmed.ncbi.nlm.nih.gov/20854677/)]
- Gross T, Taylor AG, Joudrey DN. Still a lot to lose: the role of controlled vocabulary in keyword searching. *Catalog Classif Q* 2014 Nov 25;53(1):1-39. [doi: [10.1080/01639374.2014.917447](https://doi.org/10.1080/01639374.2014.917447)]
- Valaitis R, Martin-Misener R, Wong ST, MacDonald M, Meagher-Stewart D, Austin P, Strengthening Primary Health Care through Public HealthPrimary Care Collaboration Team. Methods, strategies and technologies used to conduct a scoping literature review of collaboration between primary care and public health. *Prim Health Care Res Dev* 2012 Jul;13(3):219-236. [doi: [10.1017/S1463423611000594](https://doi.org/10.1017/S1463423611000594)] [Medline: [22336106](https://pubmed.ncbi.nlm.nih.gov/22336106/)]
- Haughey D. Pareto Analysis Step by Step. Project Smart. URL: <https://www.projectsmart.co.uk/pareto-analysis-step-by-step.php> [accessed 2020-01-14]
- Schwalbe K. Information Technology Project Management. Boston, MA: CENGAGE; 2018.



17. EndNote. URL: <https://endnote.com/> [accessed 2019-11-19]
18. Nvivo. URL: <https://www.qsrinternational.com/nvivo/home> [accessed 2019-11-19]
19. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
20. Schreier M. Qualitative content analysis. In: *The Sage Handbook of Qualitative Data Analysis*. London, UK: SAGE Publications Ltd; 2014:170-183.
21. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. *J Med Internet Res* 2005 Mar 31;7(1):e9 [FREE Full text] [doi: [10.2196/jmir.7.1.e9](https://doi.org/10.2196/jmir.7.1.e9)] [Medline: [15829481](https://pubmed.ncbi.nlm.nih.gov/15829481/)]
22. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. *J Med Internet Res* 2020 May 19;22(5):e19556 [FREE Full text] [doi: [10.2196/19556](https://doi.org/10.2196/19556)] [Medline: [32369026](https://pubmed.ncbi.nlm.nih.gov/32369026/)]
23. Wahbeh A, Nasrallah T, Al-Ramahi M, El-Gayar O. Mining Physicians' Opinions on Social Media to Obtain Insights Into COVID-19: Mixed Methods Analysis. *JMIR Public Health Surveill* 2020 Jun 18;6(2):e19276 [FREE Full text] [doi: [10.2196/19276](https://doi.org/10.2196/19276)] [Medline: [32421686](https://pubmed.ncbi.nlm.nih.gov/32421686/)]
24. Madanian S, Parry D. IoT, Cloud Computing and Big Data: Integrated Framework for Healthcare in Disasters. *Studies in Health Technology and Informatics* 2019 Aug 21;264:998-1002. [doi: [10.3233/SHTI190374](https://doi.org/10.3233/SHTI190374)] [Medline: [31438074](https://pubmed.ncbi.nlm.nih.gov/31438074/)]
25. Adly AS, Adly AS, Adly MS. Approaches Based on Artificial Intelligence and the Internet of Intelligent Things to Prevent the Spread of COVID-19: Scoping Review. *J Med Internet Res* 2020 Aug 10;22(8):e19104 [FREE Full text] [doi: [10.2196/19104](https://doi.org/10.2196/19104)] [Medline: [32584780](https://pubmed.ncbi.nlm.nih.gov/32584780/)]
26. Chan TC, Killeen J, Griswold W, Lenert L. Information technology and emergency medical care during disasters. *Acad Emerg Med* 2004 Nov;11(11):1229-1236. [doi: [10.1197/j.aem.2004.08.018](https://doi.org/10.1197/j.aem.2004.08.018)] [Medline: [15528589](https://pubmed.ncbi.nlm.nih.gov/15528589/)]
27. Ye J. The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. *JMIR Med Inform* 2020 Jul 14;8(7):e19866 [FREE Full text] [doi: [10.2196/19866](https://doi.org/10.2196/19866)] [Medline: [32568725](https://pubmed.ncbi.nlm.nih.gov/32568725/)]
28. Fisk M, Livingstone A, Pit SW. Telehealth in the context of COVID-19: changing perspectives in Australia, the United Kingdom, and the United States. *J Med Internet Res* 2020 Jun 9;22(6):e19264 [FREE Full text] [doi: [10.2196/19264](https://doi.org/10.2196/19264)] [Medline: [32463377](https://pubmed.ncbi.nlm.nih.gov/32463377/)]
29. Hopman J, Allegranzi B, Mehtar S. Managing COVID-19 in low- and middle-income countries. *J Am Med Assoc* 2020 Apr 28;323(16):1549-1550. [doi: [10.1001/jama.2020.4169](https://doi.org/10.1001/jama.2020.4169)] [Medline: [32176764](https://pubmed.ncbi.nlm.nih.gov/32176764/)]
30. English M, Moshabela M, Nzinga J, Barasa E, Tsofa B, Marchal B, et al. Systems and implementation science should be part of the COVID-19 response in low resource settings. *BMC Med* 2020 Jul 15;18(1):219 [FREE Full text] [doi: [10.1186/s12916-020-01696-6](https://doi.org/10.1186/s12916-020-01696-6)] [Medline: [32664950](https://pubmed.ncbi.nlm.nih.gov/32664950/)]
31. Maturana C, Scott R, Palacios M. e-Health the haddon matrix: Identifying where and how e-health can assist in disaster management. In: Jordanova M, Lievens F, editors. *Global Telemedicine and Ehealth Updates: Knowledge Resources*. Grimbergen, Belgium: International Society for Telemedicine & eHealth (ISfTeH); 2012.
32. Sieben C, Scott R, Palacios M. E-Health and disaster management cycle. In: Jordanova M, Lievens F, editors. *Global Telemedicine and Ehealth Updates: Knowledge Resources*. Grimbergen, Belgium: International Society for Telemedicine & eHealth (ISfTeH); 2012.

## Abbreviations

**CINAHL:** Cumulative Index to Nursing and Allied Health Literature  
**CRED:** Centre for Research on the Epidemiology of Disasters  
**DEH:** disaster eHealth  
**DMC:** disaster management cycle  
**DSS:** decision support system  
**EBSCO:** Elton B Stephens Company  
**EHR:** electronic health record  
**IoT:** Internet of Things

*Edited by G Eysenbach; submitted 18.02.20; peer-reviewed by N Mazlan, C Wu; comments to author 29.06.20; revised version received 13.08.20; accepted 15.09.20; published 28.10.20.*

*Please cite as:*

*Madanian S, Norris T, Parry D*

*Disaster eHealth: Scoping Review*

*J Med Internet Res 2020;22(10):e18310*

*URL: <https://www.jmir.org/2020/10/e18310>*

*doi: [10.2196/18310](https://doi.org/10.2196/18310)*

*PMID: [33112244](https://pubmed.ncbi.nlm.nih.gov/33112244/)*

©Samaneh Madanian, Tony Norris, Dave Parry. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Review

# Primary Prevention of Cardiovascular Disease and Type 2 Diabetes Mellitus Using Mobile Health Technology: Systematic Review of the Literature

Vera Helen Buss<sup>1,2</sup>, PharmB, RPh, MRes; Stuart Leesong<sup>1</sup>; Margo Barr<sup>1</sup>, BSc, MPH, GCertTertTLP, PhD; Marlien Varnfield<sup>2</sup>, BSc, MSc, PhD; Mark Harris<sup>1</sup>, MBBSMD, DRACOG, FRACGP, PhD

<sup>1</sup>Centre for Primary Health Care and Equity, University of New South Wales, Sydney, Australia

<sup>2</sup>Australian e-Health Research Centre, CSIRO, Brisbane, Australia

**Corresponding Author:**

Vera Helen Buss, PharmB, RPh, MRes  
Centre for Primary Health Care and Equity  
University of New South Wales  
Level 3, AGSM Building  
UNSW Sydney  
Sydney, 2052  
Australia  
Phone: 61 293851547  
Email: [v.buss@student.unsw.edu.au](mailto:v.buss@student.unsw.edu.au)

## Abstract

**Background:** Digital technology is an opportunity for public health interventions to reach a large part of the population.

**Objective:** This systematic literature review aimed to assess the effectiveness of mobile health-based interventions in reducing the risk of cardiovascular disease and type 2 diabetes mellitus.

**Methods:** We conducted the systematic search in 7 electronic databases using a predefined search strategy. We included articles published between inception of the databases and March 2019 if they reported on the effectiveness of an intervention for prevention of cardiovascular disease or type 2 diabetes via mobile technology. One researcher performed the search, study selection, data extraction, and methodological quality assessment. The steps were validated by the other members of the research team

**Results:** The search yielded 941 articles for cardiovascular disease, of which 3 met the inclusion criteria, and 732 for type 2 diabetes, of which 6 met the inclusion criteria. The methodological quality of the studies was low, with the main issue being nonblinding of participants. Of the selected studies, 4 used SMS text messaging, 1 used WhatsApp, and the remaining ones used specific smartphone apps. Weight loss and reduction in BMI were the most reported successful outcomes (reported in 4 studies).

**Conclusions:** Evidence on the effectiveness of mobile health-based interventions in reducing the risk for cardiovascular disease and type 2 diabetes is low due to the quality of the studies and the small effects that were measured. This highlights the need for further high-quality research to investigate the potential of mobile health interventions.

**Trial Registration:** International Prospective Register of Systematic Reviews (PROSPERO) CRD42019135405; [https://www.crd.york.ac.uk/PROSPERO/display\\_record.php?RecordID=135405](https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=135405)

(*J Med Internet Res* 2020;22(10):e21159) doi:[10.2196/21159](https://doi.org/10.2196/21159)

**KEYWORDS**

systematic review; mobile health; telemedicine; primary prevention; cardiovascular diseases; diabetes mellitus, type 2

## Introduction

**Description of the Condition**

Worldwide, chronic diseases are the main cause of death and years lived with disability [1,2]. Cardiovascular disease (CVD)

and type 2 diabetes mellitus (T2DM) are globally among the top 5 chronic conditions in terms of incidence and prevalence [2]. The behavioral risk factors for these conditions, such as smoking, harmful use of alcohol, poor diet, and physical inactivity, are highly correlated with the disease progression [3]. For example, Gellert et al [4] observed in their meta-analysis

a dose-response relationship between the number of cigarettes smoked and premature death. They also found an inverse correlation between time since cessation and all-cause mortality. Wood et al [5] reported that all-cause mortality was positively associated with the level of alcohol intake, based on data from over half a million current drinkers. Chudasama et al [6] found a negative dose-response relationship between physical activity levels and all-cause mortality in their analysis of almost half a million people. Regarding low whole-grain intake, which is the highest risk factor related to poor diet, in their meta-analysis, Zhang et al [7] showed an inverse dose-response relationship between whole-grain intake and all-cause mortality. Hence, targeting these with preventive measures could significantly reduce people's chronic disease risk [8], and behavior change interventions are well suited for preventing CVD and T2DM [2,3].

### Description of the Intervention

To stop noncommunicable diseases from rising further, the World Health Organization (WHO) developed the Global Action Plan 2013-2020 [8]. In this report, the WHO emphasized the importance of early screening and the implementation of preventive programs. Further, the WHO recommended the use of information and communication technologies, such as the internet and mobile phone technologies, to deliver health education and promotion programs. In 2019, the WHO released a guideline with recommendations on digital interventions for health system strengthening [9]. This report outlined how the implementation of technology could overcome current challenges in health care systems and help to achieve the goal of universal health coverage. Health apps have promising potential. Wilson [10] pointed out that digital health interventions have the advantage of being easily accessible and cost-effective. According to the Pew Research Center [11], many people use their smartphones daily. Riley et al [12] reported that new advancements allow apps to be tailored to personal needs and preferences, as well as the integration of dynamic feedback systems. Despite the promising potential of health apps, there is still ambiguity about their effectiveness, as outlined by the WHO guideline [9].

### Objective

The aim of this systematic literature review was to assess the current evidence regarding the effectiveness of mobile health-based interventions in reducing the risk for CVD and T2DM. The focus was on multiple behavioral risk-factor interventions, rather than single risk-factor interventions, because of the lack of evidence on their combined effectiveness compared with substantial evidence on single risk-factor interventions [13,14].

## Methods

### Review Standards

We conducted this systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [15] and registered it with International Prospective Register of Systematic Reviews (PROSPERO; registration number CRD42019135405).

### Search Strategy

We searched the following medical and bioengineering databases to retrieve all relevant articles regarding preventive mobile health intervention for CVD and T2DM: EMBASE (via Ovid), Scopus, ScienceDirect, CINAHL (via EBSCOhost), MEDLINE (via Ovid), ProQuest science and technology databases, and Ei Compindex and Inspec (both via Engineering Village 2). The search strategy ([Multimedia Appendix 1](#)) included terms relating to the 2 conditions under study and the intervention; we combined the terms using Boolean operators [16] and adapted the terms to the database-specific requirements. The search included articles published from the inception of the databases until March 25, 2019. We limited the search to English- and German-language publications because these languages were proficiently spoken by the review team. We excluded review articles, conference abstracts, comments, editorials, letters to the editor, and theses. Additionally, we identified studies using "snowballing" techniques by reviewing the reference lists of articles included in the initial search and searching for other publications by authors included in the initial search [17].

### Study Selection

#### Inclusion Criteria

The study selection followed predefined inclusion criteria according to the PICOS system ([Table 1](#)). After removing duplicate publications, we reviewed all retrieved articles for eligibility, first by examining the titles and abstracts, and then the full articles if we considered the articles to be relevant in the first step. We included in the review full articles that met the inclusion and exclusion criteria. The steps described above were performed by 1 researcher (VHB). For the title and abstract screening, a 10% random sample of all retrieved articles was validated by a second researcher (shared between the remaining researchers). If discrepancies occurred, a third researcher resolved the issue. A second researcher (SL) independently assessed which of the full articles fulfilled the inclusion and exclusion criteria. The results were compared, and discrepancies were resolved by involving a third researcher (MB).

**Table 1.** Inclusion criteria according to the PICOS system.

Criteria	Description of inclusion criteria
Participants	Adults who are free of CVD <sup>a</sup> or T2DM <sup>b</sup> .
Intervention	Health promotion interventions that use mobile health technology (ie, mobile app or SMS text messaging) aiming to change more than 1 risk factor for 1 of the 2 chronic conditions under study.
Comparator	No intervention (ie, standard care), or waitlist control, or intervention delivered in person.
Outcome	Onset of disease (CVD or T2DM) or relative risk reduction, which can be in the form of surrogate parameters.
Study design	Randomized controlled trial, case-control study, or interrupted time series.

<sup>a</sup>CVD: cardiovascular disease.

<sup>b</sup>T2DM: type 2 diabetes mellitus.

### **Types of Participants**

Participants could either be healthy or have an increased disease risk. We excluded interventions targeting adults who were already diagnosed with CVD or T2DM (depending on the aim of the intervention, eg, for CVD prevention, people diagnosed with CVD) at baseline. Further, we excluded studies intended for minors (<18 years of age). The conditions under study were CVD and T2DM, for which we applied the following WHO definitions: CVD is a “group of disorders of heart and blood vessels,” including coronary heart disease, cerebrovascular disease, peripheral vascular disease, heart failure, rheumatic heart disease, congenital heart disease, and cardiomyopathies [18]; T2DM “is a chronic disease that occurs...when the body cannot effectively use the insulin it produces” [19].

### **Types of Intervention**

We included primary studies if they evaluated the effectiveness of a mobile phone–based intervention for primary prevention of 1 of the conditions under study. The intervention had to be delivered, at least partially, via mobile health technology (ie, mobile app or SMS text messaging) with the aim of changing more than 1 risk factor for 1 or more of the chronic conditions under study. We defined a mobile app as a software program that can run on mobile devices such as smartphones, and a text message as a written message sent to a mobile phone. The type of interventions that we included needed to be aimed at health promotion using behavior change strategies, including counselling or education regarding disease-related knowledge, healthy diet, physical activity, smoking cessation, motivational messages, and goal setting. We excluded from the review studies that exclusively targeted 1 behavioral risk factor (eg, smoking only, diet only, or step count only).

### **Types of Comparator**

The comparison group could consist of either no intervention (ie, standard care), or a waitlist control, or an intervention delivered in person. Studies were eligible if they included adults who were free of CVD or T2DM at study baseline, depending on the condition targeted in the study.

### **Types of Outcome**

Studies were only eligible for inclusion if their main outcomes were disease incidence (either CVD or T2DM) or a reduction

in disease risk, which could be measured using a risk prediction tool (such as the Framingham score for CVD [20]) or surrogate parameters. Examples of surrogate parameters were weight, waist circumference, blood pressure, blood glucose, level of physical activity, dietary intake, or smoking status. Additional outcomes that we included in the review were the feasibility of mobile health interventions, disease knowledge, and quality of life. Respective outcome measures included dropout rates, participants’ acceptability of and adherence to the intervention, and questionnaires assessing disease knowledge and quality of life.

### **Types of Study Design**

We restricted the study design to randomized controlled trials (RCTs), case-control studies, and interrupted time series in order to have a measurement against which the effectiveness of the intervention could be compared.

### **Data Extraction and Synthesis**

Relevant data (study objective, study design, study population, comparator, description of the intervention, duration of the intervention or follow-up, outcomes, main results, and methodology for the assessment of the study’s quality) were extracted by 1 researcher (VHB) using a standardized form in Excel 365 (Microsoft Corporation). This was reviewed by all the other researchers. We synthesized the main results of the included studies in a narrative manner focusing on the intervention delivery and reported outcomes. A meta-analysis was not possible due to the small number of identified studies and the heterogeneity in interventions and outcomes.

### **Literature Quality Assessment**

One researcher (VHB) assessed the risk of bias using the following assessment tools: for RCTs, the Cochrane Collaboration’s tool for assessing risk of bias [21]; and for non-RCTs, the Risk of Bias In Non-randomized Studies - of Intervention assessment tool [22].

## **Results**

### **Results of the Literature Search and Study Selection**

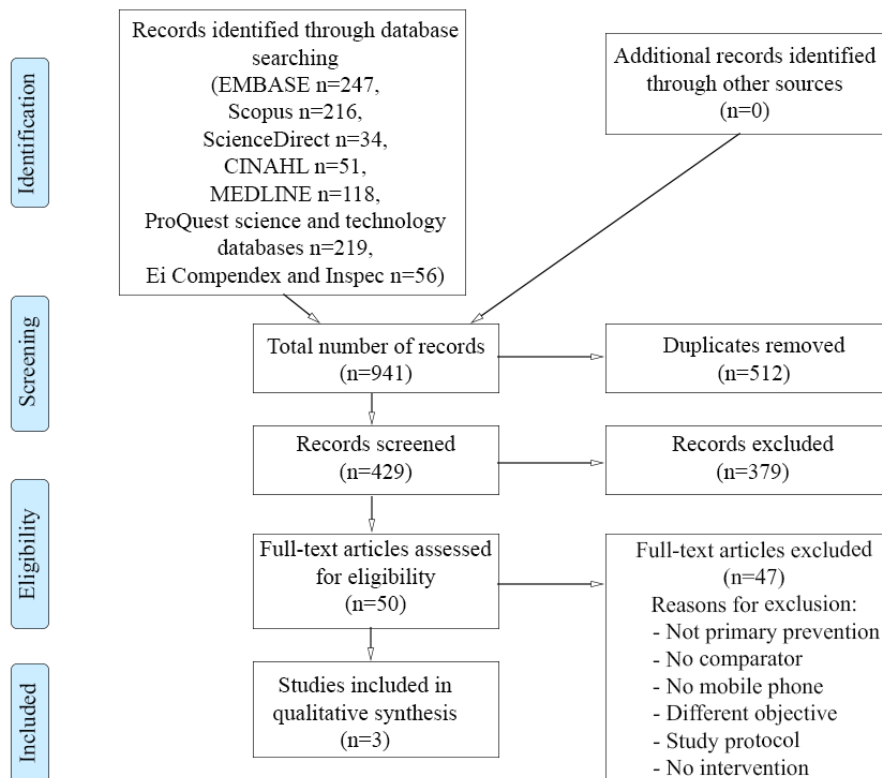
In total, we identified 941 articles using the search strategy for CVD and 732 articles using the search strategy for T2DM. In the validation of the 10% random sample of all retrieved articles,



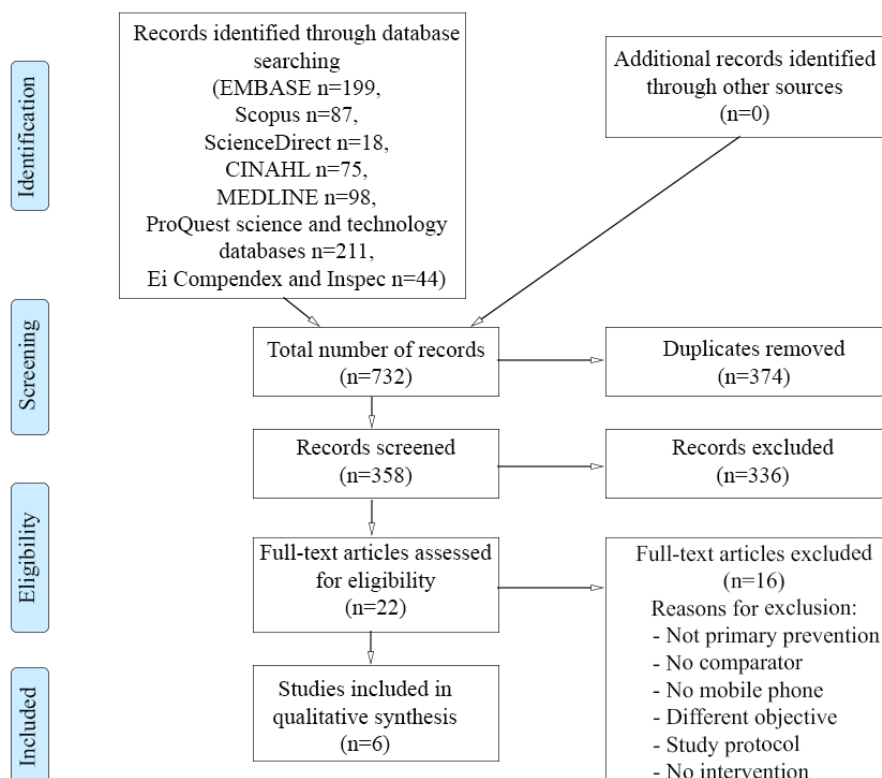
there was a 100% agreement (after initial disagreements were resolved by a third investigator) with the selection conducted by the researcher who screened all articles. Finally, 3 CVD articles [23-25] and 6 T2DM articles [26-31] fulfilled the

inclusion and exclusion criteria; we identified no additional articles through the snowballing technique (Figures 1 and 2). We excluded many articles for several of the exclusion criteria.

**Figure 1.** Full article selection process for cardiovascular disease.



**Figure 2.** Full article selection process for type 2 diabetes.



## Results of the Data Extraction

There were 3 CVD [23-25] and 6 T2DM studies [26-31]. Table 2 provides details about the CVD studies and Table 3 provides

details about the T2DM studies. For each study, the table includes the first author, year of publication, study design and duration, objectives, study population, interventions and comparators, outcomes, and the main results.

**Table 2.** Data extraction from cardiovascular disease (CVD) studies.

First author, date, reference	Study design and duration; objectives	Study population	Intervention and comparator	Outcomes	Main results
Gore, 2019 [23]	Non-RCT <sup>a</sup> for 12 months; effectiveness of an SMS text message intervention to reduce CVD risk	Adults from the United States at high risk of CVD without preexisting coronary artery disease, cerebrovascular disease, and diabetes; intervention n=204, usual care n=408	Create action plan with community health workers and return 6-12 months after initial screening for retesting; intervention: text messages once/day on advice on healthy eating, PA <sup>b</sup> , weight loss, contacting community health worker; control: usual care	Engagement, program retention, changes in risk factors (smoking, fat and fiber intake, PA, weight, BMI, BP <sup>c</sup> , low-density lipoprotein), Framingham risk score	Only statistically significant decrease in fat intake (intervention -26.3% vs control -10.6%; $P=.001$ )
Muntaner-Mas, 2017 [24]	Non-RCT for 10 weeks; effectiveness of a WhatsApp-based PA intervention to reduce CVD risk factors	Spanish adults aged 53-73 years without medical conditions or other physical problems requiring special medical attention and who were able to perform rigorous PA; mobile group n=7, training group n=16, control n=9	Intervention: twice/week functional fitness for training and mobile group; for training group face-to-face sessions, for mobile group training videos for download via WhatsApp, chat function plus motivational messages from study coordinator; control: no intervention	BP, WC <sup>d</sup> , waist to height ratio, weight, BMI, fat mass index, fat-free mass index, heart rate after exercise, balance, handgrip strength, aerobic capacity	No statistically significant differences between mobile group and control; statistically significant differences between training group and control group (systolic BP $P=.038$ ; diastolic BP $P=.005$ ; mean arterial BP $P=.006$ ; heart rate after exercise $P=.002$ )
Rubinstein, 2016 [25]	RCT for 12 months; effectiveness of preventive mobile health intervention in adults with prehypertension	Adults aged 30-60 years with prehypertension from poor urban settings in Argentina, Guatemala, and Peru, free of hypertension, diabetes, and CVD; intervention n=316, usual care n=321	Intervention: monthly motivational counselling calls (healthy diet and PA) followed by weekly text messages related to behavior goals and readiness to change; control: usual care	Changes in BP, weight, BMI, WC, PA, diet	Mean differences, baseline-adjusted (95% CI): weight -0.66 kg (-1.24 to -0.07), BMI -0.30 kg/m <sup>2</sup> (-0.54 to -0.06), daily intake of high-sugar and -fat servings -0.75 (-1.30 to -0.20); change in BP not significant

<sup>a</sup>RCT: randomized controlled trial.

<sup>b</sup>PA: physical activity.

<sup>c</sup>BP: blood pressure.

<sup>d</sup>WC: waist circumference.

**Table 3.** Data extraction from type 2 diabetes mellitus (T2DM) studies.

First author, date, reference	Study design and duration; objectives	Study population	Intervention and comparator	Outcomes	Main results
Arens, 2018 [26]	Non-RCT <sup>a</sup> for 12 months; effectiveness of app-based weight reduction program for people with metabolic syndrome	German adults aged 30-65 years treated for metabolic syndrome in 23 medical practices; intervention n=148, usual care n=85	Health goals regarding weight and PA <sup>b</sup> ; app for feedback; physicians with access to app data could give feedback, initiate messages, or modify goals; ≤9 free classes on diet and PA; control: usual care	5% weight reduction; change in BMI	5% weight reduction (adjusted for time in study) (95% CI): 44.8% (34.1 to 57.1) in intervention vs 11.5% (4.6 to 27.0) in control; Cox proportional hazard model for time to 5% weight reduction hazard ratio 6.2 (2.4 to 16.2; $P<.001$ ), baseline adjusted between groups change in weight (kg) $P=.06$ and BMI ( $\text{kg/m}^2$ ) $P=.10$
Bender, 2018 [27]	RCT for 3 months plus 3 months follow-up (no control for follow-up); effectiveness of mobile phone-based weight loss intervention to reduce T2DM risk	Filipino-American overweight or obese adults from United States at increased risk for T2DM, able to walk 20 min; intervention n=33, control n=34	5 in-person sessions, daily step count via wearable device, daily food intake and weekly weight logged in app, weekly information on weight loss, PA, and diet via private Facebook page; control: waitlist	Recruitment (goal n=50), retention, 5% weight loss, changes in weight, BMI, WC <sup>c</sup> , FBG <sup>d</sup> , HbA <sub>1c</sub> <sup>e</sup>	Weight loss ≥5%: intervention 36% vs control 6%; between-group cross-level interaction (95% CI): weight $-1.1\%/month$ ( $-1.7$ to $-0.53$ ) and $-0.85\text{ kg/month}$ ( $-1.4$ to $-0.35$ ), BMI $-0.93\text{ kg/m}^2$ ( $-1.5$ to $-0.40$ ), WC $-4.9\text{ cm}$ ( $-7.5$ to $-2.6$ ), FBG $-1.4\text{ mg/dL}$ ( $-5.9$ to $3.6$ ), HbA <sub>1c</sub> $-0.10\%$ ( $-0.21$ to $0.002$ )
Block, 2015 [28]	RCT for 6 months plus 6 months follow-up (no control for follow-up); effectiveness of digital health intervention for T2DM risk reduction in prediabetics	Prediabetics aged 30-69 years from United States with BMI ≥27 $\text{kg/m}^2$ , without diabetes medication; intervention n=163, control n=176	Tailored behavioral support for PA, diet, weight loss, stress, sleep; weekly emails with goals linked to website (tracking tools, coaching, social support, competition, health advice), app and automated phone calls; control: waitlist	Decreased HbA <sub>1c</sub> , FBG, weight, BMI, WC, triglyceride to HDL <sup>f</sup> ratio, metabolic syndrome, Framingham diabetes risk score	Mean (95% CI) HbA <sub>1c</sub> $-0.26\%$ ( $-0.27$ to $-0.24$ ) in intervention vs control $-0.18\%$ ( $-0.19$ to $-0.16$ ), FBG $-0.41\text{ mmol/L}$ ( $-0.44$ to $-0.12$ ) in intervention vs $-0.21\text{ mmol/L}$ ( $-0.15$ to $-0.10$ ) in control, all outcomes significantly greater in intervention than control ( $P<.001$ )
Fischer, 2016 [29]	RCT for 12 months; effectiveness of text message-supported T2DM prevention program	Obese and overweight adults from United States without prediabetes, English or Spanish speaking; intervention n=82, control n=81	6 text messages per week: skills, problem solving, motivation, stress reduction, recipes, web links to additional resources, PA promotion; weekly self-reported weight; eligible for individual motivational phone health coaching; control: usual care	Change in weight; percentage of participants with ≥3% or 5% weight loss, changes in HbA <sub>1c</sub> and systolic BP <sup>g</sup> , costs per participant	Weight (95% CI) in intervention $-1.2\text{ kg}$ ( $-2.5$ to $0.1$ ) vs control $-0.3\text{ kg}$ ( $-1.2$ to $0.7$ ), $P=.05$ ; 3% weight loss absolute difference between groups 17.0%, $P=.02$ ; no significant difference for 5% weight loss; HbA <sub>1c</sub> in intervention $-0.09\%$ ( $-0.2$ to $0.0$ ) vs control $0.19\%$ ( $-0.1$ to $0.5$ ), systolic BP in intervention $0.35\text{ mmHg}$ ( $-2.8$ to $3.5$ ) vs control $6.4\text{ mmHg}$ ( $3.2$ to $9.5$ )
Fukuoka, 2015 [30]	RCT for 5 months; effectiveness of mobile app-based intervention for T2DM prevention	Overweight adults aged ≥35 years from United States at high risk of diabetes; intervention n=30; control n=31	2-week run-in period before randomizing; all daily step count via pedometer; intervention: mobile version of Diabetes Prevention Program, 6 in-person sessions, app: diaries for self-monitoring of weight, PA, and caloric intake, daily reminders and messages; control: pedometer only	% change in weight and BMI; hip circumference, BP, lipid profile, glucose levels, step count, PA, caloric and fat intake	Weight (95% CI) $-6.8\%$ ( $-12.2$ to $-1.4$ ) in intervention vs $0.3\%$ ( $-2.7$ to $3.3$ ) in control; BMI $-6.6\%$ ( $-12.3$ to $-0.9$ ) in intervention vs $0.3\%$ ( $-2.7$ to $3.3$ ) in control; both $P<.001$ ; also significant differences in hip circumference, BP, step count, and PA for intervention vs control; no effect on lipid profile, glucose levels, caloric or fat intake

First author, date, reference	Study design and duration; objectives	Study population	Intervention and comparator	Outcomes	Main results
Ramachandran, 2013 [31]	RCT for 2 years; effectiveness of SMS text messaging to reduce incidence of T2DM in men with impaired glucose tolerance	Indian men aged 35-55 years with impaired glucose tolerance; intervention n=271, control n=266	All at baseline: healthy lifestyle education and written information on diet and PA, lifestyle changes prescribed; intervention: frequent reinforcing text messages, content tailored to baseline behavior; control: usual care	Incidence of T2DM; BMI, WC, BP, lipid profile, energy intake, PA	Cumulative T2DM incidence: intervention 18%, control 27%; differences in mean change (95% CI): BMI $-0.05 \text{ kg/m}^2$ ( $-0.46$ to $0.37$ ); WC $0.04 \text{ cm}$ ( $-0.56$ to $0.64$ ); systolic BP $0.04 \text{ mmHg}$ ( $-0.96$ to $1.03$ ); diastolic BP $-0.07 \text{ mmHg}$ ( $-0.64$ to $0.49$ ); total cholesterol $0.01 \text{ mmol/L}$ ( $-0.08$ to $0.10$ ); HDL $0.033 \text{ mmol/L}$ ( $0.011$ to $0.054$ ); triglycerides $-0.08 \text{ mmol/L}$ ( $-0.17$ to $-0.06$ ); energy intake $-43.7 \text{ kcal}$ ( $-65.5$ to $-22.0$ ); PA score $-1.0$ ( $-2.0$ to $0.0$ )

<sup>a</sup>RCT: randomized controlled trial.

<sup>b</sup>PA: physical activity.

<sup>c</sup>WC: waist circumference.

<sup>d</sup>FBG: fasting blood glucose.

<sup>e</sup>HbA<sub>1c</sub>: glycated hemoglobin.

<sup>f</sup>HDL: high-density lipoprotein.

<sup>g</sup>BP: blood pressure.

## Results of the Synthesis

### Summary

We synthesized the results of the data extraction according to the PICOS system. Table 4 summarizes CVD and T2DM data

individually and in total. For each parameter, we provide a count, as well as the list of relevant references.

**Table 4.** Synthesis of findings.

Finding	Cardiovascular disease		Type 2 diabetes		Total (n)
	No. of studies	Reference	No. of studies	Reference	
Target population					
General population	1	[24]	— <sup>a</sup>	—	1
At risk of the disease	2	[23,25]	6	[26-31]	8
Location					
Spain	1	[24]	—	—	1
United States	1	[23]	4	[27-30]	5
Germany	—	—	1	[26]	1
Latin America	1	[25]	—	—	1
India	—	—	1	[31]	1
Intervention delivery					
SMS text messaging	2	[23,25]	2	[29,31]	4
WhatsApp	1	[24]	—	—	1
Mobile app	—	—	4	[26-28,30]	4
Comparator					
Usual care	3	[23-25]	3	[26,29,31]	6
Waitlist	—	—	2	[27,28]	2
Pedometer only	—	—	1	[30]	1
Face-to-face training	1	[24]	—	—	1
Outcomes <sup>b</sup>					
Weight loss	1	[25]	3	[27,28,30]	4
Reduced BMI	1	[25]	3	[27,28,30]	4
Reduced waist circumference	—	—	2	[27,28]	2
Lower fasting blood glucose/glycated hemoglobin	N/A <sup>c</sup>	—	1	[28]	1
Improved diet	2	[23,25]	1	[31]	3
Improved physical activity	—	—	2	[30,31]	2
Improved blood pressure	—	—	1	[30]	1
Study design					
Randomized controlled trial	1	[25]	5	[27-31]	6
Nonrandomized controlled trial	2	[23,24]	1	[26]	3

<sup>a</sup>—: data not available.<sup>b</sup>Statistically significant compared with control group.<sup>c</sup>N/A: not applicable.

### Participants

The CVD studies were conducted in Spain, the United States, and Latin America. For the T2DM studies, 1 was conducted in Germany, 1 in India, and 4 in the United States. All studies had small to medium samples, ranging from 32 to 637 participants. For CVD, 2 of the 3 studies targeted populations at higher risk of developing CVD [23,25], whereas the study by Muntaner-Mas et al [24] included healthy people. For T2DM, all studies focused on populations at increased risk of the disease.

### Interventions

The duration of the interventions varied from 10 weeks to 2 years. In 4 studies the participants received text messages [23,25,29,31], in 1 study the intervention was delivered via WhatsApp [24], and in the remaining 4 studies a specifically developed mobile phone app was involved [26-28,30]. Only 1 intervention was delivered fully automated [28]; all other interventions included human involvement [23-27,29-31].



### Comparators

Of the studies, 6 used usual care as the control group. In 1 trial, the control group received pedometers only [30]; 1 study had a second comparator group, additional to usual care, which received face-to-face training sessions [24]; and 2 studies used waitlist controls [27,28], meaning that the control group received the intervention after the intervention group had completed it.

### Outcomes

The mobile phone interventions led to statistically significant weight loss compared with the control group in 4 studies [25,27,28,30], ranging from a difference of  $-0.66$  kg ( $P=.04$ ) over 12 months [25] to  $-6.2$  kg for the intervention compared with  $0.3$  kg for the control group ( $P<.001$ ) over 5 months [30]. The same studies reported a statistically significant decrease in BMI compared with the control group [25,27,28,30], ranging from a difference of  $-0.3$  kg/m<sup>2</sup> ( $P=.02$ ) over 12 months [25] to  $-2.2$  kg/m<sup>2</sup> for the intervention compared with  $0.1$  kg/m<sup>2</sup> for the control group ( $P<.001$ ) over 5 months [30]. A smaller waist circumference due to the intervention was measured in 2 T2DM studies [27,28], from  $-4.56$  cm for the intervention compared with  $-2.22$  cm ( $P<.001$ ) for the control group over 6 months [28] to a cross-level interaction of  $-4.9$  cm (95% CI  $-7.5$  to  $-2.6$ ) over 3 months [27]. One T2DM study reported statistically significantly lower fasting blood glucose ( $-0.41$  mmol/L in the intervention compared with  $-0.12$  mmol/L in the control group;  $P<.001$ ) and glycated hemoglobin levels ( $-0.26\%$  in the intervention compared with  $-0.18\%$  in the control group;  $P<.001$ ) over 6 months [28]. Statistically significantly greater changes in the lipid profile were observed in the intervention group than in the control group in 2 of the T2DM trials [28,31], from a difference in mean change of high-density lipoprotein cholesterol of  $0.033$  mmol/L (95% CI  $0.011$  to  $0.054$ ) and triglycerides of  $-0.080$  mmol/L (95% CI  $-0.17$  to  $-0.06$ ) over

2 years [31] to a triglyceride to high-density lipoprotein ratio of  $-0.21$  in the intervention compared with  $0.21$  in the control group ( $P=.04$ ) over 6 months [28]. Improved diet patterns that were statistically superior to the control group were observed in 3 studies [23,25,31], of which 2 studies aimed at CVD prevention. Improvements in physical activity were reported in 2 T2DM studies [30,31]. Blood pressure was statistically significantly improved in the intervention groups compared with the control group in 1 T2DM study [30].

### Study Design

A total of 6 studies were RCTs [25,27-31]; the remaining 3 were non-RCTs [23,24,26].

### Results of Literature Quality Assessment

All RCTs used acceptable methods for randomization [25,27-31], but in none of the studies were the participants blinded to the design, which is an inherent problem with this type of intervention. Figure 3 summarizes the risk of bias for the RCTs. Of the 6 studies, 3 ensured blinding of the study personnel [25,27,28] and 3 ensured the blinding of the outcome assessors [25,28,31]. Apart from the study by Fukuoka et al [30], all RCTs published study protocols on the ClinicalTrials.gov database. Overall, due to performance bias, all studies were at high risk of bias.

Of the 3 non-RCTs [23,24,26], the study by Muntaner-Mas et al [24] was at moderate risk of bias, the study by Gore et al [23] was at high risk of bias, and the study by Arens et al [26] was at critical risk of bias. Figure 4 summarizes the risk of bias for the non-RCTs. The biggest issue with the study by Arens et al [26] was that missing data were not handled adequately, putting the study at critical risk of bias. We assessed the study by Gore et al [23] to be at high risk of confounding because some of the measurements that were used to control for confounding were based on nonvalidated questionnaires.

**Figure 3.** Risk-of-bias summary table for the randomized controlled trials. The upper 1 is a cardiovascular disease study and the remainder are type 2 diabetes studies.

Rubinstein et al [25]	+	+	-	+	+	+	+
Bender et al [27]	+	+	-	+	-	+	?
Block et al [28]	+	+	-	+	+	+	?
Fischer et al [29]	+	+	-	?	?	+	+
Fukuoka et al [30]	+	+	-	?	?	+	?
Ramachandran et al [31]	+	+	-	-	+	?	+
	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants (performance bias)	Blinding of personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)

Legend: + low risk of bias, - serious risk of bias, ? unclear.

**Figure 4.** Risk-of-bias summary table for the nonrandomized controlled trials. The upper 2 are cardiovascular disease studies and the lower 1 is a type 2 diabetes study.

Gore et al [23]	-	+	+	+	+	o	o	-
Muntaner-Mas et al [24]	o	+	+	+	+	o	o	o
Arens et al [26]	-	+	+	+	---	-	o	---
	Risk of confounding	Risk of selection bias	Risk of misclassification bias	Risk of bias due to deviations from intended interventions	Risk of bias due to missing data	Risk of bias in measurement outcomes	Risk of bias in the selection of the reported results	Overall predicted direction of bias for this outcome

Legend: + low risk of bias, o moderate risk of bias, - serious risk of bias, --- critical risk of bias.

## Discussion

### Principal Findings

We identified only a small number ( $n=9$ ) of articles that fulfilled the preset inclusion and exclusion criteria. We assessed most of the studies to be at high risk of bias. Additionally, 3 studies were underpowered (sample size  $<100$ ), and 2 studies had short follow-up times ( $<6$  months). Ideally, to show the effectiveness in reducing the risk of CVD or T2DM, the studies should have reported disease incidence rates. The only study that did this was that by Ramachandran et al [31], with their primary outcome being a decrease of T2DM incidence due to the SMS text messaging intervention over 2 years. Block et al [28] reported the percentage of people with metabolic syndrome as defined by the International Diabetes Federation Task Force on Epidemiology and Prevention [32]. Block et al [28] also measured change in the Framingham 8-year diabetes risk score [33], and Gore et al [23] measured change in the Framingham 10-year CVD risk score [20]. All other studies reported single risk factors rather than multivariable absolute risk of disease. None of the identified studies directly targeted tobacco smoking cessation or responsible alcohol intake. Rubinstein et al [25] mentioned in their article that their original protocol included both lifestyle factors, but they were later excluded. According to the authors, alcohol intake was considered a sensitive matter requiring face-to-face interactions, while tobacco smoking was excluded because supposedly, compared with physical activity and diet, it had less effect on the onset of hypertension and was more difficult to target via a mobile health intervention [25]. Overall, there were some positive findings suggesting that mobile health-based interventions can achieve at least small to moderate reductions in CVD and T2DM risk, although these were based on weak evidence.

### Strengths and Limitations

The strength of this literature review was that it followed the PRISMA statement. We systematically searched several databases to identify all relevant published articles. Further, we conducted a manual search through the snowballing technique. For the title and abstract screening, a 10% random sample of all retrieved articles was validated by a second researcher, and 2 reviewers independently performed the full article selection. However, only 1 researcher conducted the database search, the data extraction, and the risk-of-bias assessment. Although we a priori restricted the search to English- and German-language articles, we did not exclude any articles because they were not available in these 2 languages. We did not perform a meta-analysis due to the small number of publications that met the inclusion criteria and the differences in their interventions and outcome measures.

### Comparison With Prior Work

Previous mobile health research has focused more on self-management of chronic diseases than on prevention. In their systematic review and meta-analysis, Wu et al [34] investigated the effectiveness of mobile phone apps for diabetes self-management (including prediabetes, gestational diabetes, type 1 diabetes, and T2DM). They identified 3 studies that targeted prediabetes, 2 of which we also included in this review.

The overall conclusion of Wu et al [34] was that there was evidence for the effectiveness of app interventions in T2DM self-management, but not for prediabetes. Lunde et al [35] conducted a systematic review looking at various types of noncommunicable diseases and lifestyle advice. Most of the identified studies (8 out of 9) targeted T2DM patients for whom the authors measured improvements in lifestyle factors, particularly reduced glycated hemoglobin levels (in 5 of the 8 studies). For CVD patients, Lunde et al [35] found only 2 relevant articles, and these were without statistically significant improvements in any of the outcomes of interest (weight, BMI, waist circumference, physical activity, and quality of life). A systematic review by Coorey et al [36] focused on self-management of CVD via mobile apps, in which they concluded that short-term improvements in behavior and risk factors were possible but there was insufficient evidence for long-term effects. Alessa et al [37] reported from their systematic review of 21 studies that mobile apps could reduce blood pressure, although the evidence originated mainly from studies that had a high risk of bias.

Palmer et al [14] conducted a systematic review of noncommunicable disease prevention through smoking cessation, alcohol reduction, physical activity, and diet using mobile technology. In total, they found 71 articles, but only 2 of the studies were aimed at the combination of physical activity, diet, and smoking cessation, with both studies targeting secondary prevention of CVD. Among the studies they reviewed, 8 RCTs focused on alcohol reduction but did not include any other lifestyle advice, with these studies specifically targeting heavy drinkers [14]. In general, it appears that many interventions are designed to provide advice for 1 or 2 behavioral risk factors that are associated with increased chronic disease risk, whereas there were only a few evaluation studies of comprehensive mobile health interventions addressing the 4 common behavioral risk factors (ie, tobacco smoking, excessive alcohol consumption, physical inactivity, and poor diet) [14]. Noble et al [13] stated in their systematic review that there were clustering patterns between the 4 behavioral risk factors—tobacco smoking, excessive alcohol consumption, physical inactivity, and poor diet—which indicated similar or the same reasons causing these behaviors. Hence, the authors suggested that future interventions should apply a holistic approach instead of targeting single risk factors. Similarly, Geller et al [38] called for future research studies to focus on improved lifestyles, meaning a change in multiple health behaviors rather than 1, even if it might be harder to achieve. Meader et al [39] reported in their systematic review that targeting smoking simultaneously with other behaviors resulted in negative outcomes for diet and physical activity, suggesting that it might be more beneficial to apply a sequential approach. In a Cochrane review published in 2016, Whittaker et al [40] stated that studies have demonstrated that mobile phone-based interventions can be effective in achieving smoking cessation over 6 months, particularly SMS text messaging in high-income countries.

### Implications and Future Directions

Most studies that were conducted according to the review's inclusion criteria were at high risk of bias. This review only

considered studies of multirisk factor interventions, which resulted in only 9 studies being included. There is a lack of research evaluating interventions that address the 4 common behavioral risk factors (ie, tobacco smoking, excessive alcohol consumption, physical inactivity, and poor diet) in a single mobile health intervention. Researchers may have preferred to focus on 1 risk factor at a time due to simplicity for participants and clarity of intervention-outcome relationships. Hence, future

studies should further explore the use of mobile technology for primary disease prevention, by applying a rigorous study design.

## Conclusions

According to the findings of this systematic review, evidence for the effectiveness of mobile health-based interventions in reducing the risk of CVD and T2DM is scarce due to the quality of the included studies and the small effects that were measured. This highlights the need for further high-quality research to investigate the potential of mobile health interventions.

## Acknowledgments

This research was supported by a joint stipend from the University of New South Wales and the Commonwealth Scientific and Industrial Research Organisation.

## Authors' Contributions

VHB participated in research design, data collection, data analysis, and writing of the manuscript. SL participated in data collection, data analysis, and revision of the manuscript. MV, MB, and MH contributed to research design, data collection, data analysis, and revision of the manuscript. All authors provided final approval of the version to be published.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Search strategy.

[DOC File, 48 KB - [jmir\\_v22i10e21159\\_app1.doc](#)]

## References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: [10.1016/S0140-6736\(18\)32203-7](#)] [Medline: [30496103](#)]
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018 Dec 10;392(10159):1789-1858 [FREE Full text] [doi: [10.1016/S0140-6736\(18\)32279-7](#)] [Medline: [30496104](#)]
3. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018 Nov 10;392(10159):1923-1994 [FREE Full text] [doi: [10.1016/S0140-6736\(18\)32225-6](#)] [Medline: [30496105](#)]
4. Gellert C, Schöttker B, Brenner H. Smoking and all-cause mortality in older people: systematic review and meta-analysis. *Arch Intern Med* 2012 Jun 11;172(11):837-844. [doi: [10.1001/archinternmed.2012.1397](#)] [Medline: [22688992](#)]
5. Wood AM, Kaptoge S, Butterworth AS, Willeit P, Warnakula S, Bolton T, Emerging Risk Factors Collaboration/EPIC-CVD/UK Biobank Alcohol Study Group. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *Lancet* 2018 Dec 14;391(10129):1513-1523. [doi: [10.1016/S0140-6736\(18\)30134-X](#)] [Medline: [29676281](#)]
6. Chudasama YV, Khunti KK, Zaccardi F, Rowlands AV, Yates T, Gillies CL, et al. Physical activity, multimorbidity, and life expectancy: a UK Biobank longitudinal study. *BMC Med* 2019 Jun 12;17(1):108 [FREE Full text] [doi: [10.1186/s12916-019-1339-0](#)] [Medline: [31186007](#)]
7. Zhang B, Zhao Q, Guo W, Bao W, Wang X. Association of whole grain intake with all-cause, cardiovascular, and cancer mortality: a systematic review and dose-response meta-analysis from prospective cohort studies. *Eur J Clin Nutr* 2018 Dec;72(1):57-65. [doi: [10.1038/ejcn.2017.149](#)] [Medline: [29091078](#)]
8. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Geneva, Switzerland: World Health Organization; 2013. URL: [https://apps.who.int/iris/bitstream/handle/10665/94384/9789241506236\\_eng.pdf?sessionid=75D5722BE3A73FAE254F414C58825563?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/94384/9789241506236_eng.pdf?sessionid=75D5722BE3A73FAE254F414C58825563?sequence=1) [accessed 2019-12-19]



9. World Health Organization. WHO guideline recommendations on digital interventions for health system strengthening. Geneva, Switzerland: World Health Organization; 2019. URL: <https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1> [accessed 2019-12-19]
10. Wilson K. Mobile cell phone technology puts the future of health care in our hands. CMAJ 2018 Apr 03;190(13):E378-E379. [doi: [10.1503/cmaj.180269](https://doi.org/10.1503/cmaj.180269)] [Medline: [29615420](https://pubmed.ncbi.nlm.nih.gov/29615420/)]
11. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Washington, DC: Pew Research Center; 2019 Feb 05. URL: <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/> [accessed 2019-10-12]
12. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: [10.1007/s13142-011-0021-7](https://doi.org/10.1007/s13142-011-0021-7)] [Medline: [21796270](https://pubmed.ncbi.nlm.nih.gov/21796270/)]
13. Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. Prev Med 2015 Dec;81:16-41. [doi: [10.1016/j.ypmed.2015.07.003](https://doi.org/10.1016/j.ypmed.2015.07.003)] [Medline: [26190368](https://pubmed.ncbi.nlm.nih.gov/26190368/)]
14. Palmer M, Sutherland J, Barnard S, Wynne A, Rezel E, Doel A, et al. The effectiveness of smoking cessation, physical activity/diet and alcohol reduction interventions delivered by mobile phones for the prevention of non-communicable diseases: a systematic review of randomised controlled trials. PLoS One 2018;13(1):e0189801 [FREE Full text] [doi: [10.1371/journal.pone.0189801](https://doi.org/10.1371/journal.pone.0189801)] [Medline: [29304148](https://pubmed.ncbi.nlm.nih.gov/29304148/)]
15. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [Medline: [19622511](https://pubmed.ncbi.nlm.nih.gov/19622511/)]
16. Lefebvre C, Manheimer E, Glanville J. Boolean operators (AND, OR and NOT). In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Oxford, UK: Cochrane Collaboration; 2011. URL: [https://handbook-5-1.cochrane.org/index.htm#chapter\\_6/6\\_4\\_7\\_boolean\\_operators\\_and\\_or\\_and\\_not.htm](https://handbook-5-1.cochrane.org/index.htm#chapter_6/6_4_7_boolean_operators_and_or_and_not.htm) [accessed 2019-09-26]
17. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ 2005 Nov 05;331(7524):1064-1065 [FREE Full text] [doi: [10.1136/bmj.38636.593461.68](https://doi.org/10.1136/bmj.38636.593461.68)] [Medline: [16230312](https://pubmed.ncbi.nlm.nih.gov/16230312/)]
18. World Health Organization. Cardiovascular disease. Geneva, Switzerland: WHO; 2019. URL: [https://www.who.int/cardiovascular\\_diseases/about\\_cvd/en/](https://www.who.int/cardiovascular_diseases/about_cvd/en/) [accessed 2019-06-19]
19. World Health Organization. Diabetes. Geneva, Switzerland: WHO; 2018 Oct 30. URL: <https://www.who.int/news-room/fact-sheets/detail/diabetes> [accessed 2019-12-16]
20. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008 Feb 12;117(6):743-753. [doi: [10.1161/CIRCULATIONAHA.107.699579](https://doi.org/10.1161/CIRCULATIONAHA.107.699579)] [Medline: [18212285](https://pubmed.ncbi.nlm.nih.gov/18212285/)]
21. Cochrane Collaboration. The Cochrane Collaboration's tool for assessing risk of bias: 8.5.1 Overview. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Oxford, UK: Cochrane Collaboration; 2011. URL: [https://handbook-5-1.cochrane.org/index.htm#chapter\\_8/8\\_5\\_1\\_overview.htm](https://handbook-5-1.cochrane.org/index.htm#chapter_8/8_5_1_overview.htm) [accessed 2019-10-24]
22. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919. [Medline: [27733354](https://pubmed.ncbi.nlm.nih.gov/27733354/)]
23. Gore MO, Krantz MJ, Albright K, Beaty B, Coronel-Mockler S, Bull S, et al. A controlled trial of mobile short message service among participants in a rural cardiovascular disease prevention program. Prev Med Rep 2019 Mar;13:126-131 [FREE Full text] [doi: [10.1016/j.pmedr.2018.11.021](https://doi.org/10.1016/j.pmedr.2018.11.021)] [Medline: [30568871](https://pubmed.ncbi.nlm.nih.gov/30568871/)]
24. Muntaner-Mas A, Vidal-Conti J, Borràs PA, Ortega FB, Palou P. Effects of a Whatsapp-delivered physical activity intervention to enhance health-related physical fitness components and cardiovascular disease risk factors in older adults. J Sports Med Phys Fitness 2017;57(1-2):90-102. [doi: [10.23736/S0022-4707.16.05918-1](https://doi.org/10.23736/S0022-4707.16.05918-1)] [Medline: [26364690](https://pubmed.ncbi.nlm.nih.gov/26364690/)]
25. Rubinstein A, Miranda JJ, Beratarrechea A, Diez-Canseco F, Kanter R, Gutierrez L, et al. Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Jan;4(1):52-63. [doi: [10.1016/S2213-8587\(15\)00381-2](https://doi.org/10.1016/S2213-8587(15)00381-2)] [Medline: [26653067](https://pubmed.ncbi.nlm.nih.gov/26653067/)]
26. Arens JH, Hauth W, Weissmann J. Novel app- and web-supported diabetes prevention program to promote weight reduction, physical activity, and a healthier lifestyle: observation of the clinical application. J Diabetes Sci Technol 2018 Jul;12(4):831-838 [FREE Full text] [doi: [10.1177/1932296818768621](https://doi.org/10.1177/1932296818768621)] [Medline: [29584454](https://pubmed.ncbi.nlm.nih.gov/29584454/)]
27. Bender MS, Cooper BA, Flowers E, Ma R, Arai S. Filipinos Fit and Trim - a feasible and efficacious DPP-based intervention trial. Contemp Clin Trials Commun 2018 Dec;12:76-84 [FREE Full text] [doi: [10.1016/j.conctc.2018.09.004](https://doi.org/10.1016/j.conctc.2018.09.004)] [Medline: [30294699](https://pubmed.ncbi.nlm.nih.gov/30294699/)]
28. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: [10.2196/jmir.4897](https://doi.org/10.2196/jmir.4897)] [Medline: [26499966](https://pubmed.ncbi.nlm.nih.gov/26499966/)]



29. Fischer HH, Fischer IP, Pereira RI, Furniss AL, Rozwadowski JM, Moore SL, et al. Text message support for weight loss in patients with prediabetes: a randomized clinical trial. *Diabetes Care* 2016 Dec;39(8):1364-1370. [doi: [10.2337/dc15-2137](https://doi.org/10.2337/dc15-2137)] [Medline: [26861922](https://pubmed.ncbi.nlm.nih.gov/26861922/)]
30. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. *Am J Prev Med* 2015 Aug;49(2):223-237 [FREE Full text] [doi: [10.1016/j.amepre.2015.01.003](https://doi.org/10.1016/j.amepre.2015.01.003)] [Medline: [26033349](https://pubmed.ncbi.nlm.nih.gov/26033349/)]
31. Ramachandran A, Snehalatha C, Ram J, Selvam S, Simon M, Nanditha A, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. *Lancet Diabetes Endocrinol* 2013 Nov;1(3):191-198. [doi: [10.1016/S2213-8587\(13\)70067-6](https://doi.org/10.1016/S2213-8587(13)70067-6)] [Medline: [24622367](https://pubmed.ncbi.nlm.nih.gov/24622367/)]
32. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, International Diabetes Federation Task Force on Epidemiology and Prevention, National Heart, Lung, and Blood Institute; American Heart Association, American Heart Association, World Heart Federation, International Atherosclerosis Society, International Association for the Study of Obesity. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 2009 Oct 20;120(16):1640-1645. [doi: [10.1161/CIRCULATIONAHA.109.192644](https://doi.org/10.1161/CIRCULATIONAHA.109.192644)] [Medline: [19805654](https://pubmed.ncbi.nlm.nih.gov/19805654/)]
33. Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 2007 May 28;167(10):1068-1074. [doi: [10.1001/archinte.167.10.1068](https://doi.org/10.1001/archinte.167.10.1068)] [Medline: [17533210](https://pubmed.ncbi.nlm.nih.gov/17533210/)]
34. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: [10.2196/12297](https://doi.org/10.2196/12297)] [Medline: [30664494](https://pubmed.ncbi.nlm.nih.gov/30664494/)]
35. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. *J Med Internet Res* 2018 May 04;20(5):e162 [FREE Full text] [doi: [10.2196/jmir.9751](https://doi.org/10.2196/jmir.9751)] [Medline: [29728346](https://pubmed.ncbi.nlm.nih.gov/29728346/)]
36. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: systematic review with meta-synthesis of quantitative and qualitative data. *Eur J Prev Cardiol* 2018 Mar;25(5):505-521. [doi: [10.1177/2047487317750913](https://doi.org/10.1177/2047487317750913)] [Medline: [29313363](https://pubmed.ncbi.nlm.nih.gov/29313363/)]
37. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile apps to support the self-management of hypertension: systematic review of effectiveness, usability, and user satisfaction. *JMIR Mhealth Uhealth* 2018 Jul 23;6(7):e10723 [FREE Full text] [doi: [10.2196/10723](https://doi.org/10.2196/10723)] [Medline: [30037787](https://pubmed.ncbi.nlm.nih.gov/30037787/)]
38. Geller K, Lippke S, Nigg CR. Future directions of multiple behavior change research. *J Behav Med* 2017 Feb;40(1):194-202. [doi: [10.1007/s10865-016-9809-8](https://doi.org/10.1007/s10865-016-9809-8)] [Medline: [27785652](https://pubmed.ncbi.nlm.nih.gov/27785652/)]
39. Meader N, King K, Wright K, Graham HM, Petticrew M, Power C, et al. Multiple risk behavior interventions: meta-analyses of RCTs. *Am J Prev Med* 2017 Jul;53(1):e19-e30 [FREE Full text] [doi: [10.1016/j.amepre.2017.01.032](https://doi.org/10.1016/j.amepre.2017.01.032)] [Medline: [28258777](https://pubmed.ncbi.nlm.nih.gov/28258777/)]
40. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2016;4:CD006611. [doi: [10.1002/14651858.CD006611.pub4](https://doi.org/10.1002/14651858.CD006611.pub4)] [Medline: [27060875](https://pubmed.ncbi.nlm.nih.gov/27060875/)]

## Abbreviations

**CVD:** cardiovascular disease

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**PROSPERO:** International Prospective Register of Systematic Reviews

**RCT:** randomized controlled trial

**T2DM:** type 2 diabetes mellitus

**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 06.06.20; peer-reviewed by C Hockham, V Haldane; comments to author 25.06.20; revised version received 10.08.20; accepted 02.09.20; published 29.10.20.*

*Please cite as:*

Buss VH, Leesong S, Barr M, Varnfield M, Harris M

Primary Prevention of Cardiovascular Disease and Type 2 Diabetes Mellitus Using Mobile Health Technology: Systematic Review of the Literature

*J Med Internet Res* 2020;22(10):e21159

URL: <http://www.jmir.org/2020/10/e21159/>

doi: [10.2196/21159](https://doi.org/10.2196/21159)

PMID: [33118936](https://pubmed.ncbi.nlm.nih.gov/33118936/)

©Vera Helen Buss, Stuart Leesong, Margo Barr, Marlien Varnfield, Mark Harris. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

[Viewpoint](#)

# Private Video Consultation Services and the Future of Primary Care

Chris Salisbury<sup>1</sup>, MSc, MD, FRCGP; Anna Quigley<sup>2</sup>, LLB (Hons), MSc; Nick Hex<sup>3</sup>, BA, CIPFA; Camille Aznar<sup>2</sup>, PhD

<sup>1</sup>Centre for Academic Primary Care, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

<sup>2</sup>Ipsos MORI Social Research Institute, London, United Kingdom

<sup>3</sup>York Health Economics Consortium, York, United Kingdom

**Corresponding Author:**

Chris Salisbury, MSc, MD, FRCGP

Centre for Academic Primary Care, Department of Population Health Sciences

Bristol Medical School

University of Bristol

Canynge Hall,

39 Whatley Road

Bristol, BS8 2PS

United Kingdom

Phone: 44 1173314530

Email: [c.salisbury@bristol.ac.uk](mailto:c.salisbury@bristol.ac.uk)

## Abstract

In many countries, private companies provide primary care services based predominantly on offering video consultations via smartphones. One example is Babylon GP at Hand (BGPaH), which offers video consultations to National Health Service patients, 24 hours a day, and has grown rapidly in London over the last 3 years. The development of this type of service has been controversial, particularly in the United Kingdom, but there has been little formal published evaluation of these services in any country. This paper outlines the main controversies about the use of privately provided video consultation services for primary care and shows how they are informed by the limited evaluations that have been conducted, particularly the evaluation of BGPaH. This paper describes the advantages of these services in terms of convenience, speed of access, the ability to consult without traveling or face-to-face patient-doctor contact, and the possibility of recruiting doctors who cannot work in conventional settings or do not live near the patients. It also highlights the concerns and uncertainties about quality and safety, demand, fragmentation of care, impact on other health services, efficiency, and equity. There are questions about whether private primary care services based on video consultations have a sustainable business model and whether they will undermine other health care providers. During the recent COVID-19 pandemic, the use of video consulting has become more widespread within conventional primary care services, and this is likely to have lasting consequences for the future delivery of primary care. It is important to understand the extent to which lessons from the evaluation of BGPaH and other private services based on a *video-first* model are relevant to the use of video consulting within conventional general practices, and to consider the advantages and disadvantages of these developments, before video consultation-based services in primary care become more widely established.

(*J Med Internet Res* 2020;22(10):e19415) doi:[10.2196/19415](https://doi.org/10.2196/19415)

**KEYWORDS**

remote consultation; primary health care; general practice; delivery of health care; access to health care; mobile phone

## Primary Care Based on Video Consultations

In the United Kingdom, recent policy has strongly promoted *digital-first* health care under the National Health Service (NHS), arguing that this will improve access and convenience for patients while also increasing efficiency [1-3]. Private

companies such as Livi and Pushdoctor have contracted with groups of general practices to offer video consultations free to patients under the NHS. In these models, video consultations are provided by doctors working anywhere in the country, in addition to conventional care provided by patients' usual doctors. The best-known example in the United Kingdom is Babylon GP at Hand (BGPaH), which provides almost all services on behalf of a general practice in London, under a

digital-first model whereby almost all consultations are initially conducted using a smartphone [4]. Key features of BGPaH are shown in [Textbox 1](#). Driven by a strong advertising campaign,

this service has grown very rapidly in London since 2017 and has recently expanded to Birmingham.

**Textbox 1.** Summary of Babylon GP at Hand. GP: general practitioner.

Key features of Babylon GP at Hand:

- Provides a web-based symptom checker.
- Offers video or telephone consultations with a General Practitioner (GP) 24 hours a day, usually within 2 hours of making a request.
- Babylon GP at Hand (BGPaH) GPs make referrals and provide prescriptions in the same way as conventional GPs, sending them electronically.
- When necessary (usually only after a video or telephone consultation), patients can have a face-to-face consultation at one of 7 locations across London or one in Birmingham.
- Patients have to deregister from their previous National Health Service (NHS) general practice when they register with BGPaH.
- Under the NHS GP Choice policy, patients can register with BGPaH wherever they live, providing they can travel to one of the clinic sites within 40 min.
- If patients need urgent help and are not able to visit a clinic, or need a home visit and do not live near a BGPaH clinic, they have to contact the NHS 111 national telephone helpline to arrange help from other NHS providers.
- Patients who need frequent face-to-face consultations or have difficulty traveling to a clinic are encouraged by BGPaH to register elsewhere.
- Patients requiring a face-to-face consultation for specific purposes (eg, a cervical smear) contact a support team to make an appointment at one of the five clinics.
- Consultations are provided by over a hundred GPs, mostly working from home from anywhere in the country. Some GPs work at a central hub in London, which also manages test results and repeat prescriptions.

Similar developments have occurred in many other countries, with private companies offering video consultations via smartphones instead of conventional face-to-face appointments for primary care. Examples include Dr on Demand in the United States; Curon in Japan; Ping An Good Doctor in China; and KRY, which operates in several European countries. Often backed by venture capital, these companies are expanding rapidly into new regions. The aim of this paper is to consider the implications of these developments for the future of primary care. It particularly focuses on the United Kingdom, and lessons learnt from the evaluation of the BGPaH service in London, but also refers to relevant findings from evaluations of similar services in other countries.

## Controversies

The growth of private companies offering video consultations for primary care has been controversial. Concerns have been expressed about the safety and effectiveness of managing patients by telephone and video, and there have also been concerns that services focused on speed of access will undermine continuity of care [5,6]. New commercial services could compete for staff at a time when shortages of doctors threaten the viability of some general practices. An underlying concern in the United Kingdom is that the involvement of commercial companies offering video consultations undermines the public service ethos of the NHS and increases fragmentation of services.

However, others have argued that video consulting represents the future of health care [7,8] and that criticisms are based on resistance to change, with health services having been slow to exploit the potential of technology to improve efficiency and convenience for consumers [7,9].

Some of these arguments have conflated 3 different issues. First, there is the issue of private companies providing a new model of care, either in competition with conventional services or as an adjunct to conventional care. Second, there is debate about the appropriate role of video consultations in primary care. The third issue is about *digital-first* access models in which patients are expected to have an initial contact by video, telephone, or asynchronous web-based message before being offered a face-to-face consultation only when necessary.

These 3 issues are related because the new model provided by services such as BGPaH depends on video consultations and digital-first approaches. However, the recent COVID-19 pandemic has greatly accelerated the use of remote consultations and digital-first approaches by conventional general practices. In the past, the introduction of these approaches has been slow and patchy, but during the pandemic, doctors in many different countries have rapidly turned to telephone and video consultations because of the need to manage large numbers of patients without face-to-face contact [10]. Video consultations may be offered as an option or within a digital-first model. In this paper, we focus on private companies offering primary care on a *video-first* model, particularly lessons learnt from the evaluation of BGPaH. However, some of the points discussed will also be relevant to the provision of video consultations within a conventional primary care model.

## Evaluation

Despite the rapid growth of companies offering primary care video consultation services and the controversy surrounding them, there has been almost no rigorous evaluation of these services in any of the countries where they have been established. Some of the most detailed evidence comes from

an independent evaluation of BGPaH, which was commissioned by the NHS and published on the website of a local clinical commissioning group in May 2019 [11]. **Textbox 2** summarizes the key components of this evaluation, which was designed to

understand the impact of BGPaH on patients, the general practitioner (GP) workforce, and the wider health care system. The evaluation report and an annex providing details of the full results are available elsewhere [11,12].

**Textbox 2.** Components of Babylon GP at Hand evaluation. GP: general practitioner.

- *Patient experience survey:* Web-based survey of 1452 Babylon GP at Hand (BGPaH) patients (1452/23,073, 6.29% response rate) to quantitatively assess their experience, compared with a similar patient cohort responding to the National Health Service (NHS) GP Patient Survey [13], matched using propensity score matching.
- *Qualitative interviews:* In-depth interviews with 12 general practitioners, a nurse and a member of operational staff from BGPaH, 32 current patients (including 3 in the process of deregistering), and 4 patients who had deregistered from BGPaH, along with site visits to the BGPaH hub and one clinic.
- *Analysis of secondary data:* Analysis of routine data sets about activity and synthesis of NHS England analytical work using nationally held data.

There were significant limitations to the BGPaH evaluation due to the availability of data and the short timescale of the evaluation. The patient survey had a very low response rate, which raises concerns about the representativeness of the views expressed. It was not possible to make comparisons between responders and nonresponders, but survey participants were matched using propensity score matching to patients with similar characteristics receiving conventional care. Babylon was only able to provide limited data about patients' face-to-face consultations, and no information was available about patients' presenting problems, health outcomes, or referrals to other NHS services.

Despite these limitations, the BGPaH evaluation provides useful information as a case study to inform current debates about the implementation of video-first services in primary care. In the United States, there have been brief reports about digital primary care based on video consultations from organizations such as Kaiser Permanente [14-16], Jefferson Health [17,18], and the Veteran's Administration [19]. Several digital health platforms have been established in Sweden, and these have been reported in a descriptive evaluation, but this does not distinguish between video, audio, and text-based web-based consultations [20]. None of these reports represent a comprehensive evaluation, and all have significant limitations, but they do provide insights into some of the controversies.

## Advantages of Video Consultation Services in Primary Care

A primary care service based mainly on a digital-first model (including video and/or telephone consultations) has obvious

advantages. It provides patients with convenient access to health care advice without leaving home or work, offering savings in time and travel costs to patients and productivity gains for society [21]. It improves access in sparsely populated areas where physically getting to a GP may be problematic. As clinicians can work from anywhere, this makes it possible to recruit doctors who cannot work in conventional settings and to provide care in areas where it is difficult to recruit sufficient GPs. In the recent COVID-19 pandemic, remote consultations by video and telephone made it possible to avoid direct patient-doctor contact.

## Patient Experience

Patients choosing to use private video consultation services appear to be generally very satisfied with their care, although it is important to remember the limitations of the evidence (particularly nonresponse bias). **Table 1** shows responses from participants in the survey of users of BGPaH [11] compared with matched patients responding to the national GP Patient Survey [13]. Overall, 71.49% (1038/1452) of BGPaH survey respondents described the quality of care they received at BGPaH as being better than that they received at their previous general practice, whereas only 10.67% (155/1452) described it as worse. This greater satisfaction appeared to be driven by the speed and convenience of the service, the quality of interpersonal care, and the length of consultations. These advantages were not specific to video consultations, with many patients choosing telephone consultations with BGPaH.



**Table 1.** Patients' experience of Babylon GP at Hand compared with matched respondents from the national GP Patient Survey (summary data). GP: general practitioner.

Question	Response option	GPPS <sup>a</sup> respondents <sup>b</sup>		BGPaH <sup>c</sup> respondents	
		n (%)	N	n (%)	N
Overall, how would you describe your experience of making an appointment?	Very good	258 (23.39)	1103	717 (65.48)	1095
Overall experience	Very good	402 (34.69)	1159	682 (58.19)	1172
Thinking about the reason for your last general practice appointment, were your needs met?	Yes, definitely	622 (55.83)	1114	663 (63.69)	1041
At your last appointment, how good was the doctor at giving you enough time <sup>d</sup>	Very good	361 (40.7)	887	546 (61.8)	884
At your last appointment, how good was the doctor at listening to you <sup>d</sup>	Very good	395 (44.6)	885	571 (64.6)	884
At your last appointment, how good was the doctor at treating you with care and concern <sup>d</sup>	Very good	385 (43.7)	882	553 (62.8)	880
During your last general practice appointment, did you have confidence and trust in the health care professional you saw or spoke to? <sup>d</sup>	Yes, definitely	507 (57.2)	886	638 (71.9)	887

<sup>a</sup>GPPS: General Practitioner Patient Survey.

<sup>b</sup>Propensity matched sample of patients from the GP Patient Survey [13] who were resident in London and had similar characteristics to respondents in the BGPaH survey. Matching variables used data available in both surveys, including age, gender, ethnicity, religion, sexuality, work status, whether there were children in the household, whether the respondent was a carer, and whether the respondent had a limiting long-term illness.

<sup>c</sup>BGPaH: Babylon GP at Hand.

<sup>d</sup>Responses from patients who had a consultation with a general practitioner.

Less positive views were expressed in interviews by some patients who needed face-to-face consultations or had complex needs (Textbox 3). Although some patients were concerned about a lack of continuity of care, most had actively chosen speed of access over continuity. Some patients experienced problems with integration between the digital and face-to-face

services, particularly issues such as providing urine samples, following up test results, and arranging hospital referrals.

High levels of patient satisfaction have also been found in evaluations conducted in the United States [16-19], similarly driven by the increased speed and convenience of video consultations compared with waiting for a face-to-face appointment.

**Textbox 3.** The views of Babylon GP at Hand patients and doctors. GP: general practitioner.

Patients	
•	"I can only get doctors' appointments on weekdays during office hours so I couldn't register with the doctors. So I thought this was the best idea I've ever seen."
•	"I would never have made an effort to see a GP - too much effort to leave work, make appointment and plan around it. Having phone consultations makes things a lot easier - convenience, evening weekends, not having to miss work or having to trek to where GP is."
•	"You click a button and get an appointment in 10 minutes, that's their selling point, but there's nothing for long-term health management."
•	App experience great for arranging telephone or video consultation but when referred for a face-to-face examination it was a nightmare getting an appointment! No convenient times outside of working hours during the week. Had to wait ages to get something that would work for me."
General Practitioners	
•	"I was dissatisfied with my practice that I'd been at. (...) I felt underappreciated. I felt overworked and under-remunerated for the services that I gave in my own time, whereas all of those things were a bit different. The ethos here is very, very different."
•	"General practice has suffered a lot in the last ten years, and it's really hard for GPs to do the job they want to do as best as they can. (...) I was doing two years in different salaried jobs and it was really hard and I didn't feel valued, didn't feel like I was delivering what I wanted to for patients. Actually, this was the first time as a GP I've actually felt like I can do that, I can deliver the care."

## Impact on Emergency Departments

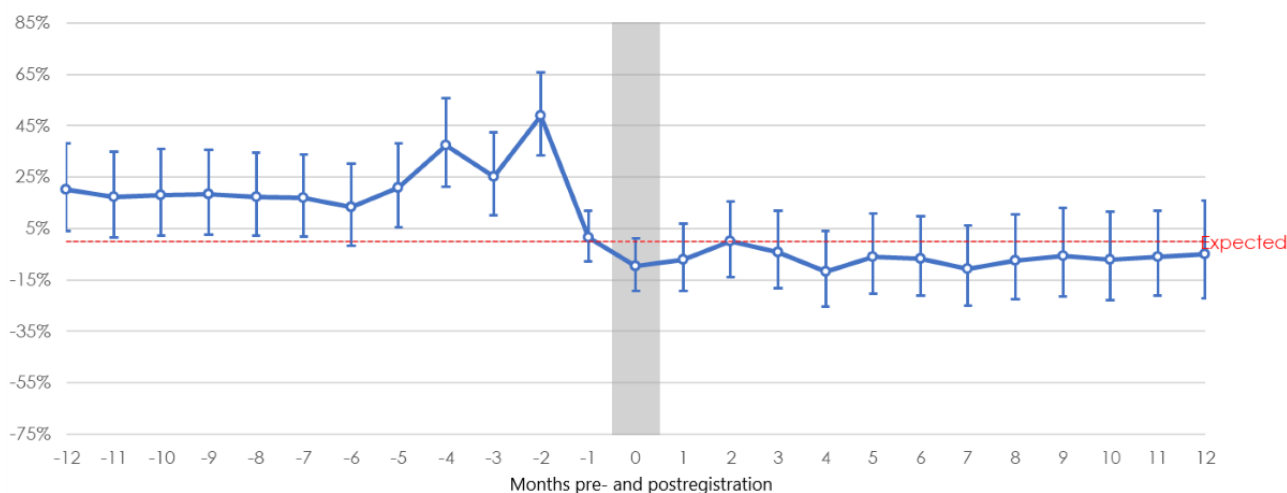
One potential advantage of video consultation services is that by improving access to general practice, they could reduce the use of emergency departments for primary care problems.

However, there is very limited evidence to support this claim. For example, although analysis in the BGPaH study showed an apparent reduction in the use of emergency departments (Figure 1), the evaluators noted that there were a number of potential confounders in the data that meant robust conclusions could not

be drawn. It was also notable that BGPaH patients had below average use of hospital admissions and outpatient attendances

both before and after registration (age-sex standardized), reflecting that they are a generally healthy population.

**Figure 1.** Babylon GP at Hand emergency department attendance rates compared with newly registered London patients (age-sex standardized). Source: National Health Service England. GP: general practitioner.



## Uptake

One of the most notable features of BGPaH is the speed at which it has attracted patients, when other studies have highlighted very low uptake of video consultations. For example, Kaiser Permanente has offered video consultations since 2014, but between 2015 and 2017, these accounted for less than 1% of all office consultations [15]. In the United Kingdom, uptake of video consultations was very low when offered by conventional general practices before the COVID-19 pandemic [22,23]. In Sweden, the use of digital consultations has been increasing for several years but still represented only 2% of all primary care consultations by 2017 [20]. A consistent finding has been that introduction of video consulting has been impeded by technical problems [23,24]. The apparent success of BGPaH may be because the whole process of care has been redesigned around a new technology-enabled model with a very strong focus on convenience for patients. Experience of technological innovations in other settings shows that these are only likely to improve services when they are used to enable whole-system redesign, rather than being bolted on to existing systems and processes [9]. It is notable how, in the recent pandemic, many of the issues that impeded the introduction of video consultations seem to have been rapidly overcome through a combination of technical innovation, entrepreneurship, an obvious incentive for clinicians to avoid face-to-face contact, and a pragmatic relaxation of regulatory objections.

## Equity

The vast majority (94%) of patients registered with BGPaH are aged under 45 years; two-thirds live in more affluent areas; and 84% are in paid work, and they have substantially fewer long-term health conditions than age-sex matched controls, except for asthma. Currently, only 0.28% of BGPaH patients are aged over 70 years, compared with 12% in an average English general practice [25]. These findings suggest that this

type of video-first service is particularly suitable for patients with less complex health care needs but also raise questions about equity of access in relation to need. In Kaiser Permanente, the use of electronic tools is highest among young adults and white patients and lower among older patients and those from African American or Latino ethnic groups [14]. In Sweden, there is a significantly positive relationship between the use of digital care and affluence [20].

## Quality and Safety

A number of authors have questioned the quality and safety of digital consultations [26,27] and argued that these should require the same level of evidence as other medical innovations [26]. Although there are published frameworks for the evaluation of digital health systems, which include the need for evidence about usability, effectiveness, and safety in real-world implementation [27,28], this evidence is largely absent for video consultation systems. A recent Scottish study concluded that video consultations dealt with fewer problems and achieved lower scores on measures of consultation quality than face-to-face consultations and were most suitable for simple problems [23]. Babylon has published their own analysis of the safety of the automated triage software used in their digital health platform [29], but this was not peer-reviewed and the findings have been subject to criticism [27]. However, the evaluation of BGPaH highlights several positive aspects of the service [11], and a recent inspection by the Care Quality Commission rated BGPaH as good overall [30].

## Supply-Induced Demand

There is some evidence that, by reducing the threshold to access care, video consultation services may lead to increased consultation rates through supply-induced demand. In a patient survey, 47% of BGPaH patients reported that they used the service more regularly than their previous practice (8% said

they used it less regularly), and interviews with patients also gave examples of increased frequency of use. The estimated mean annual consultation rate of patients with BGPaH was 4.3 consultations per annum, higher than the national average for patients of their age [31]. In Kaiser Permanente, the number of *virtual visits* after offering internet, mobile, and video access routes increased from 4.1 million in 2008 to 10.5 million in 2013, but there was no associated reduction in face-to-face consultations [14]. Similarly, an evaluation of BlueShield in California demonstrated a substantial increase in utilization and health care spending [32]. This increased use of primary service utilization may be an advantage if it addresses unmet need but is problematic if it diverts time and resources away from patient groups with greater needs. This is an issue of cost-effectiveness but cannot be assessed without information about patient outcomes or impacts on other health services.

### *Cream-Skimming?*

In the United Kingdom, the growth of BGPaH has been particularly controversial because of claims that it threatens the viability of conventional practices by targeting young and healthy patients, leaving conventional practices to see older patients and those with more complex problems who generate most work [33,34]. However, this is arguably a problem with the way in which GPs in the United Kingdom are paid under the NHS, rather than the BGPaH model itself. The current payment formula for general practice in the United Kingdom takes into account the age and sex and the health needs of a patient population at an area level, but it does not effectively adjust for the different health needs of individual patients. The problem of how to fairly reimburse health providers in ways that support the appropriate use of video consultations without increasing health system costs is being debated in the United Kingdom [35,36] and several countries [14,20]. In relation to the argument about cream-skimming, it is notable that BGPaH does not appear to attract patients who generate less work—if anything, their patients have higher consultation rates than average patients of similar age and sex, although they have fewer long-term health conditions.

### *Is the Improved Access Due to Video Technology or Greater Investment?*

The main aim of primary care services based on a video-first model is to improve access to care. Video (and telephone) consultations make it possible for patients to see a doctor without traveling and losing time from other activities. It is noteworthy that although the BGPaH model defaults to a video consultation, the evaluation found that almost as many patients had used a telephone consultation as had used a video consultation. The main advantage for patients appeared to be the ability to get any sort of consultation quickly, rather than the mode of consultation. It is not obvious why providing consultations by video should reduce delays to speak to a doctor, if this is due to a shortage of GP availability. GPs working for BGPaH see fewer patients in a 4-hour shift than GPs working in most conventional NHS practices. It is possible that the key issue that improves access is the re-engineering of the

organizational model and access pathway, rather than the use of technology. Alternatively, it could be that the improved access is due to greater investment in additional capacity and that similar improvements could be made with increased investment in conventional care.

### *The Sustainability of New Models of Care*

One of the reasons that the growth of BGPaH has been controversial in the United Kingdom has been a concern about a gradually increasing provision of segmented services by private companies and a lack of transparency about their business models. There are assumptions that digital primary care will be more efficient than conventional models of care, but the evidence available so far does not support this hypothesis [37]. No details of the costs and productivity of staff were available for the evaluation of BGPaH because they were considered commercially confidential. Although the per-consultation cost of a video consultation at Kaiser Permanente is less than a face-to-face consultation, increased consultation rates mean that cost savings have not materialized [14]. Many of the new video-based primary care services are provided by start-up companies that are trying to disrupt what they see as entrenched service models. Backed by large investments of cash [38], they are competing to gain market share, and it is not clear whether their current operations are profitable or instead being provided as loss-leaders. In the context of a state-funded service such as the NHS, it is important to understand far more about the costs of delivering smartphone-based services to ensure that they are sustainable, before they become indispensable after having displaced other services.

A further concern has been that private services such as BGPaH could strip the NHS of staff by offering more flexible and convenient working conditions. In the BGPaH evaluation, we observed that BGPaH GPs tended to be younger than the general GP workforce, and more than two-thirds are female. Many were attracted to the job because it offered a good work-life balance. Most were very satisfied with working for Babylon, contrasting the autonomy, training, flexibility, and good working conditions with the long hours and increasing workloads they experienced in conventional general practice. These findings may have important implications for how conventional services need to improve working conditions to retain staff.

### *Continuity of Care*

Conventionally, the existence of a relationship built over time between a patient and their doctor has been viewed as one of the defining features of general practice. Continuity of care is associated with benefits in patient satisfaction, adherence to medical advice, use of hospital services, and health outcomes [39]. However, continuity of care in conventional general practices in the United Kingdom is declining rapidly [40], and in the evaluation of BGPaH, it was clear that continuity was not a priority for most patients choosing this service. The growth of BGPaH and other private video consultation services is one manifestation of a change in relationships between patients and their general practitioners, with many patients choosing a more

impersonal but potentially more convenient service. It is possible that there will be a loss of the benefits associated with continuity of care. Although no such adverse effects were observed in the evaluation of BGPaH, it will be important to monitor these effects over a longer period.

## Implications of Private Video-First Consultation Services for the Future of Primary Care

Evaluation suggests that video-first consultation services for primary care are associated with high levels of satisfaction among the type of patients who prefer this approach, and this is generally a young and healthy population. However, this may not necessarily provide a model for designing services for the majority of patients who use primary care and who may have different needs or priorities. Private video consultation-based services provide some aspects of primary care to only a limited extent and rely on the continued existence of conventional practices to provide these. Smartphone-based services focusing on addressing specific presenting complaints may not provide opportunistic health care, which is a central role of general practice. Older patients with chronic diseases make up the bulk of the general practice workload. These patients are much less likely to choose to use video consultations than younger patients [11], and they also place a higher priority on relational continuity of care from a doctor they know [41].

One possible scenario for the future of primary care is the development of different types of services tailored to the needs of different population groups. However, this also has several potential disadvantages, including the loss of a single point of contact for health care provided by conventional general practice; the potential for individuals to fall into the gaps between different services, each with a defined remit; and the need for patients to transfer between services as their health care needs change. This particularly applies to older patients, but is also relevant to many other patient groups, particularly those who are vulnerable or have complex needs.

Due to the COVID-19 pandemic, many conventional general practices have rapidly implemented similar *digital-first* access pathways and remote consultations, including by video. What lessons can we learn from services such as BGPaH that are relevant to video consulting in conventional general practices? First, it is feasible to use video consultations for many types of consultations, and it is likely that a greater use of these consultations will continue in the future. Second, although the rapid implementation of video consulting during the pandemic has been driven by expediency, it should not be assumed that it necessarily provides optimal or more efficient care. Third, it is notable that even in a service such as BGPaH marketed around the offer of video consultations, many patients prefer the simpler technology of the telephone. Fourth, it is important to distinguish between video consultations as a technology and a service delivery model in which video consultation is the default mode of consulting. If digital consultations are offered as a choice, within a general practice that offers all types of consultation, they should improve access for patients who prefer this option. However, if patients are required to use digital services before accessing a face-to-face consultation (the *digital-first* model as used by BGPaH), this could favor patients with the fewest needs and increase health inequalities.

Private companies providing video consultations for primary care have acted as a challenge to conventional general practice, and the COVID-19 pandemic has further accelerated similar changes in service delivery in these conventional practices. However, there remain questions about important issues such as the quality, safety, and efficiency of video consultations; supply-induced demand; continuity of care; and the appropriateness of video consultations for different types of problems and patients, and these questions require further research. There are also concerns about the implications for the wider health care system of private companies providing video-first services that tend to address the needs of a specific segment of the population. In advance of evidence about these key issues, the growth of private companies providing video-first primary care should be managed cautiously and accompanied by much more extensive and rigorous independent evaluation than has been undertaken so far.

## Acknowledgments

The authors would like to thank Josh Keith who contributed to the design of the evaluation of BGPaH, Joanna Barry and Heather Ashford who provided research and administrative support, James Hebblethwaite (Information Manager, Hammersmith and Fulham Clinical Commissioning Group [CCG]) and Dilwyn Sheers and Matthew Dearing (on behalf of NHS England Data, Analysis and Intelligence Service) who designed and carried out a range of analyses, including the analysis of impacts on secondary care presented in the text and in Figure 1. CS is a National Institute for Health Research (NIHR) Senior Investigator. The views expressed in this paper are those of the authors and not necessarily those of the NIHR, Hammersmith and Fulham CCG, NHS England, or the Department of Health and Social Care. The BGPaH evaluation was commissioned by the Hammersmith and Fulham Clinical Commissioning Group and NHS England. It was conducted independently by Ipsos MORI and the York Health Economics Consortium with CS.

## Authors' Contributions

CS was an academic adviser to the evaluation of BGPaH and provided topic and methodological expertise and contributed to the design of the evaluation and interpretation of the findings. AQ was responsible for the evaluation and delivery of the project on behalf of Ipsos MORI. NH led the economic aspects of the evaluation. CA was responsible for day-to-day delivery supported by



a delivery team of research executives and team members drawn from across Ipsos MORI's operations teams. CS drafted this paper, and all authors provided intellectual input and approved the final version. CS is the guarantor for this paper.

## Conflicts of Interest

None declared.

## References

1. NHS Long Term Plan. 2019. URL: <https://www.longtermplan.nhs.uk/> [accessed 2020-09-21]
2. Five Year Forward View. National Health Service. National Health Service; 2014. URL: <https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf> [accessed 2020-09-21]
3. Investment and Evolution: a Five-year Framework for GP Contract Reform to Implement the NHS Long Term Plan. 2019. URL: <https://www.england.nhs.uk/wp-content/uploads/2019/01/gp-contract-2019.pdf> [accessed 2020-09-21]
4. GP at Hand. 2019. URL: <https://www.gpathand.nhs.uk/> [accessed 2020-09-21]
5. Marshall M, Shah R, Stokes-Lampard H. Online consulting in general practice: making the move from disruptive innovation to mainstream service. *Br Med J* 2018 Mar 26;360:k1195. [doi: [10.1136/bmj.k1195](https://doi.org/10.1136/bmj.k1195)] [Medline: [29581174](https://pubmed.ncbi.nlm.nih.gov/29581174/)]
6. Bhatti N. Seeing a GP on a Smartphone Sounds Wonderful - But It's Not. *The Guardian*. 2017 Nov 16. URL: <https://www.theguardian.com/healthcare-network/views-from-the-nhs-frontline/2017/nov/16/seeing-gp-smartphone-sounds-wonderful-its-not> [accessed 2020-09-21]
7. Gerada C. GPs have been Luddites for too long and must embrace new technology. *Br Med J* 2018 May 22;361:k2206. [doi: [10.1136/bmj.k2206](https://doi.org/10.1136/bmj.k2206)] [Medline: [29789309](https://pubmed.ncbi.nlm.nih.gov/29789309/)]
8. The Future of Healthcare: Our Vision for Digital, Data and Technology in Health and Care. Department of Health and Social Care. 2018 Oct 17. URL: <https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care> [accessed 2020-09-21]
9. Imison C, Castle-Clarke S, Watson R, Edwards N. Delivering the Benefits of Digital Health Care. Nuffield Trust. 2016 Feb. URL: <https://www.nuffieldtrust.org.uk/files/2017-01/delivering-the-benefits-of-digital-technology-web-final.pdf> [accessed 2020-09-21]
10. Webster P. Virtual health care in the era of COVID-19. *Lancet* 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30818-7](https://doi.org/10.1016/S0140-6736(20)30818-7)] [Medline: [32278374](https://pubmed.ncbi.nlm.nih.gov/32278374/)]
11. Quigley A, Hex N, Aznar C, Salisbury C. Evaluation of Babylon GP at Hand: Final Evaluation Report. 2019 May. URL: <https://www.hammersmithfulhamccg.nhs.uk/media/156123/Evaluation-of-Babylon-GP-at-Hand-Final-Report.pdf> [accessed 2020-09-21]
12. Ipsos MORI, York Health Economics Consortium, Salisbury C. Evaluation of Babylon GP at hand: Technical Annexes to the final report. 2019 May. URL: <https://www.hammersmithfulhamccg.nhs.uk/media/156120/Evaluation-of-Babylon-GP-at-Hand-Annexes.pdf> [accessed 2020-09-21]
13. GP Patient Survey. NHS England and NHS Improvement. 2019. URL: <http://www.gp-patient.co.uk/surveysandreports> [accessed 2020-09-21]
14. Pearl R. Kaiser Permanente Northern California: current experiences with internet, mobile, and video technologies. *Health Aff (Millwood)* 2014 Feb;33(2):251-257. [doi: [10.1377/hlthaff.2013.1005](https://doi.org/10.1377/hlthaff.2013.1005)] [Medline: [24493768](https://pubmed.ncbi.nlm.nih.gov/24493768/)]
15. Reed ME, Parikh R, Huang J, Ballard DW, Barr I, Wargon C. Real-time patient-provider video telemedicine integrated with clinical care. *N Engl J Med* 2018 Oct 11;379(15):1478-1479. [doi: [10.1056/NEJMc1805746](https://doi.org/10.1056/NEJMc1805746)] [Medline: [30304654](https://pubmed.ncbi.nlm.nih.gov/30304654/)]
16. Reed ME, Huang J, Parikh R, Millman A, Ballard DW, Barr I, et al. Patient-provider video telemedicine integrated with clinical care: patient experiences. *Ann Intern Med* 2019 Aug 6;171(3):222-224. [doi: [10.7326/M18-3081](https://doi.org/10.7326/M18-3081)] [Medline: [31035294](https://pubmed.ncbi.nlm.nih.gov/31035294/)]
17. Powell RE, Stone D, Hollander JE. Patient and health system experience with implementation of an enterprise-wide telehealth scheduled video visit program: mixed-methods study. *JMIR Med Inform* 2018 Feb 13;6(1):e10 [FREE Full text] [doi: [10.2196/medinform.8479](https://doi.org/10.2196/medinform.8479)] [Medline: [29439947](https://pubmed.ncbi.nlm.nih.gov/29439947/)]
18. Powell RE, Henstenburg JM, Cooper G, Hollander JE, Rising KL. Patient perceptions of telehealth primary care video visits. *Ann Fam Med* 2017 May;15(3):225-229. [doi: [10.1370/afm.2095](https://doi.org/10.1370/afm.2095)] [Medline: [28483887](https://pubmed.ncbi.nlm.nih.gov/28483887/)]
19. Wennergren J, Munshi I, Fajardo AD, George VV. Implementation of clinical video telemedicine (CVT) within a VA medical center is cost effective and well received by veterans. *Int J Clin Med* 2014;05(12):711-716. [doi: [10.4236/ijcm.2014.512097](https://doi.org/10.4236/ijcm.2014.512097)]
20. Ekman B, Thulesius H, Wilkens J, Lindgren A, Cronberg O, Arvidsson E. Utilization of digital primary care in Sweden: descriptive analysis of claims data on demographics, socioeconomic, and diagnoses. *Int J Med Inform* 2019 Jul;127:134-140. [doi: [10.1016/j.ijmedinf.2019.04.016](https://doi.org/10.1016/j.ijmedinf.2019.04.016)] [Medline: [31128825](https://pubmed.ncbi.nlm.nih.gov/31128825/)]
21. Ekman B. Cost analysis of a digital health care model in Sweden. *Pharmacoecon Open* 2018 Sep;2(3):347-354 [FREE Full text] [doi: [10.1007/s41669-017-0059-7](https://doi.org/10.1007/s41669-017-0059-7)] [Medline: [29623633](https://pubmed.ncbi.nlm.nih.gov/29623633/)]



22. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. The potential of alternatives to face-to-face consultation in general practice, and the impact on different patient groups: a mixed-methods case study. *Health Services and Delivery Research* 2018;6(20). [doi: [10.3310/hsdr06200](https://doi.org/10.3310/hsdr06200)] [Medline: [29889485](https://pubmed.ncbi.nlm.nih.gov/29889485/)]
23. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. *Br J Gen Pract* 2019 Sep;69(686):e595-e604. [doi: [10.3399/bjgp19X704573](https://doi.org/10.3399/bjgp19X704573)] [Medline: [31262846](https://pubmed.ncbi.nlm.nih.gov/31262846/)]
24. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. *J Med Internet Res* 2018 Apr 17;20(4):e150 [FREE Full text] [doi: [10.2196/jmir.9897](https://doi.org/10.2196/jmir.9897)] [Medline: [29625956](https://pubmed.ncbi.nlm.nih.gov/29625956/)]
25. Patients Registered at a Gp Practice. NHS Digital. URL: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/general-practice-data-hub/patients-registered-at-a-gp-practice> [accessed 2020-09-21]
26. The Lancet. Is digital medicine different? *Lancet* 2018 Jul 14;392(10142):95. [doi: [10.1016/S0140-6736\(18\)31562-9](https://doi.org/10.1016/S0140-6736(18)31562-9)] [Medline: [30017135](https://pubmed.ncbi.nlm.nih.gov/30017135/)]
27. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. *Lancet* 2018 Nov 24;392(10161):2263-2264. [doi: [10.1016/S0140-6736\(18\)32819-8](https://doi.org/10.1016/S0140-6736(18)32819-8)] [Medline: [30413281](https://pubmed.ncbi.nlm.nih.gov/30413281/)]
28. Evidence Standards Framework for Digital Health Technologies. National Institute for Health and Care Excellence. 2019. URL: <https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies> [accessed 2020-09-21]
29. Razzaki S, Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, et al. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. *ArXiv* 2018 [FREE Full text]
30. GP at Hand Inspection Report. Care Quality Commission. 2019 May 21. URL: [https://www.cqc.org.uk/sites/default/files/new\\_reports/AAA1557.pdf](https://www.cqc.org.uk/sites/default/files/new_reports/AAA1557.pdf) [accessed 2020-09-21]
31. Hobbs FD, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. *Lancet* 2016 Jun 04;387(10035):2323-2330 [FREE Full text] [doi: [10.1016/S0140-6736\(16\)00620-6](https://doi.org/10.1016/S0140-6736(16)00620-6)] [Medline: [27059888](https://pubmed.ncbi.nlm.nih.gov/27059888/)]
32. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-to-consumer telehealth may increase access to care but does not decrease spending. *Health Aff (Millwood)* 2017 Mar 1;36(3):485-491. [doi: [10.1377/hlthaff.2016.1130](https://doi.org/10.1377/hlthaff.2016.1130)] [Medline: [28264950](https://pubmed.ncbi.nlm.nih.gov/28264950/)]
33. Roland M. General practice by smartphone. *Br Med J* 2019 Jul 24;366:l4713. [doi: [10.1136/bmj.l4713](https://doi.org/10.1136/bmj.l4713)] [Medline: [31340952](https://pubmed.ncbi.nlm.nih.gov/31340952/)]
34. Rosen R. Are disruptive innovators in GP provision strengthening or weakening the NHS? *Br Med J* 2017 Nov 29;359:j5470. [doi: [10.1136/bmj.j5470](https://doi.org/10.1136/bmj.j5470)] [Medline: [29187352](https://pubmed.ncbi.nlm.nih.gov/29187352/)]
35. Digital-first Primary Care: Policy Consultation on Patient Registration, Funding and Contracting Rules. NHS England. 2019. URL: <https://www.engage.england.nhs.uk/consultation/digital-first-primary-care-policy/> [accessed 2020-09-21]
36. Digital-First Primary Care and Its Implications for General Practice Payments. NHS England. 2018 Jul 04. URL: [https://www.engage.england.nhs.uk/survey/digital-first-primary-care/user\\_uploads/digital-first-access-to-gp-care-engagement-v2.pdf](https://www.engage.england.nhs.uk/survey/digital-first-primary-care/user_uploads/digital-first-access-to-gp-care-engagement-v2.pdf) [accessed 2020-09-21]
37. Salisbury C, Murphy M, Duncan P. The impact of digital-first consultations on workload in general practice: modeling study. *J Med Internet Res* 2020 Jun 16;22(6):e18203 [FREE Full text] [doi: [10.2196/18203](https://doi.org/10.2196/18203)] [Medline: [32543441](https://pubmed.ncbi.nlm.nih.gov/32543441/)]
38. Field M. NHS Chatbot App Babylon Raises \$550m With Saudi Backing as MP Demands Inquiry Into Funding. *The Telegraph*. 2019 Aug 02. URL: <https://www.telegraph.co.uk/technology/2019/08/02/nhs-chatbot-app-babylon-raises-550m-help-saudi-backing/> [accessed 2020-09-21]
39. Gray DJ, Sidaway-Lee K, White E, Thorne A, Evans PH. Continuity of care with doctors-a matter of life and death? A systematic review of continuity of care and mortality. *BMJ Open* 2018 Jun 28;8(6):e021161. [doi: [10.1136/bmjopen-2017-021161](https://doi.org/10.1136/bmjopen-2017-021161)] [Medline: [29959146](https://pubmed.ncbi.nlm.nih.gov/29959146/)]
40. Levene LS, Baker R, Walker N, Williams C, Wilson A, Bankart J. Predicting declines in perceived relationship continuity using practice deprivation scores: a longitudinal study in primary care. *Br J Gen Pract* 2018 Jun;68(671):e420-e426. [doi: [10.3399/bjgp18X696209](https://doi.org/10.3399/bjgp18X696209)] [Medline: [29739778](https://pubmed.ncbi.nlm.nih.gov/29739778/)]
41. Aboulghate A, Abel G, Elliott MN, Parker RA, Campbell J, Lyratzopoulos G, et al. Do English patients want continuity of care, and do they receive it? *Br J Gen Pract* 2012 Aug;62(601):e567-e575. [doi: [10.3399/bjgp12X653624](https://doi.org/10.3399/bjgp12X653624)] [Medline: [22867681](https://pubmed.ncbi.nlm.nih.gov/22867681/)]

## Abbreviations

**BGPaH:** Babylon GP at Hand  
**CCG:** Clinical Commissioning Group  
**GP:** general practitioner  
**NHS:** National Health Service  
**NIHR:** National Institute for Health Research

*Edited by G Eysenbach; submitted 20.04.20; peer-reviewed by J Kronenfeld, S Shaw, M Kang; comments to author 12.06.20; revised version received 29.07.20; accepted 13.08.20; published 01.10.20.*

*Please cite as:*

Salisbury C, Quigley A, Hex N, Aznar C

Private Video Consultation Services and the Future of Primary Care

J Med Internet Res 2020;22(10):e19415

URL: <https://www.jmir.org/2020/10/e19415>

doi: [10.2196/19415](https://doi.org/10.2196/19415)

PMID: [32812887](https://pubmed.ncbi.nlm.nih.gov/32812887/)

©Chris Salisbury, Anna Quigley, Nick Hex, Camille Aznar. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 01.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Viewpoint

# The Need for Sustainable Teleconsultation Systems in the Aftermath of the First COVID-19 Wave

Guido Giunti<sup>1,2</sup>, MD, PhD; Richard Goossens<sup>2</sup>, Prof Dr; Antoinette De Bont<sup>3</sup>, Prof Dr; Jacob J Visser<sup>4</sup>, MD, PhD; Mark Mulder<sup>4</sup>, MD, PhD; Stephanie C E Schuit<sup>4</sup>, Prof Dr

<sup>1</sup>University of Oulu, Oulu, Finland

<sup>2</sup>TU Delft, Delft, Netherlands

<sup>3</sup>Erasmus University Rotterdam, Rotterdam, Netherlands

<sup>4</sup>Erasmus University Medical Center, Rotterdam, Netherlands

**Corresponding Author:**

Guido Giunti, MD, PhD

University of Oulu

Pentti Kaiteran katu 1

Oulu, 90014

Finland

Phone: 358 294480000

Email: [drguidogiunti@gmail.com](mailto:drguidogiunti@gmail.com)

## Abstract

The physical and social distancing measures that have been adopted worldwide because of COVID-19 will probably remain in place for a long time, especially for senior adults, people with chronic conditions, and other at-risk populations. Teleconsultations can be useful in ensuring that patients continue to receive clinical care while reducing physical crowding and avoiding unnecessary exposure of health care staff. Implementation processes that typically take months of planning, budgeting, pilot testing, and education were compressed into days. However, in the urgency to deal with the present crisis, we may be forgetting that the introduction of digital health is not exclusively a technological issue, but part of a complex organizational change problem. This viewpoint offers insight regarding issues that rapidly adopted teleconsultation systems may face in a post-COVID-19 world.

(*J Med Internet Res* 2020;22(10):e21211) doi:[10.2196/21211](https://doi.org/10.2196/21211)

**KEYWORDS**

telemedicine; COVID-19; telehealth; teleconsultation; exposure; software; digital health; organization

## Introduction

The ongoing COVID-19 pandemic has dramatically changed the global landscape in general, and health care systems in particular. Over the past few months, varying states of lockdown have been declared worldwide [1], and teleconsultation has become the “new normal” way of accessing health care.

The social distancing measures that have been adopted worldwide because of COVID-19 [2] will probably remain in place for a long time, especially for senior adults, people with chronic conditions, and other at-risk populations. Teleconsultations can be useful to ensure that patients continue to receive clinical care while reducing physical crowding [3] and avoiding unnecessary exposure of health care staff [4].

Despite their promise, digital tools had proven difficult to implement until very recently [5]. Implementation processes that would typically take months of planning, budgeting, pilot

testing, and education are now compressed into days. However, in the urgency to deal with the present crisis, we risk forgetting that the introduction of digital health is not exclusively a technological issue, but part of a complex organizational change problem [6].

In theory, all that is required for a telehealth visit is the right equipment and platform. In practice, the same barriers that existed in 2019 still exist today: proper protocols for patient care, equipment deployment, attitudes, and even legislation [7-10]. The lack of digital health training of health professionals is an increasingly recognized barrier. Although there are initiatives worldwide that provide specialization training on medical informatics (such as the American Medical Informatics Association in the United States, with nearly 1700 board-certified professionals [11], and England's efforts with the Topol review [12]), Europe lags behind, with less than one-third of medical schools covering these topics [9].

## The Netherlands and Erasmus University Medical Center

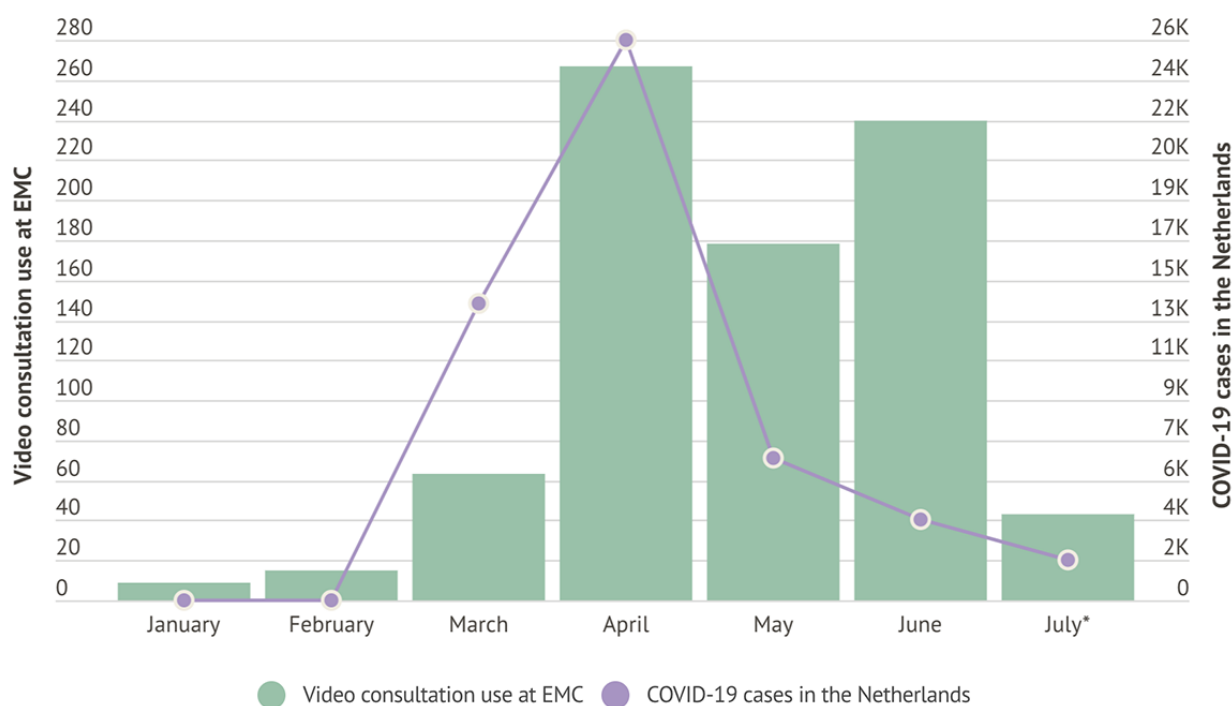
On February 26, 2020, the first patient with COVID-19 was diagnosed in the Netherlands and admitted to hospital [13]. From then onwards, the number of infections, hospital admissions, and deaths due to COVID-19 infection increased rapidly, with a peak incidence of around 1500 daily new confirmed infections, over 500 daily hospital admissions, and close to 200 daily deaths two months after the admission of patient zero [14]. In that same two-month period, the Netherlands went into a state of “lockdown,” as did many other countries in the world. Standard hospital care was drastically reduced to deal with the surge of patients with COVID-19, especially in ICUs. At the time of writing, the Netherlands seems to have weathered this first wave of COVID-19, with current infection rates, hospital admissions, and deaths down to a trickle and lockdown restrictions increasingly eased.

Erasmus University Medical Center (Erasmus MC) played a central role in managing the COVID-19 crisis in the Netherlands by serving as the national coordination center for the distribution

of patients with COVID-19 nationwide. The medical center is based in Rotterdam, and is one of the largest and most authoritative scientific university medical centers in Europe. Erasmus MC is a 1125-bed academic medical center offering specialized health care services in neurosurgery, cardiothoracic surgery, pediatric oncology, and neonatal and intensive care; in addition, it is a level I trauma center. A Microsoft-based, integrated electronic health record (EHR) solution called HiX (ChipSoft) has been fully deployed since June 2017.

As became common practice worldwide at the beginning of the COVID-19 pandemic, most of our clinical work had to be reorganized and priorities changed. In-person visits had to be postponed and nonemergency procedures rescheduled, with outpatient care shifting toward teleconsultation where possible. At that time, Erasmus MC’s teleconsultations occurred through telephone calls and an integrated EHR Skype for Business feature. Between March and June 2020, a clear spike in telephone consultations could be seen, from around 20,000 to an average of over 35,000. A similar fluctuation was observed for video consultations, which grew from approximately 10 per month to an average of over 187 (Figure 1).

**Figure 1.** Video consultation use over time and COVID-19 cases in 2020. EMC: Erasmus University Medical Center.



The rapid demand for teleconsultations required that more health care personnel be rerouted to cover the need, and steps had to be taken to increase the bandwidth of the service. It is in this context that our institution needed to develop an internal strategy for communication and training regarding the use of these digital

tools that were not commonly used before the pandemic. To instruct staff and patients on the use of the platform, didactic material targeting the two groups separately was designed (Figures 2 and 3).

**Figure 2.** Didactic materials for patients.

## Starten van een videoconsult

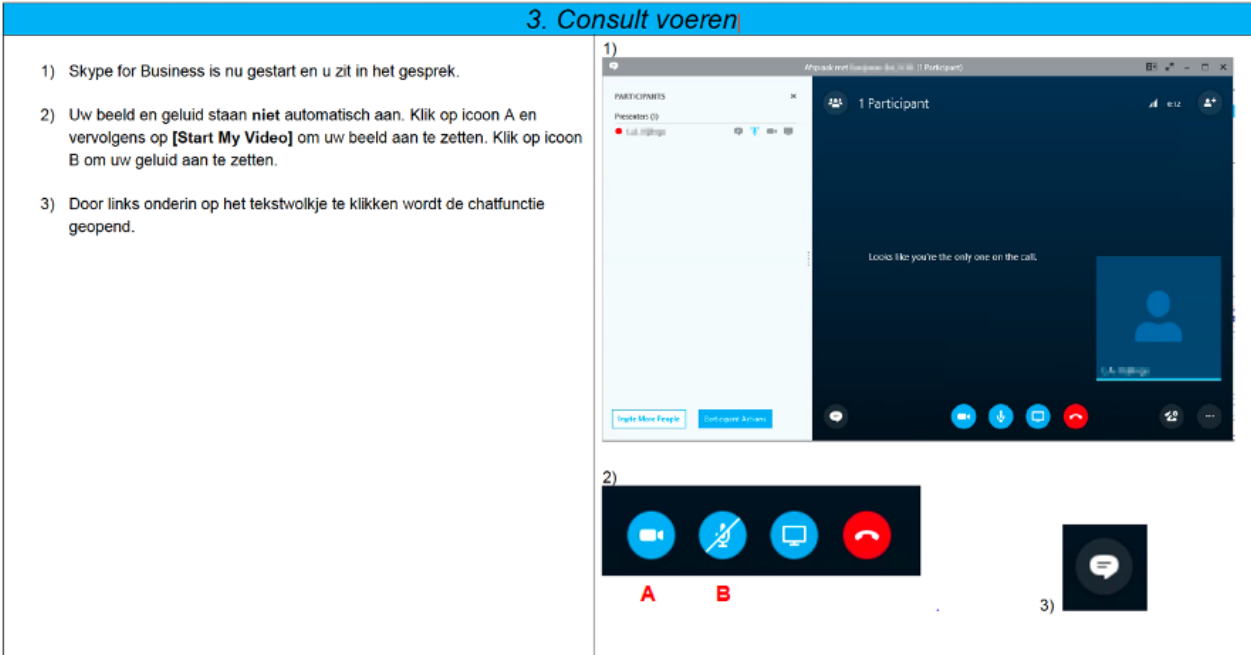
Neem voldoende tijd om u voor te bereiden op uw videoconsult en om in te loggen. Zorg dat u op tijd klaar zit voor uw afspraak wanneer uw arts zich (online) meldt. Volg deze stappen:

- Ga naar [www.erasmusmc.nl/mijnerasmusmc](http://www.erasmusmc.nl/mijnerasmusmc), bij voorkeur via de standaardbrowser van uw apparaat
- Log in met uw DigiD
- Na het inloggen opent u uw afsprakenoverzicht – klik hiervoor op één van deze rode vlakken:


**Figure 3.** Didactic materials for health care professionals.

## 3. Consult voeren

- 1) Skype for Business is nu gestart en u zit in het gesprek.
- 2) Uw beeld en geluid staan **niet** automatisch aan. Klik op icoon A en vervolgens op [Start My Video] om uw beeld aan te zetten. Klik op icoon B om uw geluid aan te zetten.
- 3) Door links onderin op het tekstvolkje te klikken wordt de chatfunctie geopend.



As seen in [Figure 1](#), after the first wave of COVID-19, the number of telephone call consultations went back to pre-pandemic numbers, whereas video consultations have diminished but are still higher than before the crisis. Taking advantage of the benefit of hindsight, certain lessons have

emerged from Erasmus MC's approach to the sudden teleconsultation demand increase.



## Design and Implementation Considerations

Designing protocols for telehealth requires proper clinical and technical scoping with careful testing. Erasmus MC's patients and health care professionals were able to adapt to the new teleconsultation approach, but attention should be paid to the care pathway flow. There are still concerns about how these tools could impact the professional–patient relationship [15], and the need to involve stakeholders in design and development is raised time and time again [16]. Even in situations where the urgency is great, implementation without the involvement of frontline care providers results in unanticipated incidents and disruptions to daily practice [17].

There are many other aspects that we must consider in this rapid implementation. Our choice of platform needs to be aligned with interoperability and security requirements such as EHR integration and General Data Protection Regulation (GDPR) or Health Insurance Portability and Accountability Act (HIPAA) compliance. Partnerships with new vendors should attempt to

meet not only current but also future needs. Patient privacy and data ownership are still underexplored territories from policy and regulatory perspectives that need to be treaded carefully [18]. Patient-generated data is no longer just part of the health care system but rather belongs within a new context of “consumer” health care services [18]. Usage of data requires rights to be renegotiated where transparency and open dialog are paramount for a balanced agenda.

Returns on investment, consultation fees, costs, and telehealth rules in general are nascent at best, and in many countries, clinicians are not reimbursed for virtual consultations or online prescriptions [19]. Lack of reimbursement and revenue has been regularly cited as a significant barrier to the adoption and implementation of telemedicine services [20]. COVID-19 has been a great promotor for changes in this area. For example, in the United States, Medicare took an important step this past March to reduce the payment obstacle by covering telehealth in many more settings, at least temporarily [21].

Textbox 1 presents some important questions we need to be asking ourselves to prepare our teleconsultation system for a post-COVID-19 age.

### Textbox 1. Questions for a post-COVID-19 teleconsultation system.

- What are the advantages of developing a system in-house versus partnering with a vendor?
- How well can our institution's current infrastructure handle in-house development?
- Will our institution benefit from an assortment of tools or a unified platform?
- What processes would need to be changed to accommodate a new teleconsultation system?
- How will the workforce be trained in the use of the teleconsultation system? What about the patients?
- What kind of reactions would adopting a teleconsultation platform generate?
- How will our institution communicate the policies and protocols for after COVID-19?

## Final Thoughts

It is imperative that the solutions we develop today attempt to be as sustainable as possible in the long run. It would be unfortunate if advances made today were to be discarded once the crisis is over. In this new era, the use of truly multidisciplinary teams that bring together not only health professionals but also designers, engineers, social scientists, and health economists, among others, is a critical first step in

developing open, productive, and sustainable implementations [17].

We have the opportunity to start acting on a vision of the future of medicine, placing the building blocks for the consultation room of 2030. The impact the current pandemic will have on our society is greater than we can currently understand, and we must be responsible in our approach to the changes that are coming. Every step makes a footprint—let's make ours count.

## Acknowledgments

We would like to thank Prof Joke Hendricks and Eva Hofmeester, MSc for their cooperation and support.

## Authors' Contributions

All authors contributed extensively to the work presented in this paper. All authors reviewed and approved the final version of the manuscript.

## Conflicts of Interest

None declared.

## References

1. Kaplan J, Frias L, McFall-Johnsen M. A third of the global population is on coronavirus lockdown — here's our constantly updated list of countries and restrictions. Business Insider. 2020 Jul 11. URL: <https://www.businessinsider.in/international/news/a-third-of-the-global-population-is-on-coronavirus-lockdown-x2014-hereaposs-our-constantly-updated-list-of-countries-and-restrictions/slidelist/75208623.cms> [accessed 2020-09-30]
2. Lewnard JA, Lo NC. Scientific and ethical basis for social-distancing interventions against COVID-19. *Lancet Infect Dis* 2020 Jun;20(6):631-633 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30190-0](https://doi.org/10.1016/S1473-3099(20)30190-0)] [Medline: [32213329](https://pubmed.ncbi.nlm.nih.gov/32213329/)]
3. Langabeer J, Gonzalez M, Alqusairi D, Champagne-Langabeer T, Jackson A, Mikhail J, et al. Telehealth-Enabled Emergency Medical Services Program Reduces Ambulance Transport to Urban Emergency Departments. *West J Emerg Med* 2016 Nov 1;17(6):713-720 [FREE Full text] [doi: [10.5811/westjem.2016.8.30660](https://doi.org/10.5811/westjem.2016.8.30660)] [Medline: [27833678](https://pubmed.ncbi.nlm.nih.gov/27833678/)]
4. Mehrohta A, Ray K, Brockmeyer DM, Barnett ML, Bender JA. Rapidly Converting to “Virtual Practices”: Outpatient Care in the Era of Covid-19. *NEJM Catal Innov Care Deliv* 2020 Apr 01:1. [doi: [10.1056/CAT.20.0091](https://doi.org/10.1056/CAT.20.0091)]
5. Barsom E, Feenstra T, Bemelman W, Bonjer J, Schijven M. Coping with COVID-19: scaling up virtual care to standard practice. *Nat Med* 2020 May;26(5):632-634. [doi: [10.1038/s41591-020-0845-0](https://doi.org/10.1038/s41591-020-0845-0)] [Medline: [32405054](https://pubmed.ncbi.nlm.nih.gov/32405054/)]
6. Kruse CS, Regier V, Rheinboldt KT. Barriers over time to full implementation of health information exchange in the United States. *JMIR Med Inform* 2014;2(2):e26 [FREE Full text] [doi: [10.2196/medinform.3625](https://doi.org/10.2196/medinform.3625)] [Medline: [25600635](https://pubmed.ncbi.nlm.nih.gov/25600635/)]
7. Gagnon M, Légaré F, Labrecque M, Frémont P, Pluye P, Gagnon J, et al. Interventions for promoting information and communication technologies adoption in healthcare professionals. *Cochrane Database Syst Rev* 2009 Jan 21(1):CD006093 [FREE Full text] [doi: [10.1002/14651858.CD006093.pub2](https://doi.org/10.1002/14651858.CD006093.pub2)] [Medline: [19160265](https://pubmed.ncbi.nlm.nih.gov/19160265/)]
8. OECD. Achieving efficiency improvements in the health sector through the implementation of information and communication technologies. 2010. URL: <https://www.digitalhealthnews.eu/publications/latest/1933-achieving-efficiency-improvements-in-the-health-sector-through-the-implementation-of-information-and-communication-technologies> [accessed 2020-09-30]
9. Giunti G, Guisado-Fernandez E, Belani H, Lacalle-Remigio JR. Mapping the Access of Future Doctors to Health Information Technologies Training in the European Union: Cross-Sectional Descriptive Study. *J Med Internet Res* 2019 Aug 12;21(8):e14086 [FREE Full text] [doi: [10.2196/14086](https://doi.org/10.2196/14086)] [Medline: [31407668](https://pubmed.ncbi.nlm.nih.gov/31407668/)]
10. Policy for Device Software Functions and Mobile Medical Applications. 2019. URL: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf> [accessed 2020-09-30]
11. Detmer DE, Lumpkin JR, Williamson JJ. Defining the medical subspecialty of clinical informatics. *J Am Med Inform Assoc* 2009;16(2):167-168 [FREE Full text] [doi: [10.1197/jamia.M3094](https://doi.org/10.1197/jamia.M3094)] [Medline: [19074293](https://pubmed.ncbi.nlm.nih.gov/19074293/)]
12. Health Education England. The Topol Review: Preparing the healthcare workforce to deliver the digital future. URL: <https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf> [accessed 2020-09-30]
13. Alderweireld CEA, Buiting AGM, Murk JLAN, Verweij JJ, van Kasteren M. COVID-19: patiënt nul in Nederland. *Ned Tijdschr Geneesk*. 2020. URL: <https://www.ntvg.nl/artikelen/covid-19-patient-nul-nederland/icmje> [accessed 2020-09-30]
14. Rijksinstituut voor Volksgezondheid en Milieu. Epidemiologische situatie COVID-19 in Nederland. 2020 Jun 10. URL: [https://www.rivm.nl/sites/default/files/2020-06/COVID-19\\_WebSite\\_rapport\\_dagelijks20200610\\_1033.pdf](https://www.rivm.nl/sites/default/files/2020-06/COVID-19_WebSite_rapport_dagelijks20200610_1033.pdf) [accessed 2020-09-30]
15. Visvanathan A, Gibb AP, Brady RRW. Increasing clinical presence of mobile communication technology: avoiding the pitfalls. *Telemed J E Health* 2011 Oct;17(8):656-661. [doi: [10.1089/tmj.2011.0018](https://doi.org/10.1089/tmj.2011.0018)] [Medline: [21780941](https://pubmed.ncbi.nlm.nih.gov/21780941/)]
16. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. 2014 May;39(5):356-364 [FREE Full text] [Medline: [24883008](https://pubmed.ncbi.nlm.nih.gov/24883008/)]
17. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. *J Am Med Inform Assoc* 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: [10.1093/jamia/ocaa037](https://doi.org/10.1093/jamia/ocaa037)] [Medline: [32208481](https://pubmed.ncbi.nlm.nih.gov/32208481/)]
18. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who Owns the Data? Open Data for Healthcare. *Front Public Health* 2016;4:7 [FREE Full text] [doi: [10.3389/fpubh.2016.00007](https://doi.org/10.3389/fpubh.2016.00007)] [Medline: [26925395](https://pubmed.ncbi.nlm.nih.gov/26925395/)]
19. Hollander JE, Sites FD. The Transition from Reimagining to Recreating Health Care Is Now. *NEJM Catalyst* 2020 Apr 08:1 [FREE Full text] [doi: [10.1056/CAT.20.0093](https://doi.org/10.1056/CAT.20.0093)]
20. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. *J Telemed Telecare* 2018 Jan;24(1):4-12. [doi: [10.1177/1357633X16674087](https://doi.org/10.1177/1357633X16674087)] [Medline: [29320966](https://pubmed.ncbi.nlm.nih.gov/29320966/)]
21. Butler SM. After COVID-19: Thinking Differently About Running the Health Care System. *JAMA* 2020 Jun 23;323(24):2450-2451. [doi: [10.1001/jama.2020.8484](https://doi.org/10.1001/jama.2020.8484)] [Medline: [32573658](https://pubmed.ncbi.nlm.nih.gov/32573658/)]

## Abbreviations

**EHR:** Electronic Health Records

**Erasmus MC:** Erasmus University Medical Center

**GDPR:** General Data Protection Regulation

**HIPAA:** Health Insurance Portability and Accountability Act

*Edited by G Eysenbach, G Fagherazzi; submitted 08.06.20; peer-reviewed by F Muhammad, E Da Silva, F Saigí-Rubió, I Mircheva, E Burner, K Goniewicz, C Basch, S Pit; comments to author 30.06.20; revised version received 28.07.20; accepted 09.09.20; published 05.10.20.*

*Please cite as:*

*Giunti G, Goossens R, De Bont A, Visser JJ, Mulder M, Schuit SCE*

*The Need for Sustainable Teleconsultation Systems in the Aftermath of the First COVID-19 Wave*

*J Med Internet Res 2020;22(10):e21211*

URL: <http://www.jmir.org/2020/10/e21211/>

doi: [10.2196/21211](https://doi.org/10.2196/21211)

PMID: [32997642](https://pubmed.ncbi.nlm.nih.gov/32997642/)

©Guido Giunti, Richard Goossens, Antoinette De Bont, Jacob J Visser, Mark Mulder, Stephanie C E Schuit. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Viewpoint

# Covidom, a Telesurveillance Solution for Home Monitoring Patients With COVID-19

Youri Yordanov<sup>1</sup>, MD, PhD; Agnes Dechartres<sup>2</sup>, MD, PhD; Xavier Lescure<sup>3</sup>, MD, PhD; Caroline Apra<sup>4</sup>, MD; Pascaline Villie<sup>5</sup>, MS; Jerome Marchand-Arvier<sup>5</sup>, MS; Erwan Debuc<sup>6</sup>, MD; Aurélien Dinh<sup>7</sup>, MD; Patrick Jourdain<sup>8</sup>, MD, PhD; On Behalf Of The AP-HP / Universities / Inserm COVID-19 Research Collaboration<sup>9</sup>

<sup>1</sup>Sorbonne Université, AP-HP, Hôpital Saint Antoine, Service d'Accueil des Urgences, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, UMR-S 1136, Paris, France

<sup>2</sup>Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, UMR-S 1136, AP-HP, Hôpital Pitié Salpêtrière, Département de Santé Publique, Centre de Pharmacopidémie de l'AP-HP (Cephepi), Paris, France

<sup>3</sup>Department of Infectious and Tropical Diseases, Bichat-Claude Bernard University Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France

<sup>4</sup>Service de Neurochirurgie, Hôpital Pitié Salpêtrière, Assistance Publique-Hôpitaux de Paris – Sorbonne Université, Paris, France

<sup>5</sup>Assistance Publique-Hôpitaux de Paris, Paris, France

<sup>6</sup>Service d'Accueil des Urgences, Hôpital Saint Antoine, Assistance Publique-Hôpitaux de Paris – Sorbonne Université, Paris, France

<sup>7</sup>Infectious Disease Department, R Poincaré University Hospital, Garches, Assistance Publique-Hôpitaux de Paris – Paris Saclay University, Paris, France

<sup>8</sup>Université Paris Saclay, AP-HP, Hôpital du Kremlin Bicêtre, Service de Cardiologie, Département médico-universitaire Coeur Vaisseaux Reins, Kremlin Bicêtre, France, Kremlin-Bicêtre, France

<sup>9</sup>AP-HP / Universities / Inserm COVID-19 Research Collaboration, Paris, France

**Corresponding Author:**

Youri Yordanov, MD, PhD

Sorbonne Université, AP-HP, Hôpital Saint Antoine, Service d'Accueil des Urgences, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, UMR-S 1136

34 rue Crozatier

Paris

France

Phone: 33 171970869

Email: [youri.yordanov@aphp.fr](mailto:youri.yordanov@aphp.fr)

## Abstract

In a matter of months, COVID-19 has escalated from a cluster of cases in Wuhan, China, to a global pandemic. As the number of patients with COVID-19 grew, solutions for the home monitoring of infected patients became critical. This viewpoint presents a telesurveillance solution—Covidom—deployed in the greater Paris area to monitor patients with COVID-19 in their homes. The system was rapidly developed and is being used on a large scale with more than 65,000 registered patients to date. The Covidom solution combines an easy-to-use and free web application for patients (through which patients fill out short questionnaires on their health status) with a regional control center that monitors and manages alerts (triggered by questionnaire responses) from patients whose health may be deteriorating. This innovative solution could alleviate the burden of health care professionals and systems while allowing for rapid response when patients trigger an alert.

(*J Med Internet Res* 2020;22(10):e20748) doi:[10.2196/20748](https://doi.org/10.2196/20748)

**KEYWORDS**

COVID-19; coronavirus disease; home monitoring; telesurveillance; monitoring; patient; infectious disease; app

## Introduction

In less than 7 months, COVID-19 escalated from a cluster of cases in Wuhan, China, to a global pandemic with more than 15 million infected people and 630,000 deaths in over 200 countries [1,2]. The clinical characteristics of patients with

COVID-19 are well described, with most presenting mild symptoms and fatalities occurring mainly in chronically ill and older patients [3-7]. In addition to being a therapeutic challenge for physicians and health care workers, the exponential increase in patients with COVID-19, and their considerable length of stay in a hospital, could exceed health care systems' capacities

[8-12]. To allow hospitals to focus on vulnerable and the most severely ill patients, those with COVID-19 but presenting no serious symptoms are being sent home [13]. However, for 10% to 15% of these patients, the disease will become severe, which requires surveillance [13,14]. Various systems have been set up to carry out this surveillance, often involving general practitioners (GPs) and telephone-based and/or home visits, or the use of telehealth technologies for virtual consultations [15,16]. However, all these systems rely on the individual management of every patient by a single doctor (GP, infectious disease specialist, or any other specialist involved in COVID-19 management). In a pandemic situation, GPs and infectious disease specialists are scarce resources and should be mobilized wisely [15,17,18].

To offer alternatives to patients while reserving medical resources for the situations that require it, the Greater Paris University Hospitals (Assistance Publique-Hôpitaux de Paris, [APHP]), in collaboration with regional GP organizations and a software company specializing in patient digital pathways, quickly developed a remote telesurveillance solution named Covidom for the home monitoring of patients with COVID-19.

## The Covidom Solution

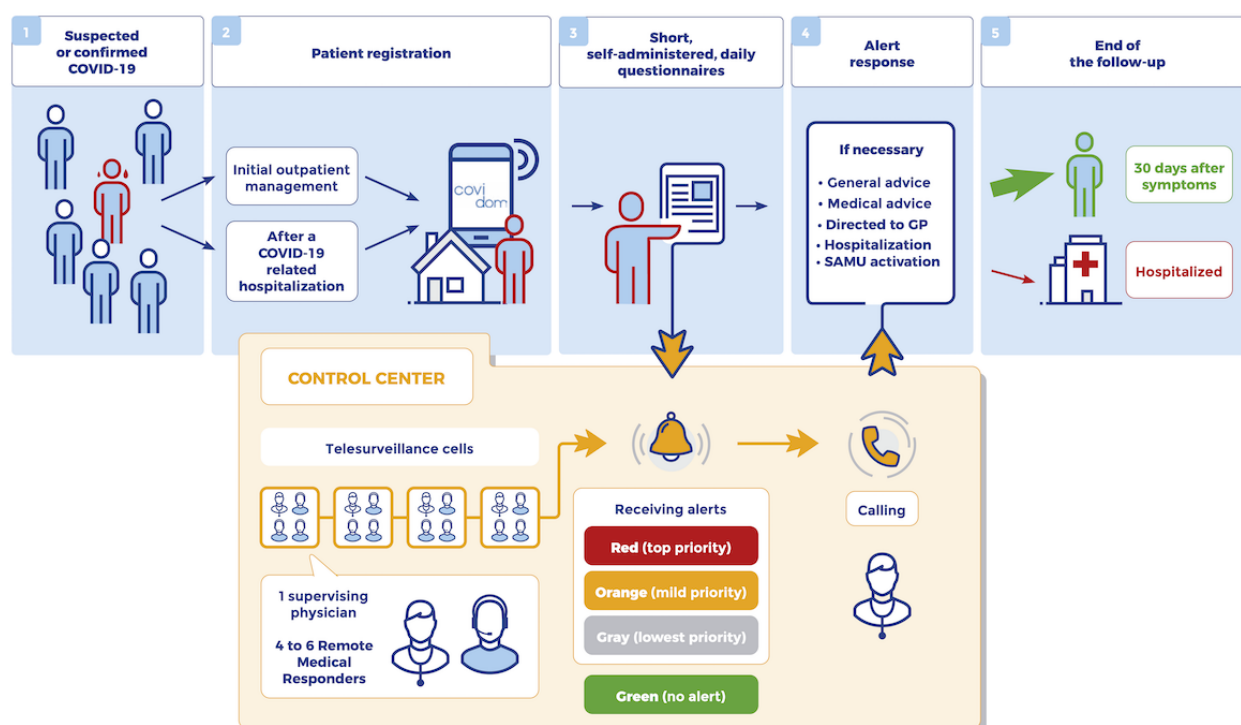
Covidom combines an easy-to-use and free web application for patients with a regional control center to manage alerts (Figure

1). Patients with a suspected or confirmed case of COVID-19, according to the French public health authorities' definition of COVID-19 infection [19], are registered by a physician after receiving a brief description of the Covidom solution and providing oral consent to participate. Registration can be performed either as part of outpatient management after diagnosis (ie, after a visit to an emergency department or consultation with a GP or another specialist) or at the time of discharge after COVID-19-related hospitalization. Registration is a simple procedure where patients provide simple baseline characteristics, including age, gender, phone number or email address, date of first symptoms, and risk profile. A high-risk profile includes the presence of cardiovascular disease, diabetes, chronic lung disease, immunodeficiency (transplant, active cancer treatment, uncontrolled HIV infection, etc), third trimester of pregnancy, or age >65 years [19].

Patients then receive a registration link via a short mobile message or email, through which they complete registration and provide electronic consent for the Covidom telesurveillance program. They are informed of the potential use of their anonymized data for research purposes. This use was approved by the scientific and ethical committee of APHP (IRB00011591).

The data is available upon request for academic researchers.

**Figure 1.** The Covidom solution: patient pathway and regional control center organization. GP: general practitioner, SAMU: Service d'Aide Médicale Urgente.



## The Covidom Web Application

The web application was designed to be straightforward and intuitive to use for patients. The interface of the application can be seen in Figure 2. Patients complete one or two self-administered daily monitoring questionnaires for a duration of 30 days after symptom onset. These questionnaires involve

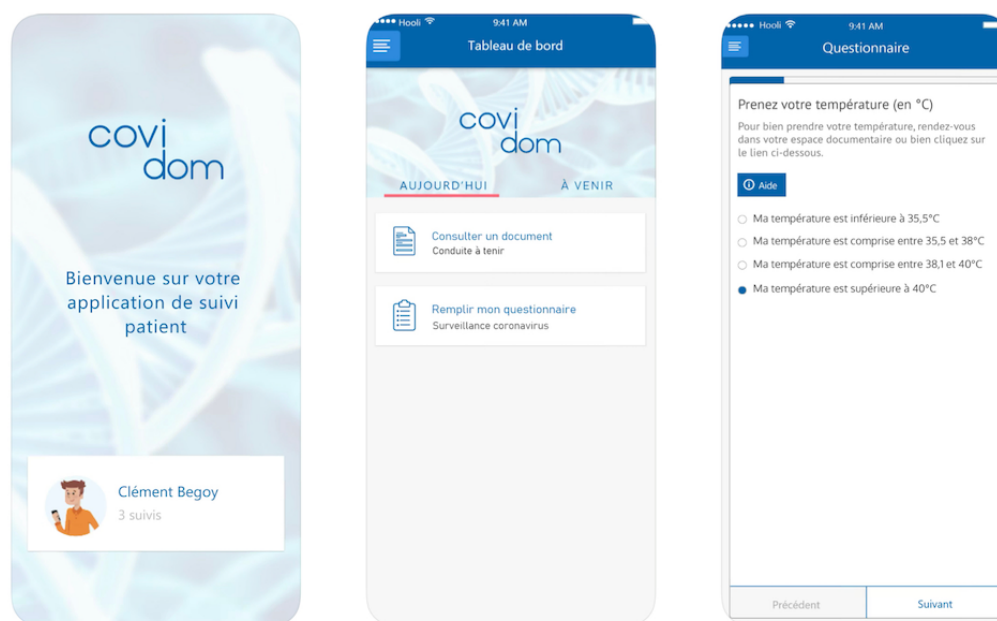
fewer than 10 short and standardized questions. The questionnaires can be accessed via computer or smartphone, and patients are informed by mobile message or email to complete them with a reminder in case of no response. The questions ask patients to self-report their respiratory rate, heart rate, temperature, respiratory uneasiness (adapted from the modified Borg dyspnea scale [20]), nausea, malaise, as well as



psychological discomfort and difficulties dealing with lockdown measures. Patients assessed as high risk by the physician who performed their initial evaluation need to complete the questionnaires twice a day, while low-risk patients respond only once a day. The answers to these monitoring questionnaires can trigger different types of alerts at the regional control center. These questionnaires were elaborated and tested by a panel of multidisciplinary health care professionals (infectious diseases, emergency physicians, GPs, and telesurveillance specialists).

In the web application, patients can also find information on the virus and how to mitigate transmission risk (ie, French health ministry documents); how to measure one's own temperature, heart rate, and respiratory rate; and how to seek psychological support if needed. In case of an emergency, patients are advised to directly contact the national emergency number by dialing 15 (Service d'Aide Médicale Urgente [SAMU]).

**Figure 2.** Covidom web application screenshots.



### The Covidom Regional Control Center

The Covidom regional control center is open from 8 AM to 8 PM, 7 days a week, and covers all patients using the Covidom system in the greater Paris area (12 million inhabitants). It is built on the concept of autonomous remote monitoring cells. Each cell is made up of 4 to 6 trained remote monitoring responders (RMRs) and a supervising physician, all physically colocated at the Covidom regional control center, equipped with face masks and adhering to physical distancing measures. The mission of the cells is to handle the alerts generated by patient answers to the daily or twice-daily questionnaires. Patient answers are classified into 4 categories by an automated algorithm:

- No alert: everything is considered normal, no need for further action;
- Orange alert: some of the answers are above a certain threshold. These alerts need a response from the regional control center;
- Red alert: some of the answers suggest that the patient's condition may be deteriorating. These alerts need a response from the control center, with the highest level of priority;
- Gray alert: the patients did not answer the questionnaire. These alerts need a response from the control center and patients need to be called, but the level of priority is low.

To handle the alerts, RMRs access the patient record and contact patients by phone to identify the cause of the alert. If needed,

the supervising physician of the cell can intervene and assess the patient over the phone. An alert is considered handled once the RMR or the physician offers a solution to the patient: general advice, medical advice, directed to their GP, hospitalization, or contact with the SAMU. Infectious disease wards, emergency departments, or the SAMU can be contacted directly by the control center using dedicated phone numbers. If necessary, these contacts could result in a mobile intensive care unit staffed with emergency physicians sent to the patient's home or a regular ambulance with a paramedic sent to assess the patient and transport them to a hospital. Of note, in the case of remote medical assessment, Covidom personnel does not charge a fee.

The control center cell physicians and RMRs are volunteers from different backgrounds ([Multimedia Appendix 1](#)). Physicians are rarely infectious disease specialists, GPs, or emergency physicians since those individuals are on the frontline caring for patients in need of acute care. Covidom personnel are mostly other specialists with decreased activity because of the lockdown who wanted to contribute to crisis management. All physicians and RMRs receive theoretical and practical training, the intensity of which depends on the person's profile. They do not receive any financial incentives, but nonfinancial incentives are offered, such as meals or transportation solutions if public transport is not available. All volunteers have to sign an individual contract with the APHP for medical confidentiality, insurance, and liability reasons. On-site psychological support is available if needed.

## Overview of Covidom as of May 19, 2020

In the period between March 9 to May 19, 2020, 57,182 patients were registered in Covidom with a suspected or confirmed case of COVID-19. These patients were referred by 1709 physicians working in 30 public and 70 private hospitals and by 2131 GPs (in private or public medical practices). Most patients were referred as part of their initial outpatient care (50,020/57,182, 87.5%) while 7162 were included at hospitalization discharge. Out of these patients, 84.5% (48,290/57,182) confirmed their registration, 8.4% (4057/48,290) never answered a surveillance questionnaire, and 70.6% (34,104/48,290) answered questionnaires for more than 7 days. A total of 104 patients were offered alternatives, by contacting patients' GPs to organize a follow-up, as they had trouble using the system (eg, uncomfortable using the required technologies or language issues). Patients' mean age was 43.7 years (SD 15.8) and a majority were female (33,542/48,290, 58.7%) (Table 1). The

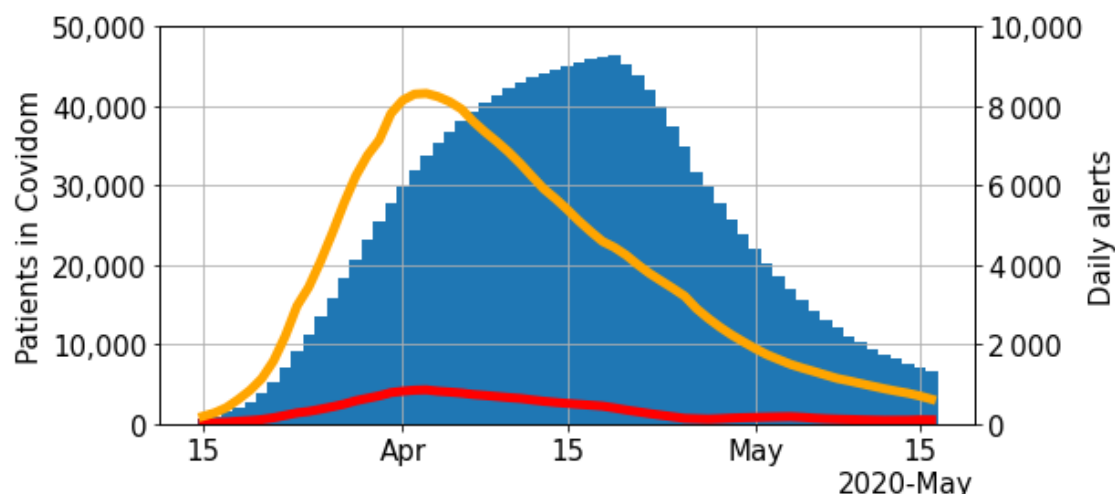
patients' risk profile was recorded as high in 60.3% (3315/5493) of cases included at hospital discharge and in 39.9% (17,082/42,797) of cases included as part of their initial outpatient care.

During follow-up (phases 3 and 4 of the patient's pathway in Figure 1), patients triggered 21,873 red alerts and 211,160 orange alerts. Red alerts were handled in a median time of 2 min and 20 s (IQR 46 s to 6 min and 54 s) and orange alerts were handled in 10 min and 34 s (IQR 1 min and 28 s to 93 min and 51 s). We present in Figure 3 the weekly averaged counts of patients managed using Covidom and the alerts generated. From March 30, 368 alerts resulted in contact with SAMU (via the national emergency number) through the regional control center (over 215,056 alerts by 41,758 patients). As of May 19, 72.0% (34,767/48,290) of patients had their follow-up terminated, 1.1% (544/48,290) had been hospitalized or rehospitalized, and 0.1% had died (39/48,290).

**Table 1.** General characteristics of and reasons for end of follow-up among patients using Covidom, as of May 19, 2020.

Characteristic	Posthospital discharge management (n=5493)	Initial outpatient management (n=42,797)	Total (N=48,290)
<b>General characteristics</b>			
Age (years), mean (SD)	48.5 (17.2)	42.3 (14.9)	43.7 (15.8)
<b>Gender, n (%)</b>			
Male	2669 (48.6)	16,260 (38.0)	23,564 (41.2)
Female	2818 (51.3)	26,488 (61.9)	33,542 (58.7)
<b>Risk profile, n (%)</b>			
High-risk profile	3315 (60.3)	17,082 (39.9)	24,756 (43.3)
<b>Reason for end of follow-up, n (%)</b>			
Automatic termination of follow-up after 30 days	3957 (72.0)	30,810 (72.0)	34,767 (72.0)
Follow-up ended early at patient's request <sup>a</sup>	831 (15.1)	7473 (17.5)	8304 (17.2)
Ongoing follow-up	590 (10.74)	4046 (9.45)	4636 (9.60)
Hospital admission	111 (2.0)	433 (1.0)	544 (1.1)
Death	4 (0.1)	35 (0.1)	39 (0.1)

<sup>a</sup>Follow-up ended early at patient's request: no more symptoms, no longer felt like answering questionnaires, or any other reason left at the patient's discretion.

**Figure 3.** Number of patients and alerts over time.

### Implications, Future Works, and Limitations

To our knowledge, Covidom is the first and largest telesurveillance solution (65,202 patients as of July 24) described for the home monitoring of COVID-19 cases with the aim to alleviate the burden of health care professionals and systems. Telesurveillance has never been used in acute infectious diseases at this scale until now [21]; previously, it had been mostly used in chronic diseases [22,23]. Most health care systems are based on in-person interactions between patients and their clinicians, but in a pandemic context, this situation is highly challenged, and health digital solutions are of interest both to patients and to the health care system [15-17]. From patients' perspective, this system can provide reassurance by daily monitoring their condition with a procedure in place in case of worsening symptoms. Patients often worry about the potential and sudden worsening of symptoms during a lockdown, due to limited social contact [24,25]. This telesurveillance system allows for close but minimally invasive surveillance using daily short questionnaires with fewer than 10 questions, which is easily accepted by patients [26]. From the public health perspective, this system may offer a partial virtual safety net to rapidly detect any signs of deterioration in patients with COVID-19, while making proper use of scarce resources via a 2-step process where automated alert algorithms can trigger a medical response when needed. Automatic algorithms and health care professionals based in a regional control center could help reserve health care workers and hospital beds for the patients who need them most and alleviate pressure on the health care system.

Such tools also have the major advantage of ensuring close surveillance while avoiding physical contact, which can help limit the spread of the virus and possible health care worker contamination. Providing appropriate care while preserving one's own health is a strong motivation for health care workers to rapidly develop and widely use virtual health care solutions [27].

Finally, Covidom represents an important source of epidemiologic data, providing an opportunity to increase our

knowledge of the disease, in particular of its common but least studied mild form. Covidom was initially deployed in the greater Paris area but is being extended to other French regions using the same principles: use of a web application with a dedicated regional control center whose functioning may depend on the region. Because of its simplicity and quick response, this solution could be easily adapted in other countries.

Of course, the Covidom solution needs to be thoroughly evaluated; in particular, the efficiency and ability of the alerts to detect patients at high risk of deterioration and the medico-economic impact of such a solution need evaluation. To do this, we will link the Covidom database with hospitals and national social security databases to identify patients who directly contacted them or were self-referred to an emergency department.

The Covidom system was sustainable during the lockdown due to the personnel availability that resulted from nonurgent elective procedures or appointments being rescheduled; most of the workforce comprised salaried employees (as opposed to a pay-per-service system). We observed significant fluctuations in the availability of human resources. At first, and due to the lockdown, many volunteers offered their help. Since lockdown measures were lifted (May 11, 2020) and as control center cell physicians and RMRs progressively resumed their usual activities, finding enough personnel has become more of a challenge. Adapting the Covidom solution and offering alternatives to patients who had difficulty with the system (unfamiliar with these technologies or language issues) was done by connecting these patients to their GPs. However, additional options could have included translated versions of the questionnaires and the other available documents to help patients with language issues. Surveillance could also have been more flexible as it was found to be short for some patients with recurrent or persistent symptoms, while others would have preferred to stop the follow-up as soon as the symptoms disappeared. Finally, sharing the patients' Covidom file, or a summary of it, with patients' GPs was not done.

## Conclusion

Covidom is an innovative solution for the home monitoring of patients with COVID-19. The model could be easily transposed to other countries or contexts. Most patients have been included as part of their initial outpatient management, making Covidom

the largest cohort to date of patients with a mild case of COVID-19, which is the form that occurs in a majority of patients but is least studied. Using telesurveillance solutions like Covidom could augment health care systems' abilities by allowing them to monitor patients and promptly identify worsening symptoms, while limiting the need to travel and the risk of contamination.

## Acknowledgments

We thank Falzon Alexandre, Fayolle Guillaume, Laporte Fanny, Amélie Tortel, and the Nouveaux-e Santé team for their help with the web application and regional center surveillance interface development. We also thank Debastard Laurent, Grenier Alexandre, Hody Julien, Penn Thomas, and the Paris region URPS (Union régionale des professionnels de santé) for their help in the development and dissemination of the Covidom solution.

This study received funding from the Programme Hospitalier de Recherche Clinique 2020 of the French Ministry of Health and a research grant from the APHP-Fondation de France. The Covidom platform received funding from the EIT Health for COVID-19 research.

Members of the AP-HP / Universities / Inserm COVID-19 research collaboration are as follows: the Writing Committee (Youri Yordanov, Agnes Dechartres, François-Xavier Lescure, Caroline Apra, Pascaline Villie, Jérôme Marchand-Arvier, Erwan Debuc, Aurélien Dinh, and Patrick Jourdain); the Data-Sciences Committee (Caroline Apra, Charlotte Caucheteux, Luc Jaulmes, Alexandre Gramfort, Jenny Mansour, and Arthur Mensch); the Scientific Committee (Amélie Aime-Eusebi, Caroline Apra, Alexandre Bleibtreu, Erwan Debuc, Agnes Dechartres, Laurène Deconinck, Aurélien Dinh, Patrick Jourdain, Christine Katlama, Josselin Lebel, François-Xavier Lescure, and Youri Yordanov); the Covidom Regional Center Steering Committee (Yves Artigou, Amélie Banzet, Elodie Boucheron, Christiane Boudier, Edouard Buzenac, Marie-Claire Chapron, Dalhia Chekaoui, Laurent De Bastard, Erwan Debuc, Aurélien Dinh, Alexandre Grenier, Pierre-Etienne Haas, Julien Hody, Michèle Jarraya, Patrick Jourdain, Louis Lacaille, Aurélie Le Guern, Jeremy Leclert, Fanny Male, Jérôme Marchand-Arvier, Emmanuel Martin-Blondet, Apollinne Nassour, Oussama Ourahou, Thomas Penn, Ambre Ribardiere, Nicolas Robin, Camille Rouge, Nicolas Schmidt, and Pascaline Villie); and the Centre de Pharmaco-Epidémiologie de l'AP-HP (Sofia Zemouri).

## Authors' Contributions

YY and AD were involved in the study conception, data extraction, data analysis, interpretation of results, and drafting of the manuscript. XL was involved in the Covidom solution development, interpretation of results, and critical revision of the manuscript. CA was involved in the data extraction, data analysis, interpretation of results, and critical revision of the manuscript. PV, JM-A, ED, ADi, and PJ were involved in the Covidom solution development, study conception, interpretation of results, and critical revision of the manuscript.

YY (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study have been explained.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

List of physicians, supervising physicians, and remote monitoring responders part of the Covidom system.

[DOC File, 172 KB - [jmir\\_v22i10e20748\\_app1.doc](#)]

## References

1. Lu H, Stratton CW, Tang Y. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *J Med Virol* 2020 Apr;92(4):401-402 [FREE Full text] [doi: [10.1002/jmv.25678](#)] [Medline: [31950516](#)]
2. Coronavirus Update (Live), 433,760 Cases and 631,402 Deaths from COVID-19 Virus Pandemic. Worldometer. 2020. URL: <https://www.worldometers.info/coronavirus/> [accessed 2020-07-23]
3. Wang X, Fang J, Zhu Y, Chen L, Ding F, Zhou R, et al. Clinical characteristics of non-critically ill patients with novel coronavirus infection (COVID-19) in a Fangcang Hospital. *Clin Microbiol Infect* 2020 Aug;26(8):1063-1068 [FREE Full text] [doi: [10.1016/j.cmi.2020.03.032](#)] [Medline: [32251842](#)]
4. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* 2020 Mar 28;395(10229):1054-1062. [doi: [10.1016/S0140-6736\(20\)30566-3](#)] [Medline: [32171076](#)]



5. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)] [Medline: [31986264](https://pubmed.ncbi.nlm.nih.gov/31986264/)]
6. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* 2020 Jul 01;180(7):934-943 [FREE Full text] [doi: [10.1001/jamainternmed.2020.0994](https://doi.org/10.1001/jamainternmed.2020.0994)] [Medline: [32167524](https://pubmed.ncbi.nlm.nih.gov/32167524/)]
7. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: [10.1001/jama.2020.6775](https://doi.org/10.1001/jama.2020.6775)] [Medline: [32320003](https://pubmed.ncbi.nlm.nih.gov/32320003/)]
8. Legido-Quigley H, Mateos-García JT, Campos VR, Gea-Sánchez M, Muntaner C, McKee M. The resilience of the Spanish health system against the COVID-19 pandemic. *Lancet Public Health* 2020 May;5(5):e251-e252 [FREE Full text] [doi: [10.1016/S2468-2667\(20\)30060-8](https://doi.org/10.1016/S2468-2667(20)30060-8)] [Medline: [32199083](https://pubmed.ncbi.nlm.nih.gov/32199083/)]
9. Watkins J. Preventing a covid-19 pandemic. *BMJ* 2020 Feb 28;368:m810. [doi: [10.1136/bmj.m810](https://doi.org/10.1136/bmj.m810)] [Medline: [32111649](https://pubmed.ncbi.nlm.nih.gov/32111649/)]
10. Armocida B, Formenti B, Ussai S, Palestra F, Missoni E. The Italian health system and the COVID-19 challenge. *Lancet Public Health* 2020 May;5(5):e253 [FREE Full text] [doi: [10.1016/S2468-2667\(20\)30074-8](https://doi.org/10.1016/S2468-2667(20)30074-8)] [Medline: [32220653](https://pubmed.ncbi.nlm.nih.gov/32220653/)]
11. Lewnard J, Liu V, Jackson M, Schmidt M, Jewell B, Flores J, et al. Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington: prospective cohort study. *BMJ* 2020 May 22;369:m1923. [doi: [10.1136/bmj.m1923](https://doi.org/10.1136/bmj.m1923)] [Medline: [32444358](https://pubmed.ncbi.nlm.nih.gov/32444358/)]
12. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet* 2020 Apr 11;395(10231):1225-1228 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30627-9](https://doi.org/10.1016/S0140-6736(20)30627-9)] [Medline: [32178769](https://pubmed.ncbi.nlm.nih.gov/32178769/)]
13. Gandhi RT, Lynch JB, Del Rio C. Mild or Moderate Covid-19. *N Engl J Med* 2020 Apr 24:e. [doi: [10.1056/NEJMcip2009249](https://doi.org/10.1056/NEJMcip2009249)] [Medline: [32329974](https://pubmed.ncbi.nlm.nih.gov/32329974/)]
14. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). World Health Organization. 2020 Feb 28. URL: [https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)) [accessed 2020-06-01]
15. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. *N Engl J Med* 2020 Apr 30;382(18):1679-1681. [doi: [10.1056/NEJMp2003539](https://doi.org/10.1056/NEJMp2003539)] [Medline: [32160451](https://pubmed.ncbi.nlm.nih.gov/32160451/)]
16. Mahase E. Covid-19: out-of-hours providers are drafted in to manage non-urgent patients in community. *BMJ* 2020 Mar 09;368:m959. [doi: [10.1136/bmj.m959](https://doi.org/10.1136/bmj.m959)] [Medline: [32152010](https://pubmed.ncbi.nlm.nih.gov/32152010/)]
17. Lurie N, Carr BG. The Role of Telehealth in the Medical Response to Disasters. *JAMA Intern Med* 2018 Jun 01;178(6):745-746. [doi: [10.1001/jamainternmed.2018.1314](https://doi.org/10.1001/jamainternmed.2018.1314)] [Medline: [29710200](https://pubmed.ncbi.nlm.nih.gov/29710200/)]
18. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. *N Engl J Med* 2020 Jun 04;382(23):e82 [FREE Full text] [doi: [10.1056/nejmp2005835](https://doi.org/10.1056/nejmp2005835)]
19. Infection au nouveau Coronavirus (SARS-CoV-2), COVID-19, France et Monde. Sante Publique France. URL: <https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/articles/infection-au-nouveau-coronavirus-sars-cov-2-covid-19-france-et-monde> [accessed 2020-04-27]
20. Muza SR, Silverman MT, Gilmore GC, Hellerstein HK, Kelsen SG. Comparison of Scales Used to Quantitate the Sense of Effort to Breathe in Patients with Chronic Obstructive Pulmonary Disease. *Am Rev Respir Dis* 1990 Apr;141(4\_pt\_1):909-913. [doi: [10.1164/ajrccm/141.4\\_pt\\_1.909](https://doi.org/10.1164/ajrccm/141.4_pt_1.909)]
21. Eron L. Telemedicine: the future of outpatient therapy? *Clin Infect Dis* 2010 Sep 15;51 Suppl 2:S224-S230. [doi: [10.1086/653524](https://doi.org/10.1086/653524)] [Medline: [20731581](https://pubmed.ncbi.nlm.nih.gov/20731581/)]
22. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. *N Engl J Med* 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: [10.1056/NEJMoa1010029](https://doi.org/10.1056/NEJMoa1010029)] [Medline: [21080835](https://pubmed.ncbi.nlm.nih.gov/21080835/)]
23. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. *BMJ* 2013 Oct 17;347:f6070. [doi: [10.1136/bmj.f6070](https://doi.org/10.1136/bmj.f6070)] [Medline: [24136634](https://pubmed.ncbi.nlm.nih.gov/24136634/)]
24. Horesh D, Brown AD. Traumatic stress in the age of COVID-19: A call to close critical gaps and adapt to new realities. *Psychol Trauma* 2020 May;12(4):331-335. [doi: [10.1037/tra0000592](https://doi.org/10.1037/tra0000592)] [Medline: [32271070](https://pubmed.ncbi.nlm.nih.gov/32271070/)]
25. Zhang Y, Ma ZF. Impact of the COVID-19 Pandemic on Mental Health and Quality of Life among Local Residents in Liaoning Province, China: A Cross-Sectional Study. *Int J Environ Res Public Health* 2020 Mar 31;17(7):e [FREE Full text] [doi: [10.3390/ijerph17072381](https://doi.org/10.3390/ijerph17072381)] [Medline: [32244498](https://pubmed.ncbi.nlm.nih.gov/32244498/)]
26. Sahlqvist S, Song Y, Bull F, Adams E, Preston J, Ogilvie D, iConnect consortium. Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: randomised controlled trial. *BMC Med Res Methodol* 2011 May 06;11(1):62 [FREE Full text] [doi: [10.1186/1471-2288-11-62](https://doi.org/10.1186/1471-2288-11-62)] [Medline: [21548947](https://pubmed.ncbi.nlm.nih.gov/21548947/)]
27. Webster P. Virtual health care in the era of COVID-19. *Lancet* 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30818-7](https://doi.org/10.1016/S0140-6736(20)30818-7)] [Medline: [32278374](https://pubmed.ncbi.nlm.nih.gov/32278374/)]



## Abbreviations

**APHP:** Assistance Publique-Hôpitaux de Paris  
**GP:** general practitioner  
**RMR:** remote monitoring responder  
**SAMU:** Service d'Aide Médicale Urgente  
**URPS:** Union régionale des professionnels de santé

*Edited by G Eysenbach; submitted 28.05.20; peer-reviewed by F Marongiu, A Sheikhtaheri, B Smith, M Simons; comments to author 13.07.20; revised version received 28.07.20; accepted 16.08.20; published 22.10.20.*

*Please cite as:*

Yordanov Y, Dechartres A, Lescure X, Apra C, Villie P, Marchand-Arvier J, Debuc E, Dinh A, Jourdain P, On Behalf Of The AP-HP / Universities / Inserm COVID-19 Research Collaboration  
Covidom, a Telesurveillance Solution for Home Monitoring Patients With COVID-19  
J Med Internet Res 2020;22(10):e20748  
URL: <http://www.jmir.org/2020/10/e20748/>  
doi: [10.2196/20748](https://doi.org/10.2196/20748)  
PMID: [33006938](https://pubmed.ncbi.nlm.nih.gov/33006938/)

©Youri Yordanov, Agnes Dechartres, Xavier Lescure, Caroline Apra, Pascaline Villie, Jerome Marchand-Arvier, Erwan Debuc, Aurélien Dinh, Patrick Jourdain, On Behalf Of The AP-HP / Universities / Inserm COVID-19 Research Collaboration. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Viewpoint

# Portals of Change: How Patient Portals Will Ultimately Work for Safety Net Populations

Alejandra Casillas<sup>1</sup>, MD, MSc; Anshu Abhat<sup>2</sup>, MD, MPH; Anish Mahajan<sup>2</sup>, MD, MSc; Gerardo Moreno<sup>3</sup>, MD, MSc; Arleen F Brown<sup>1</sup>, MD, PhD; Sara Simmons<sup>2</sup>, BSc; Peter Szilagyi<sup>4</sup>, MD, MPH

<sup>1</sup>Division of General Internal Medicine and Health Services Research, David Geffen School of Medicine, Los Angeles, CA, United States

<sup>2</sup>Los Angeles County Department of Health Services, Los Angeles, CA, United States

<sup>3</sup>Department of Family Medicine, David Geffen School of Medicine, Los Angeles, CA, United States

<sup>4</sup>Department of Pediatrics, David Geffen School of Medicine, Los Angeles, CA, United States

**Corresponding Author:**

Alejandra Casillas, MD, MSc

Division of General Internal Medicine and Health Services Research

David Geffen School of Medicine

1100 Glendon Avenue

Ste 850

Los Angeles, CA, 90024

United States

Phone: 1 310 794 0150

Email: [acasillas@mednet.ucla.edu](mailto:acasillas@mednet.ucla.edu)

## Abstract

Despite the implementation of internet patient portals into the safety net after the introduction of the Affordable Care Act in the United States, little attention has been paid to the process of engaging vulnerable patients into these portals. The portal is a health technology tool that was developed with a mainstream, English-speaking audience in mind. Thus, there are valid concerns that such technologies will actually exacerbate health care disparities, conferring further advantages to the already advantaged. In this paper, we describe a framework for portal engagement (awareness, registration, and use) among safety net patients. We incorporate the experiences in the Los Angeles County Department of Health Services to illustrate important contextual factors for portal outreach in our safety net. Finally, we discuss considerations for moving forward with health technology in the safety net as the next version of patient portals are being developed.

(*J Med Internet Res* 2020;22(10):e16835) doi:[10.2196/16835](https://doi.org/10.2196/16835)

**KEYWORDS**

patient portal; safety net; health disparities; digital divide

## Introduction: A Patient Portal Arrives to the Safety Net

Online patient health portals, also known as “patient portals” or “portals,” are two-way communication systems that are tethered to a patient’s electronic health record (EHR) [1]. Portals allow patients to manage many aspects of their health care from the convenience of their internet-connected device. Some studies have found that use of patient portals may improve quality of care and health outcomes, particularly for patients with chronic conditions such as diabetes mellitus [1-10].

Largely driven by the financial incentives of the Health Information Technology for Economic and Clinical Health Meaningful Use program (as part of the 2014 US federal health

care reform), patient portals have rapidly expanded [2,11-13]. In the last few years, many safety net systems—which are health systems that provide a significant level of care to minority, low-income, limited English proficient (LEP), and other vulnerable patients—have begun implementing patient portals. This innovation has given safety net health systems a new mechanism to directly share information and communicate with patients online. Since many vulnerable patients face barriers to in-person visits (such as taking unpaid time off from needed work), this telemedicine mechanism has potential for enhancing their care. Most notably, the coronavirus-19 (COVID-19) disease pandemic forced health systems to scale back on physical patient visits, and in-person patient education and engagement dramatically. With no established telemedicine workflows in place, this situation can exacerbate the health disparities for

these patients who are already at higher risk of poor disease management outcomes. Recent events have therefore highlighted the need to prioritize the integral role of the patient portal for care delivery in the Los Angeles safety net, and safety nets across the country.

Much of the prior literature related to patient portals and low-income populations has focused on the barriers posed by the digital divide; that is, the fact that many vulnerable populations lack the digital access, capacity, and interest to use a portal [14-28]. However, prior national studies suggest that the racial/ethnic and socioeconomic digital divide is shrinking, and that there are no racial/ethnic differences among people accessing the internet via mobile phones (about 60% of US adults) [28,29]. Furthermore, new data corroborate and expand upon prior work showing a high level of interest in portals among low-income, LEP, Medicaid, and public hospital patients [30-34]. Thus, the aforementioned barriers to the use of portals by safety net populations may be diminishing, and this population may be increasingly ready and eager to use the patient portal.

Despite these findings, little attention has been paid to the *process of engaging vulnerable patients* into an online health portal. This is especially problematic as the portal is a health technology tool that was developed with a mainstream, English-speaking audience in mind [13,35,36]. There are valid concerns that such technologies, including the patient portal, will actually exacerbate health care disparities, conferring “further advantages to the already advantaged,” a tenet of the Inverse Care Law [26].

As safety net health systems continue to implement patient portals [37], important questions remain about the factors that influence safety net portal registration and use, and the portal education strategies that will be effective among these vulnerable patients [13,15,35,38-40]. Health systems that are developing portals for vulnerable populations might benefit from an underlying framework that takes into consideration the unique needs of these populations.

In this paper, we describe our portal engagement process (awareness, registration, and use) for safety net patients. We then incorporate the experiences in the Los Angeles County Department of Health Services (LAC DHS) between 2015 and 2019 to illustrate important contextual factors in our safety net’s portal development. Finally, we discuss considerations for moving forward with health technology in the safety net as the next version of patient portals are being developed.

## Development of a Framework for Portal Engagement Among Vulnerable Patients in the LAC DHS

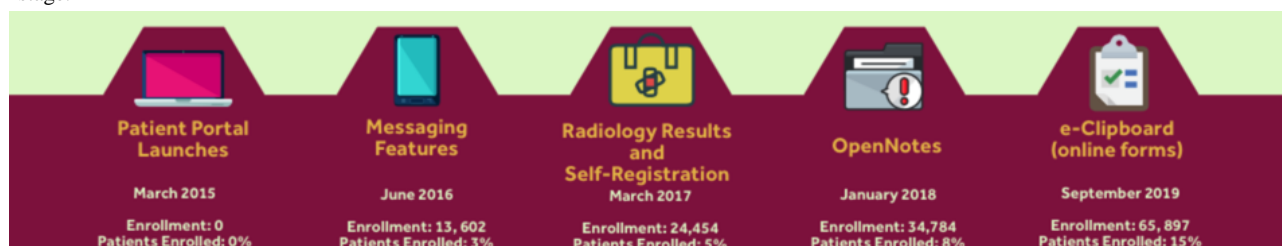
### Overview of Formative Findings

The LAC DHS, as the second largest municipal health safety net system in the United States, launched its English-Spanish patient portal in March 2015, with one of the few bilingual interfaces in the nation. The LAC DHS serves 600,000 unique patients each year, with 400,000 patients empaneled to its primary care clinics. Over half of the LAC DHS population is LEP, with the majority of these patients being Spanish speakers. At its inception, patients were able to view lab results, medication lists, and vital signs; additional features were added progressively. Figure 1 outlines the timing of various portal features in the LAC DHS *MyWellness* patient portal and accompanying portal registration data.

Many of these portal developments/improvements stemmed from patient input that we received on the frontlines. From January to March 2016, the LAC DHS performed a system-wide quality improvement internal survey of patients waiting in line for medical records. Patients were asked to fill out a paper or tablet survey to self-describe any internet access (including public and private access), knowledge about the *MyWellness* patient portal, and interest in health information on the *MyWellness* patient portal. We systematically randomly surveyed almost 200 patients, 73.0% (n=146) of whom reported having access to the internet. Only 20.0% (n=40) of patients were aware of the patient portal, and 45.0% (n=90) of those surveyed indicated interest in learning more about the portal.

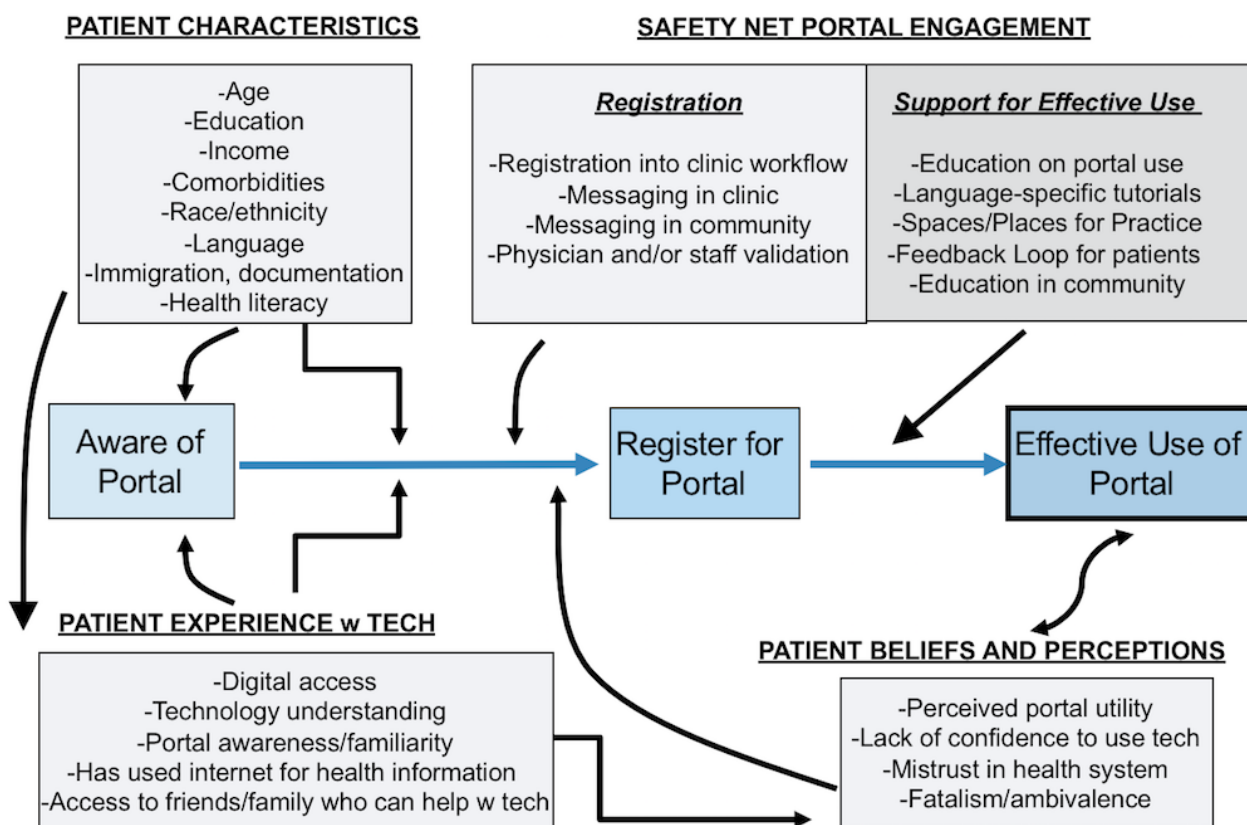
We conducted focus groups in the summer of 2017 among LAC DHS patients with chronic conditions to better understand perceptions of the portal and to obtain patient-centered recommendations for implementation [41]. Important themes from these focus groups were that we needed to provide dedicated patient guidance for both portal enrollment and portal navigation. One of the LAC DHS sites received a Catalyst human-centered design grant through the Center for Care Innovations to better understand the patient process and experience around trying to register for the patient portal. This formative work culminated in the development of a novel framework to outline factors affecting portal registration and use in the safety net, which is partly based on the technology acceptance model [42]. As shown in Figure 2, the framework outlines factors under the domains of *patient characteristics*, *patient experience with technology*, *patient beliefs and perceptions*, and *engagement initiatives by safety net health systems* that will impact (1) patient awareness of a patient portal, leading to (2) patient registration, and ultimately result in (3) sustained use of the portal by patients.

**Figure 1.** Los Angeles County Department of Health Services portal features timeline and percentage of empaneled primary care patients enrolled at each stage.



**Figure 2.** Implementation framework for portal registration and use among safety net patients.

## Factors that influence Portal Use in the Los Angeles Safety Net



Our formative findings and framework coincide with some of the conclusions provided by Grossman et al [43] in 2019. After reviewing over 100 studies about patient portal use among vulnerable populations, they found that individually focused interventions had the most evidence for increasing use in vulnerable populations. They recommended that research should “move beyond identifying disparities to systematically addressing them at multiple levels” for these patients [43]. In that vein, our work over the last 5 years has led to three actionable steps in facilitating portal uptake among our patients, which we now recommend as a model to other safety nets and EHR companies serving these patient populations:

### Step 1: Remove Unnecessary Patient Process Barriers in Enrolling for the Portal

Recent studies have provided recommendations on patient- and system-level interventions to increase registration and use of

the portal (ie, opt-out enrollment strategies, patient portal training) [43]. However, many of these recommendations will depend on the specific restrictions and parameters of the EHR vendor. For our patient population, the LAC DHS originally required an email address and a social security number (SSN) as a unique patient identifier (to receive a patient portal activation code). Privacy is paramount as it pertains to health data and security; however, authentication that requires an email address for portal enrollment can widen digital disparities to data access in the safety net. An acceptable alternative could be sending the activation code via a cell phone number (although this does not overcome the barrier that patients of low socioeconomic status may share phones). At one of our largest LAC DHS sites, patients between the ages of 18-64 have 75%-87% cell phone ownership and 49% of the patients aged >65 years report cell phone ownership. Authentication processes that require an SSN in safety net populations are



especially problematic, given that many immigrant patients may lack this information or even fear a process asking for it, owing to their (or their family members') documentation status. In response to these issues, we worked with our vendor to eliminate these steps and replace them with acceptable enrollment alternatives (eg, self-registration, medical record number identifiers).

*To improve equity to data access, we suggest that EHR vendors improve the ease of enrollment and give patients options for multifactor authentication that do not require an email address or secure/stigmatized information.*

## Step 2: Rethink Portal Engagement With a Patient-Centered Approach

The creation of LAC DHS Patient Portal Summits have brought together leadership, patient advocate groups, front desk staff, and clinicians to uniformly design engagement materials and staff incentive programs that best advance portal registration among our patients (even in low-resource settings). With support from a 2018 Center for Care Innovation's Catalyst grant, we created portal engagement materials that could eventually be administered by any LAC DHS staff member. Patients, nursing, staff, and physicians from all LAC DHS sites provided input over the course of 18 months. To date, this collaboration has resulted in a training curriculum for LAC DHS staff on portal engagement and engagement videos featuring the portal, developed by (and showing) patient and staff members and displayed in waiting rooms across our system in multiple languages [42,44]. From November 2018 to January 2019, a patient portal enrollment competition took place to further engage frontline staff in enrolling and engaging patients in the portal.

*Based on concepts that were prominent in our own formative work with patients, we suggest that safety net systems improve their portal messaging approaches to focus on: (a) validation of the portal by health care workers, (b) messaging about the portal that is useful to a patient's daily life and relevant to their personal health [45], and (c) educational scripts that incorporate family/community members.*

## Step 3: Partner With the EHR Vendor to Focus on an Appropriate User Interface

Our data show that the majority of LAC DHS patients who access the portal do so via their mobile devices (70%), whereas portals from nonsafety-net patients tend to more frequently be accessed via a desktop [28]. Pew Research surveys on smartphone use describe a phenomenon of "smartphone dependence" among low-income and minority populations [28]. Cable internet is increasingly expensive, and low-income communities increasingly depend on smartphone internet access for their overall online access. This is particularly important to keep in mind as patients are accessing health data. At the LAC DHS, we recognize that our internet strategies must be "mobile first." Accordingly, true mobile-friendly experiences became a ripe area for co-design with our patients, health system, and EHR vendors. In addition to prioritizing the improvement of our mobile app version of the *MyWellness* portal over the desktop app, and ensuring that *all* portal features are available

on the mobile version, we are also working with our EHR vendor on the following aspects: improving usability of the portal website for non-English speakers, making messaging more obvious on the website, improving patient education options to "learn more" about a lab or health condition, using multiple languages, creating a patient virtual feedback group, and forming a staff "super-user and champions" group to report back to our EHR vendor regarding usability concerns for our patients.

Another limitation has been the unwillingness of large portal vendors to allow the use of images from the portal in creation of training materials. This is a simple barrier that must be addressed, and just another example of how health systems should better partner with their EHR vendors on relevant patient portal operations and research [46]. We recognize that improving the user interface and usability will be a key challenge for this partnership, although there are published models that can help us better understand how users interact with this technology and the resources that may be necessary to support its use [47]. Portals are not alone in this limitation, as most digital health apps and platforms do not meet basic health literacy or language standards. Indeed, there are no clear standards for digital engagement in health care, and this gap must be addressed if the portal is to become completely accessible for diverse populations.

*Despite very specific standards for health literacy of written materials, this has not been translated to digital tools. Therefore, we recommend more transparency and inclusion around testing strategies to ensure that safety net populations are active participants in these testing phases.*

## Patient Portal 2.0: Reimagining the Next Generation of Portals for Improved Accessibility to Safety Net Patients

### Key Questions

If health technology is destined to serve even the most vulnerable patients, medical informatics research must answer the following questions: How can portals be made more accessible from the patient's perspective? What do safety net patients need from patient portals? By focusing on patients from vulnerable backgrounds, the safety net setting becomes a real-life laboratory for developing portals that will improve portal accessibility for all populations. We highlight the following patient-centered points (findings from our formative work) in reimagining the next version of portals.

### Connect to the Safety Net Patient's "Home" Team

Outside of the health care setting, there is already a "team" surrounding many of our patients, made up of family members and friends who serve as caregivers and trusted confidants for health decisions. Therefore we need to make it easier to connect digitally with these trusted team members (in addition to the individual patient) if the patient portal is to be used as a primary health management tool moving forward.

To first address this, there must be better options around "level of access" to patient data through proxy relationships, especially



when patient privacy and security remain a top level of concern among safety net patients. Our LAC DHS physicians have reported the need to provide limited portal views of “sensitive” information (eg, HIV results, intravenous drug use history) for patients who rely on family or friends as informal or formal caregivers. The option to share the portal with the patient’s team may be foregone in some cases because of the current “all-or-nothing” access approach to health information via a proxy login. One solution to this dilemma is to allow patients the ability to choose what level of proxy access a caregiver/family member will have. In reality, this is a feature that should be available to all patients, and will particularly resonate with adolescent and geriatric populations, and for patients with disabilities in other health care settings.

### Create a Virtual Home for Patient-Centered Care

What if the portal 2.0 were *more* than a data repository and messaging/scheduling hub? Patients are starting to fill out forms online through the patient portal, and such patient-generated data should be helpful to health care systems for tailored patient care and improvements to the system. What if the portal could also suggest content and experiences for patients? For example, a patient recently diagnosed with heart failure could receive suggestions from the portal for post-discharge follow-up videos, and health systems could track these engagement metrics. The portal could also push information to the patient about heart failure options for disease monitoring, such as online weight logs or a wireless scale. Patients with prediabetes who want support for lifestyle modification could be directed to locations for free or low-cost exercise programs in recreation centers or parks. Patients and providers have already suggested a desire for a portal that can collect patient-reported outcome metrics and deliver personalized feedback [48].

Currently, health systems are reactive; that is, we often wait for patients to approach our team with questions before we address health issues and socioemotional concerns. However, if a patient fills out a “Know Me” questionnaire and sets personal weight loss goals, the portal might be able to suggest culturally tailored nutrition and physical activity community resources for the patient based on their profile, even before visiting the clinic. As another example, if housing instability is listed as a problem in the EHR, the portal should potentially be able to push local resource notifications for housing or legal aid services. When “food insecurity” is recognized and coded in the EHR, could the portal then push notifications about food resources that are tailored to the patient’s home address or the geolocation on their phone? The safety net can help us re-envision the portal as a place where: (1) patients can be connected with social resources based on their needs, rooted in the social determinants of health; (2) patients can “check in” on health indicators, track the progression of personal indicators and goals; and (3) access culturally tailored and language-appropriate videos, podcasts, and written materials based on the patients’ underlying health conditions.

### Moving Beyond Meaningful Use

To achieve this vision of the patient portal, we must grade new portals based on metrics that move beyond meaningful use, which have previously only focused on patients’ log-in,

download, and exchange of data/messages with their health care team. EHR vendors are starting to understand the need to assist patients with self-management by exploring workflows to help patients fill out forms online to monitor chronic conditions, creating online interfaces for patients to log blood sugar and blood pressure levels, working to integrate devices such as smartwatches and connected devices (ie, wireless glucometers, blood pressure cuffs, weight scales), incorporating care manager and patient goals into their platforms, and integrating online platforms for community-based resources such as food banks, transportation assistance, and housing into the health record.

### *Summary: How Patient Portals Will Meet the Needs of Safety Net Patients*

Safety net health systems provide health care for our most medically and socially fragile patients: populations that include patients with multiple morbid conditions, LEP, cognitive impairment, high-risk perinatal needs, physical and mental disabilities, low literacy, homelessness, substance use, justice system-affected, and a broad range of immigrant and refugee communities. Safety nets are the ideal places to develop and refine the next iterations of the EHR and the patient portal. Because the portal is a “gateway” to the use of other digital health interventions, it is also an avenue for intervention research and a modality for better understanding of what vulnerable patients need to more effectively interface with health technology.

A limitation of this paper is that we did not delve into the continual barriers that sustain the digital divide (having access to reliable internet, limited devices and data/storage, and patients’ digital literacy). Although the digital divide seems to be shrinking, these recommendations should also be tempered with the knowledge that extreme disparities in internet access continue to exist for low-income populations, especially in certain areas of the country [49]. All of these systemic and patient-centered barriers should be addressed in a version of patient portal development that is more inclusive. To make this tool work for our most vulnerable populations, moving forward, we need to take intentional steps to ensure that the patient portal can be effectively and efficiently deployed in health systems that serve these high-risk patients. To achieve these goals, we must (1) remove unnecessary patient process barriers in enrolling for the portal; (2) redesign engagement materials with a patient-centered approach; (3) partner with EHR vendors to focus on the user interface and usability from a safety net patient perspective; (4) engage trusted family members and caregivers to create a flexible, patient-friendly mechanism for proxy access; (5) create a virtual home for patient-centered care that includes addressing social determinants, preventive care, and chronic care; and (6) redefine the metrics of portal success, as seen in the safety net.

Finally, we would be remiss in not emphasizing that health systems around the country are quickly developing remote strategies to reach out to patients for health management resources and education as in-person services shrink, secondary to the COVID-19 pandemic, and that the patient portal is an integral part of this outreach plan. Thus, redesigning the patient

portal so that it effectively reaches and impacts our most vulnerable patients is an important step to improve health care access for the entire US population. These uncertain and challenging times in our history are an opportunity to

significantly move the needle in digital health, and create a patient portal that works for even the most vulnerable patients in this new era of healthcare— making sure that no patient is left behind.

## Acknowledgments

AC is supported by the AHRQ/PCORI KL12HS0260407-01(MOI) award via the Stakeholder-Partnered Implementation Research and Innovation Translation (Spirit) Program. The authors acknowledge scholarly feedback from faculty members of the UCLA Clinical and Translational Science Institute (CTSI) Patient Portal Research Workgroup, which is supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through grant number UL1TR001881.

## Conflicts of Interest

None declared.

## References

1. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. *J Med Internet Res* 2015 Feb 10;17(2):e44 [FREE Full text] [doi: [10.2196/jmir.3171](https://doi.org/10.2196/jmir.3171)] [Medline: [25669240](https://pubmed.ncbi.nlm.nih.gov/25669240/)]
2. Kamo N, Bender AJ, Kalmady K, Blackmore CC. Meaningful use of the electronic patient portal - Virginia Mason's journey to create the perfect online patient experience. *Healthc (Amst)* 2017 Dec;5(4):221-226. [doi: [10.1016/j.hjdsi.2016.09.003](https://doi.org/10.1016/j.hjdsi.2016.09.003)] [Medline: [27727028](https://pubmed.ncbi.nlm.nih.gov/27727028/)]
3. Goldzweig C, Towfigh A, Paige N, Orshansky G, Haggstrom D, Beroes J, Evidence-based Synthesis Program. Systematic Review: Secure Messaging Between Providers and Patients, and Patients' Access to Their Own Medical Record. Evidence on Health Outcomes, Satisfaction, Efficiency and Attitudes. Washington DC: Department of Veterans Affairs (US); 2012.
4. Ralston JD, Hirsch IB, Hoath J, Mullen M, Cheadle A, Goldberg HI. Web-based collaborative care for type 2 diabetes: a pilot randomized trial. *Diabetes Care* 2009 Feb;32(2):234-239 [FREE Full text] [doi: [10.2337/dc08-1220](https://doi.org/10.2337/dc08-1220)] [Medline: [19017773](https://pubmed.ncbi.nlm.nih.gov/19017773/)]
5. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. *Med Care* 2014 Mar;52(3):194-201 [FREE Full text] [doi: [10.1097/MLR.0000000000000069](https://doi.org/10.1097/MLR.0000000000000069)] [Medline: [24374412](https://pubmed.ncbi.nlm.nih.gov/24374412/)]
6. Harris LT, Haneuse SJ, Martin DP, Ralston JD. Diabetes quality of care and outpatient utilization associated with electronic patient-provider messaging: a cross-sectional analysis. *Diabetes Care* 2009 Jul;32(7):1182-1187 [FREE Full text] [doi: [10.2337/dc08-1771](https://doi.org/10.2337/dc08-1771)] [Medline: [19366959](https://pubmed.ncbi.nlm.nih.gov/19366959/)]
7. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. *J Gen Intern Med* 2012 Apr;27(4):420-424 [FREE Full text] [doi: [10.1007/s11606-011-1889-0](https://doi.org/10.1007/s11606-011-1889-0)] [Medline: [22005937](https://pubmed.ncbi.nlm.nih.gov/22005937/)]
8. Reed M, Huang J, Graetz I, Brand R, Hsu J, Fireman B, et al. Outpatient electronic health records and the clinical care and outcomes of patients with diabetes mellitus. *Ann Intern Med* 2012 Oct 2;157(7):482-489 [FREE Full text] [doi: [10.7326/0003-4819-157-7-201210020-00004](https://doi.org/10.7326/0003-4819-157-7-201210020-00004)] [Medline: [23027319](https://pubmed.ncbi.nlm.nih.gov/23027319/)]
9. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. *J Am Med Inform Assoc* 2008;15(1):1-7 [FREE Full text] [doi: [10.1197/jamia.M2562](https://doi.org/10.1197/jamia.M2562)] [Medline: [17947615](https://pubmed.ncbi.nlm.nih.gov/17947615/)]
10. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. *Health Aff (Millwood)* 2010 Jul;29(7):1370-1375. [doi: [10.1377/hlthaff.2010.0048](https://doi.org/10.1377/hlthaff.2010.0048)] [Medline: [20606190](https://pubmed.ncbi.nlm.nih.gov/20606190/)]
11. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *N Engl J Med* 2010 Aug 05;363(6):501-504. [doi: [10.1056/NEJMp1006114](https://doi.org/10.1056/NEJMp1006114)] [Medline: [20647183](https://pubmed.ncbi.nlm.nih.gov/20647183/)]
12. 2014 National Healthcare Quality & Disparities Report. Agency for Healthcare Research and Quality. Rockville, MD; 2014 Oct. URL: <https://archive.ahrq.gov/research/findings/nhqdr/nhqdr14/index.html> [accessed 2019-10-20]
13. Lyles CR, Fruchterman J, Youdelman M, Schillinger D. Legal, Practical, and Ethical Considerations for Making Online Patient Portals Accessible for All. *Am J Public Health* 2017 Oct;107(10):1608-1611. [doi: [10.2105/AJPH.2017.303933](https://doi.org/10.2105/AJPH.2017.303933)] [Medline: [28817324](https://pubmed.ncbi.nlm.nih.gov/28817324/)]
14. Sequist TD. Health information technology and disparities in quality of care. *J Gen Intern Med* 2011 Oct;26(10):1084-1085 [FREE Full text] [doi: [10.1007/s11606-011-1812-8](https://doi.org/10.1007/s11606-011-1812-8)] [Medline: [21809173](https://pubmed.ncbi.nlm.nih.gov/21809173/)]
15. Smith SG, O'Connor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. *J Am Med Inform Assoc* 2015 Jul;22(4):888-895 [FREE Full text] [doi: [10.1093/jamia/ocv025](https://doi.org/10.1093/jamia/ocv025)] [Medline: [25914099](https://pubmed.ncbi.nlm.nih.gov/25914099/)]

16. Arora S, Ford K, Terp S, Abramson T, Ruiz R, Camilon M, et al. Describing the evolution of mobile technology usage for Latino patients and comparing findings to national mHealth estimates. *J Am Med Inform Assoc* 2016 Sep;23(5):979-983 [FREE Full text] [doi: [10.1093/jamia/ocv203](https://doi.org/10.1093/jamia/ocv203)] [Medline: [26995564](#)]
17. File T. Digital Divides: A connectivity continuum for the United States. Data from the Current Population Survey. 2012 Sep 19. URL: <http://paa2013.princeton.edu/papers/130743> [accessed 2019-10-20] [WebCite Cache ID [https://paa2013.princeton.edu/papers/130743](#)]
18. File T. Digital Divides: A Connectivity Continuum for the United States. : Social and Economic Statistics Division, U.S. Census Bureau, U.S. Department of Commerce; 2013 Apr 11 Presented at: Annual Meeting of the Population Association of America; April 2013; New Orleans, LA p. 11-13.
19. File T. Computer and Internet Use in the United States. In: *Current Population Survey Reports*, P20-568. Washington, DC: US Census Bureau; 2013:1-14.
20. File T, Ryan C. Computer and Internet Use in the United States: 2013. *American Community Survey Reports*, ACS-28. U.S. Census Bureau, Washington DC. 2013 May 05. URL: <https://www.census.gov/history/pdf/acs-internet2013.pdf> [accessed 2020-09-05]
21. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. *J Med Internet Res* 2014 Jul 16;16(7):e172 [FREE Full text] [doi: [10.2196/jmir.3117](https://doi.org/10.2196/jmir.3117)] [Medline: [25048379](#)]
22. Laz TH, Berenson AB. Racial and ethnic disparities in internet use for seeking health information among young women. *J Health Commun* 2013;18(2):250-260 [FREE Full text] [doi: [10.1080/10810730.2012.707292](https://doi.org/10.1080/10810730.2012.707292)] [Medline: [23130608](#)]
23. Livingston G. Latinos and digital technology, 2010. Pew Research Center, Hispanic Trends. URL: <http://www.pewhispanic.org/2011/02/09/latinos-and-digital-technology-2010/> [accessed 2018-02-20]
24. Perzynski AT, Roach MJ, Shick S, Callahan B, Gunzler D, Cebul R, et al. Patient portals and broadband internet inequality. *J Am Med Inform Assoc* 2017 Sep 01;24(5):927-932 [FREE Full text] [doi: [10.1093/jamia/ocx020](https://doi.org/10.1093/jamia/ocx020)] [Medline: [28371853](#)]
25. Ratanawongsa N, Barton JL, Lyles CR, Wu M, Yelin EH, Martinez D, et al. Computer use, language, and literacy in safety net clinic communication. *J Am Med Inform Assoc* 2017 Jan;24(1):106-112 [FREE Full text] [doi: [10.1093/jamia/ocw062](https://doi.org/10.1093/jamia/ocw062)] [Medline: [27274017](#)]
26. Schillinger D. Literacy and health communication: reversing the 'inverse care law'. *Am J Bioeth* 2007 Nov;7(11):15-8; discussion W1. [doi: [10.1080/15265160701638553](https://doi.org/10.1080/15265160701638553)] [Medline: [18027291](#)]
27. Turner-Lee N, Brian S, Joseph M. Minorities Mobile Broadband and the Management of Chronic Diseases. Washington, DC: Joint Center for Political and Economic Studies; 2012.
28. Zickuhr K, Smith A. Digital Differences. Pew Research Center, Internet & Technology. Washington, DC: Pew Internet and American Life Project; 2012 Apr 13. URL: <http://www.pewinternet.org/2012/04/13/digital-differences/> [accessed 2018-02-20]
29. Gonzalez M, Sanders-Jackson A, Emory J. Online Health Information-Seeking Behavior and Confidence in Filling Out Online Forms Among Latinos: A Cross-Sectional Analysis of the California Health Interview Survey, 2011-2012. *J Med Internet Res* 2016 Jul 04;18(7):e184 [FREE Full text] [doi: [10.2196/jmir.5065](https://doi.org/10.2196/jmir.5065)] [Medline: [27377466](#)]
30. Dhanireddy S, Walker J, Reisch L, Oster N, Delbanco T, Elmore JG. The urban underserved: attitudes towards gaining full access to electronic medical records. *Health Expect* 2014 Oct;17(5):724-732 [FREE Full text] [doi: [10.1111/j.1369-7625.2012.00799.x](https://doi.org/10.1111/j.1369-7625.2012.00799.x)] [Medline: [22738155](#)]
31. Lobach DF, Willis JM, Macri JM, Simo J, Anstrom KJ. Perceptions of Medicaid beneficiaries regarding the usefulness of accessing personal health information and services through a patient Internet portal. *AMIA Annu Symp Proc* 2006:509-513 [FREE Full text] [Medline: [17238393](#)]
32. Mirsky JB, Tieu L, Lyles C, Sarkar U. A Mixed-Methods Study of Patient-Provider E-Mail Content in a Safety-Net Setting. *J Health Commun* 2016;21(1):85-91 [FREE Full text] [doi: [10.1080/10810730.2015.1033118](https://doi.org/10.1080/10810730.2015.1033118)] [Medline: [26332306](#)]
33. Ochoa A, Kitayama K, Uijtdehaage S, Vermillion M, Eaton M, Carpio F, et al. Patient and provider perspectives on the potential value and use of a bilingual online patient portal in a Spanish-speaking safety-net population. *J Am Med Inform Assoc* 2017 Nov 01;24(6):1160-1164. [doi: [10.1093/jamia/ocx040](https://doi.org/10.1093/jamia/ocx040)] [Medline: [28460130](#)]
34. Schickedanz A, Huang D, Lopez A, Cheung E, Lyles CR, Bodenheimer T, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. *J Gen Intern Med* 2013 Jul;28(7):914-920 [FREE Full text] [doi: [10.1007/s11606-012-2329-5](https://doi.org/10.1007/s11606-012-2329-5)] [Medline: [23423453](#)]
35. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. *J Med Internet Res* 2015 Jun 23;17(6):e148 [FREE Full text] [doi: [10.2196/jmir.4255](https://doi.org/10.2196/jmir.4255)] [Medline: [26104044](#)]
36. Lyles C, Schillinger D, Sarkar U. Connecting the Dots: Health Information Technology Expansion and Health Disparities. *PLoS Med* 2015 Jul;12(7):e1001852 [FREE Full text] [doi: [10.1371/journal.pmed.1001852](https://doi.org/10.1371/journal.pmed.1001852)] [Medline: [26172977](#)]
37. Casillas A, Moreno G, Grotts J, Tseng C, Morales LS. A Digital Language Divide? The Relationship between Internet Medication Refills and Medication Adherence among Limited English Proficient (LEP) Patients. *J Racial Ethn Health Disparities* 2018 Dec;5(6):1373-1380 [FREE Full text] [doi: [10.1007/s40615-018-0487-9](https://doi.org/10.1007/s40615-018-0487-9)] [Medline: [29600351](#)]

38. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. *J Med Internet Res* 2015 Dec 03;17(12):e275 [FREE Full text] [doi: [10.2196/jmir.4847](https://doi.org/10.2196/jmir.4847)] [Medline: [26681155](https://pubmed.ncbi.nlm.nih.gov/26681155/)]
39. Tieu L, Schillinger D, Sarkar U, Hoskote M, Hahn KJ, Ratanawongsa N, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? *J Am Med Inform Assoc* 2017 Apr 01;24(e1):e47-e54 [FREE Full text] [doi: [10.1093/jamia/ocw098](https://doi.org/10.1093/jamia/ocw098)] [Medline: [27402138](https://pubmed.ncbi.nlm.nih.gov/27402138/)]
40. Sheon AR, Bolen SD, Callahan B, Shick S, Perzynski AT. Addressing Disparities in Diabetes Management Through Novel Approaches to Encourage Technology Adoption and Use. *JMIR Diabetes* 2017 Jul 13;2(2):e16 [FREE Full text] [doi: [10.2196/diabetes.6751](https://doi.org/10.2196/diabetes.6751)] [Medline: [30291090](https://pubmed.ncbi.nlm.nih.gov/30291090/)]
41. Casillas A, Perez-Aguilar G, Abhat A, Gutierrez G, Olmos-Ochoa TT, Mendez C, et al. Su salud a la mano (your health at hand): patient perceptions about a bilingual patient portal in the Los Angeles safety net. *J Am Med Inform Assoc* 2019 Dec 01;26(12):1525-1535 [FREE Full text] [doi: [10.1093/jamia/ocz115](https://doi.org/10.1093/jamia/ocz115)] [Medline: [31373362](https://pubmed.ncbi.nlm.nih.gov/31373362/)]
42. Scherer R, Siddiq F, Tondeur J. The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Comput Educ* 2019 Jan;128:13-35. [doi: [10.1016/j.compedu.2018.09.009](https://doi.org/10.1016/j.compedu.2018.09.009)]
43. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. *J Am Med Inform Assoc* 2019 Aug 01;26(8-9):855-870 [FREE Full text] [doi: [10.1093/jamia/ocz023](https://doi.org/10.1093/jamia/ocz023)] [Medline: [30958532](https://pubmed.ncbi.nlm.nih.gov/30958532/)]
44. Quach T. Mrs Garcia Intro. YouTube. URL: <https://www.youtube.com/watch?v=kIWZFRMtPD8&feature=youtu.be> [accessed 2019-10-20]
45. Nambisan P. Factors that impact Patient Web Portal Readiness (PWPR) among the underserved. *Int J Med Inform* 2017 Jun;102:62-70. [doi: [10.1016/j.ijmedinf.2017.03.004](https://doi.org/10.1016/j.ijmedinf.2017.03.004)] [Medline: [28495349](https://pubmed.ncbi.nlm.nih.gov/28495349/)]
46. Strudwick G, Kuziemy C, Booth R, Rossetti S, Chyjek A, Sakal M, et al. Three Methods for Engaging Patients and Care Partners in Patient Portal Research. *Stud Health Technol Inform* 2019 Aug 21;264:1984-1985. [doi: [10.3233/SHTI190746](https://doi.org/10.3233/SHTI190746)] [Medline: [31438440](https://pubmed.ncbi.nlm.nih.gov/31438440/)]
47. Yen P, Walker DM, Smith JMG, Zhou MP, Menser TL, McAlearney AS. Usability evaluation of a commercial inpatient portal. *Int J Med Inform* 2018 Feb;110:10-18. [doi: [10.1016/j.ijmedinf.2017.11.007](https://doi.org/10.1016/j.ijmedinf.2017.11.007)] [Medline: [29331248](https://pubmed.ncbi.nlm.nih.gov/29331248/)]
48. Cronin RM, Conway D, Condon D, Jerome RN, Byrne DW, Harris PA. Patient and healthcare provider views on a patient-reported outcomes portal. *J Am Med Inform Assoc* 2018 Nov 01;25(11):1470-1480 [FREE Full text] [doi: [10.1093/jamia/ocy111](https://doi.org/10.1093/jamia/ocy111)] [Medline: [30239733](https://pubmed.ncbi.nlm.nih.gov/30239733/)]
49. Callahan B. Worst Connected Cities of 2018. National Digital Inclusion Alliance. URL: <https://www.digitalinclusion.org/blog/2019/10/23/worst-connected-cities-of-2018/> [accessed 2020-06-07]

## Abbreviations

**COVID-19:** coronavirus disease 2019  
**EHR:** electronic health record  
**LAC DHS:** Los Angeles County Department of Health Services  
**LEP:** limited English proficient  
**SSN:** social security number

*Edited by G Eysenbach; submitted 29.10.19; peer-reviewed by A Sheon, G Germinario; comments to author 28.02.20; revised version received 25.06.20; accepted 07.07.20; published 23.10.20.*

### *Please cite as:*

Casillas A, Abhat A, Mahajan A, Moreno G, Brown AF, Simmons S, Szilagyi P  
 Portals of Change: How Patient Portals Will Ultimately Work for Safety Net Populations  
*J Med Internet Res* 2020;22(10):e16835  
 URL: <https://www.jmir.org/2020/10/e16835>  
 doi:[10.2196/16835](https://doi.org/10.2196/16835)  
 PMID:[33094732](https://pubmed.ncbi.nlm.nih.gov/33094732/)

©Alejandra Casillas, Anshu Abhat, Anish Mahajan, Gerardo Moreno, Arleen F Brown, Sara Simmons, Peter Szilagyi. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of

Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Viewpoint

# Proposed Implementation of Blockchain in British Columbia's Health Care Data Management

Danielle Cadoret<sup>1</sup>, BSc; Tamara Kailas<sup>1</sup>, BSc; Pedro Velmovitsky<sup>2</sup>, MSc; Plinio Morita<sup>2,3,4,5,6</sup>, PhD; Okechukwu Igboeli<sup>1</sup>, DVM, MSc, PhD, MBA

<sup>1</sup>Science and Business Program, Faculty of Science, University of Waterloo, Waterloo, ON, Canada

<sup>2</sup>School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

<sup>3</sup>Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

<sup>4</sup>Research Institute for Aging, University of Waterloo, Waterloo, ON, Canada

<sup>5</sup>Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

<sup>6</sup>eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

**Corresponding Author:**

Okechukwu Igboeli, DVM, MSc, PhD, MBA

Science and Business Program

Faculty of Science

University of Waterloo

200 University Avenue West

Waterloo, ON, N2L 3G1

Canada

Phone: 1 519 888 4567 ext 38769

Fax: 1 519 746 2543

Email: [oigboeli@uwaterloo.ca](mailto:oigboeli@uwaterloo.ca)

## Abstract

**Background:** There are several challenges such as information silos and lack of interoperability with the current electronic medical record (EMR) infrastructure in the Canadian health care system. These challenges can be alleviated by implementing a blockchain-based health care data management solution.

**Objective:** This study aims to provide a detailed overview of the current health data management infrastructure in British Columbia for identifying some of the gaps and inefficiencies in the Canadian health care data management system. We explored whether blockchain is a viable option for bridging the existing gaps in EMR solutions in British Columbia's health care system.

**Methods:** We constructed the British Columbia health care data infrastructure and health information flow based on publicly available information and in partnership with an industry expert familiar with the health systems information technology network of British Columbia's Provincial Health Services Authorities. Information flow gaps, inconsistencies, and inefficiencies were the target of our analyses.

**Results:** We found that hospitals and clinics have several choices for managing electronic records of health care information, such as different EMR software or cloud-based data management, and that the system development, implementation, and operations for EMRs are carried out by the private sector. As of 2013, EMR adoption in British Columbia was at 80% across all hospitals and the process of entering medical information into EMR systems in British Columbia could have a lag of up to 1 month. During this lag period, disease progression updates are continually written on physical paper charts and not immediately updated in the system, creating a continuous lag period and increasing the probability of errors and disjointed notes. The current major stumbling block for health care data management is interoperability resulting from the use of a wide range of unique information systems by different health care facilities.

**Conclusions:** Our analysis of British Columbia's health care data management revealed several challenges, including information silos, the potential for medical errors, the general unwillingness of parties within the health care system to trust and share data, and the potential for security breaches and operational issues in the current EMR infrastructure. A blockchain-based solution has the highest potential in solving most of the challenges in managing health care data in British Columbia and other Canadian provinces.

**KEYWORDS**

blockchain; electronic medical records; health data management; patient centric

## Introduction

### Background

Blockchain is a relatively new technology with the potential to revolutionize the way data are secured and managed. Implementing blockchain in health care opens diverse possibilities in data analytics, cross-province health data transfer, and patient-centered health care. It also helps resolve many of the existing challenges faced in health care data management, such as eliminating information silos, increasing efficiency in data transfer between health care facilities, data sharing, and data protection [1,2]. Challenges associated with blockchain implementation in the Canadian health care system include complicated regulations surrounding health and personal data, the government's universal involvement, and transitional issues between 2 information systems. Identifying gaps in the current infrastructure and the problems that blockchain can solve and creating a realistic implementation plan for blockchain systems in the Canadian health care system will require an in-depth strategic analysis that will inform viability [1,3].

### What Is Blockchain?

Blockchain technology is a distributed ledger system that can be used for the management and exchange of information among members of a network [4,5]. Each participating computer in the network forms an independent node that maintains a copy of the ledger, which is automatically updated among all members whenever there is an addition to the ledger.

A transaction in the network is timestamped and cryptographically sealed in a *block* and then added to the chain of previous blocks (representing previous transactions) by an automated validation mechanism [4]. When a block is added to the chain, a cryptographic technique called a hash is used to ensure that the connection is immutable. Hashes are one-way digital *signatures* comprising a string of numbers and letters, and each hash is based on the hash of the preceding block. Hashes are what give blockchain its immutability, an important feature that prevents malicious changes to the blockchain [4].

Another cryptographic technique, zero knowledge proof (ZKP) systems, is often combined with blockchain to reduce the amount of data needed to be shared between parties. ZKPs are interactive proofs that have the ability to yield nothing but the validity of a claim [6-8]. A public blockchain can be joined by anyone, whereas a permissioned and private blockchain requires permission from the governing nodes of the network to join, thus upholding trust and security of the network [4]. The term *distributed* refers to the way that the blockchain or ledger is stored in the network: rather than being stored in 1 database, a copy of the blockchain is stored with each member of the network [4]. It is important to indicate that this study focuses on the use of private blockchains; public blockchains are not built to accommodate or safeguard sensitive information

properly (eg, transactions involving medical records) and, as such, are not a viable solution for this scenario.

### Current Challenges in Health Care

Electronic medical records (EMRs) have a large potential to benefit from blockchain systems. Information in EMRs may include patient diagnoses and histories as well as test and imaging results. However, the fragmentation of digital data occurs when various points of care use different EMR systems to store patient data, which are not interoperable [9]. Patients visit numerous health care offices throughout their lives, leading to fragmented medical records, which in turn limit the information available to practitioners and service providers. Ultimately, this affects the quality of care received by the patients and patients' experience with the health care system. It becomes more difficult for health care practitioners to see the complete picture of a patient's care, and this often results in redundant procedures and history taking, which takes a toll on the health care system and the patient.

Data sharing between care providers improves the accuracy of diagnoses [10-12] and reduces errors in treatment plans [13]. In addition, with telemedicine on the rise, there's a need to identify a way that allows efficient and secure sharing of patient data and consent [14]. Currently, there are no existing interoperable EMRs as a pan-Canadian endeavor involving the support and collaboration of federal and provincial governments to minimize the presence of information silos. Each province and territory is responsible for developing its own EMR strategy, leading to a wide range of differences in EMRs throughout Canada, which reduces interoperability [1,2,15].

In terms of national efforts to increase interoperability between EMRs, the federal government issued a Can \$1.6 billion (US \$1.2 billion) grant in 2001 to establish the Canada Health Infoway (Infoway), an agency specifically formed to spearhead the EMR initiative [16]. Infoway partnered with the Canadian Institute for Health Information (CIHI), an existing agency working to improve Canadian health care based on informatics, to establish national standards and guidelines for provinces to follow in their EMR adoption [17]. Despite both these agencies working together, the problem of interoperability prevails both within and across provinces. This is partly because of the variability in government-approved EMRs, which prevents complete standardization [16].

The medium in which patient health data are stored varies across provinces, regions, and institutions. These data can be stored as paper records or EMRs, which can generally be referred to as a point-of-care service applications (PCSAs). At this point, data are stored at an institutional level wherein accessibility to external parties is limited [18]. External parties may include other medical clinics, hospitals, or institutions. CIHI outlines that data stored in the PCSA may feed into the integrated assessment record (IAR). The IAR is a tool that enables a centralized repository for patient health data collected from

separate centers of care with differing PCSAs [18]. Ideally, physicians and health professionals from various institutions can access the IAR, allowing interoperability. However, successful implementation and adoption of the IAR are not well documented for health departments outside long-term care, mental health care, or community health and support services [18]. These sectors capture only a portion of the required health services and do not span the Canadian health care population. Two variables that contribute to the limited success of IARs are the lack of digitization of paper health records and the lack of standardization for PCSAs.

In addition, although new technology attempts to solve these issues, there is also the problem of physicians' resistance to change [19]. Many health care professionals do not trust computers and tablets to store information because they feel it is not as reliable as a pen and paper. Data loss, computer malfunctions, and breaches of privacy are often cited by those protesting against digitizing information. Furthermore, when hospital administrations decide to implement an EMR, many health care professionals do not adopt the software correctly and continue using their paper charts and dictaphones. This leads to a strain on time and resources and increases redundancy and errors.

Finally, a major issue is the ownership of medical records. Currently, patient records are either owned by the hospital or by the doctor in a private practice setting, and the onus is on these parties to keep records private and safe. With a large amount of evidence showing that patient-centric health care improves outcomes, there is a push toward letting patients access and own their records. This gives patients the ability to give full or partial controlled access to their records whenever and to whomever they want. However, patient-centric health care is impossible with the current health care data infrastructure. EMRs were simply not designed with the patient in mind, and their workflows and interfaces completely exclude the patient. Flipping the current infrastructure around to accommodate patient-centric health care would require completely redesigned systems.

## Goal of the Study

This study aims to provide a detailed overview of the current *status quo* of the health data management infrastructure in British Columbia to identify gaps and inefficiencies in the system. This review explores whether blockchain is a viable option for the existing gaps in EMR solutions in the Canadian health care system. This study aims to analyze the benefits of a blockchain-based data solution as well as the feasibility of switching to this technology. The expected results of this study include an implementation plan for a blockchain-based solution, which can guide future parties interested in moving forward toward a more universal, efficient, and integrated patient-centric health care data management system.

## Methods

### Flow of Health Data in British Columbia

The British Columbia health care data infrastructure was chosen as a model for blockchain implementation in this study because

of the availability of an industry expert who provided insight into the nuances of British Columbia's structure and processes. This expert is well informed of the health systems information technology (IT) infrastructure of British Columbia's Provincial Health Services Authorities (PHSAs) and is knowledgeable of the realities of health data management. This consultation supplemented details regarding the *status quo* and health information flow within the province that would not have been accessible to public domains. Information flow gaps, inconsistencies, and inefficiencies were the target of the analyses.

An important metric to measure the viability of a blockchain solution is the digitization rate of paper records. Blockchain requires digital data; without digitized records, a blockchain system cannot be implemented.

## Blockchain and Existing Solutions

Existing technological solutions for health care data management include the following:

1. IT solutions: rely on host EMR systems that allow health data to be stored on local computers (designated servers). The hosted EMR system allows a facility to take ownership of their health data [20].
2. Cloud computing: uses a network of remote servers that store and manage information. Unlike onsite storage of information or physical paper charts, data are stored outside the health facility with cloud computing [20].
3. MedRec: an existing blockchain-based EMR system developed by the Massachusetts Institute of Technology (MIT) that stores references to the off-chain location of medical records [2,21,22].

The integration of blockchain with these existing solutions and Canada's existing health care data infrastructure was explored using published literature and consultations with health IT and blockchain domain experts.

The implementation plan was designed on the basis of this information. It focused on bridging the gaps in British Columbia's current infrastructure and easing pain points of patients, health care professionals, and IT personnel. It aims to move toward a blockchain-based, universal, patient-centric health care data management system in Canada, taking existing infrastructure and ongoing health data needs into account.

## Results

### Digitization and the EMR Status Quo in British Columbia

The Ministry of Health in British Columbia (BCMOH) is the jurisdictional body responsible for EMR implementation. Their goal is to deliver relevant data to health care professionals, putting in place an infrastructure for health information sharing and thus increasing interoperability. The BCMOH established the Physician Information Technology Office (PITO) to launch the transition from paper records to EMRs in British Columbia [23]. The BCMOH's primary task is to provide systems management and recommend EMR systems for physician offices to effectively replace paper-based records. These are all

performed under Infoway's guidelines to improve interoperability.

Currently, the system development, implementation, and operations of EMRs are carried out by the private sector. In 2007, the BCMOH partnered with Sun Microsystems to build, design, implement, and operate a British Columbia-specific EMR system. Although this has not yet surfaced, the province currently uses 4 government-approved EMR applications:

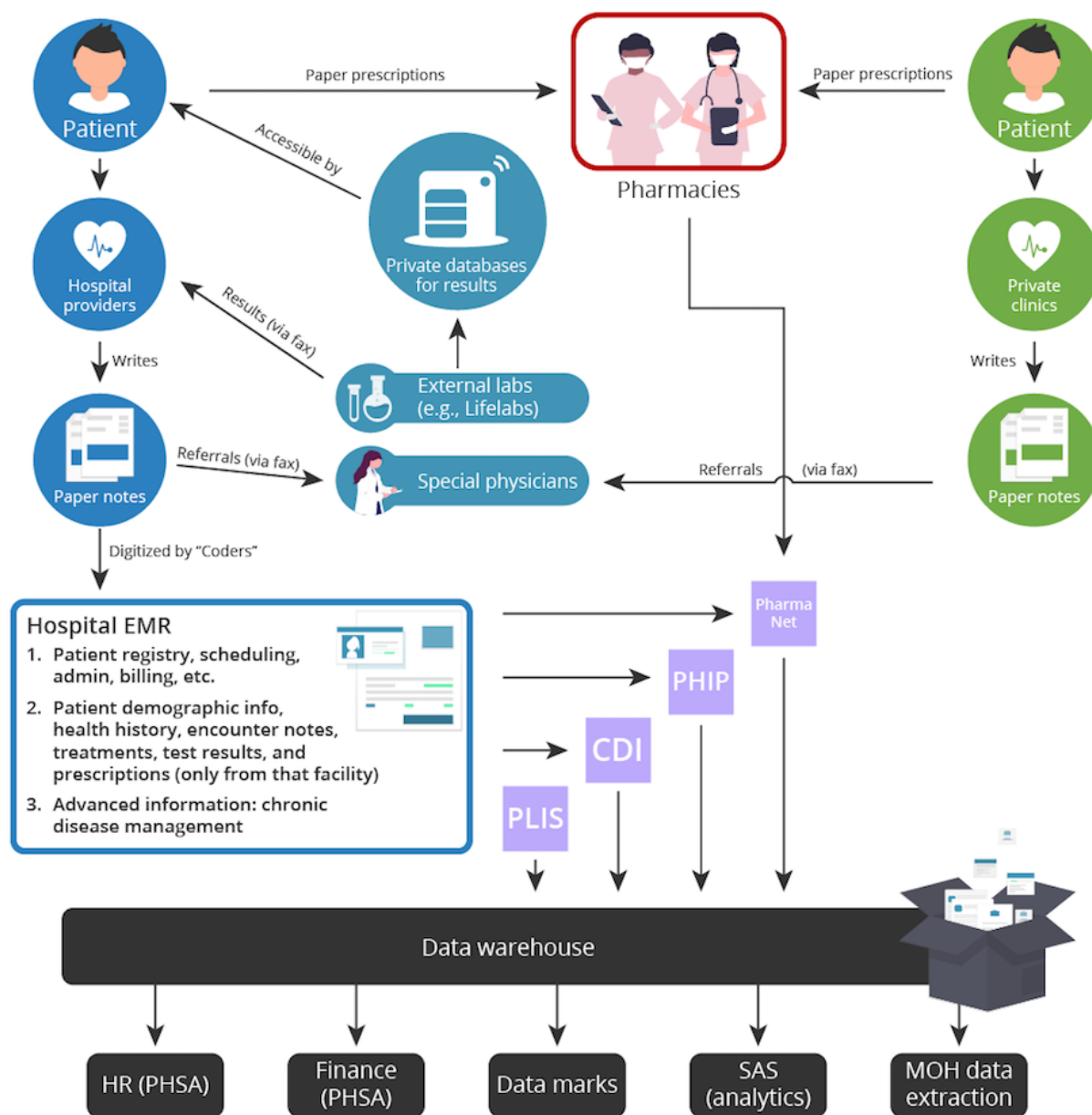
1. Intrahealth: a web-based IT solution that has various modes of deployment and databases that can be customized for the physicians' preferences [24].
2. Med Access: a web-based system that configures individual clinics and user preferences. The platform adapts to the workflow of those using it through tailoring based on role, group, and organization. The software is designed to use the internet so that patient information is not stored on the end user's computer [25]. Beyond acting as an e-chart, the system also allows for point-of-care decision support through reminders and prompts. Furthermore, interprovider communication is available with other caregivers who are also using Med Access EMR [26].
3. Wolf EMR: a cloud-based EMR system that manages and stores patient data in a customizable manner. It allows physicians to create queries and use support tools within each patient file to optimize the care provided [26].
4. Osler Systems: information on this system was not publicly available.

As of 2013, EMR adoption in British Columbia was at 80% across all hospitals [27]. Adoption was defined as the initial implementation of an EMR system in a facility but does not exclude the use of paper records. Consequently, the focus has shifted toward more optimal use of EMRs and achieving interconnectivity across various points of care.

### Flow of Health Data in British Columbia

A snapshot of the health care information flow in the British Columbia's PHSA (BCPHSA) is presented in [Figure 1](#). Patient information is introduced to the health care system when an individual visits a hospital or private clinic. These scenarios are indicated by the 2 stars. In a hospital setting, the information flow begins when a patient gives information to health care workers, such as the reason for their visit, demographic information, medical history, symptoms, and more. Patient information gathered from these consultations can then (1) be directly added into an EMR system based on technology integration at a specific hospital or (2) be first recorded on physical paper charts or dictaphones and then transcribed to an electronic format later. In the latter, the process of entering these physical records into the EMR itself could have a lag of up to 1 month. During this lag period, disease progression updates are continually written on physical paper charts and not immediately updated in the system, creating a continuous lag period and increasing the probability of errors and disjointed notes (personal communication, data consultant).

**Figure 1.** Flow of health care data in British Columbia Provincial Health Services Authority. CDI: Clinical Diagnostic Imaging; EMR: electronic medical record; HR (PHSA): Human Resources (Provincial Health Services Authority); MOH: Ministry of Health; PHIP: Public Health Improvement Plan; PHSA: Provincial Health Services Authorities; PLIS: Provincial Laboratory Information Solution.



The large blue box on the left depicts a hospital EMR and outlines the multiple functions for which it is responsible (Figure 1). At the basic level, EMRs record hospital admissions, staff, facility scheduling, and billing information. Further functionalities include patient histories, health care staff's patient notes, treatment information, and test results, among others. Some EMR systems can also provide more advanced information such as drug interaction databases, chronic disease management, and statistics on treatment outcomes. These advanced functions guide physicians to practice evidence-based medicine, an important facet of today's medical world.

In the BCPHSA, a hospital EMR will send the information it gathers to province-wide databases depicted by the light purple squares (Figure 1), as listed below:

1. PLIS: Provincial Laboratory Information Solution, a database for laboratory test results.
2. CDI: Clinical Diagnostic Imaging, a database for diagnostic imaging results and scans.
3. PHIP: Public Health Improvement Plan, a provincial body that analyzes health information and tries to improve health outcomes with conclusions from these data.
4. PharmaNet: the provincial drug information repository that also connects all pharmacies.



These databases are independent entities that maintain their systems. They are all legally required to send their data to the data warehouse (depicted in black in Figure 1), which is the central storage and integration facility for all health information in the BCPHSA. The data warehouse stores this information, formats it appropriately, and disseminates information to teams that create data marks, which are processed data sets that are ready for analysis. SAS (SAS Institute), an analytics software, is used to produce reports based on the data collected. Finally, the BCMOH requires certain data to be reported periodically, such as hospital admittances, discharge statistics, and so on.

External laboratories are groups that are not associated with the hospital system and are represented by light blue shapes. Some laboratories have their private websites, where both health care personnel and patients can log in to view their test results. Specialist clinics are also represented as external bodies separate from the hospital system or primary care physicians. For these, the flow of information comes in the form of referrals that are often faxed directly from the primary care physician to the specialist, whereas the patients themselves are commonly requested to also bring a physical copy of the referral to make an appointment with the specialist.

Primary care private practices encompass a large portion of the public's interaction with the health care system. These bodies are represented in green (Figure 1). This is the path through which patients' health care data go through when they visit their family doctor, walk-in clinics, etc. Information about patients is stored within the private clinic's own EMR and is not sent to any other facility. According to the industry expert consulted for this work, this is often the case as well for specialist clinics that are not attached to hospitals. Finally, patients and pharmacies are connected only through physical prescriptions. Patients may receive a printed or even handwritten prescription to bring to the pharmacist to collect their medications.

### Alternative Solutions for Health Data Management

Hospitals and clinics have several choices for managing electronic records of health care information, such as EMR software or cloud-based data management. Currently in development are blockchain-based EMR systems such as MedRec, but these systems are only in their testing phases in the United States and in some countries in Asia; they are neither available nor in testing in Canada at the time of publication. Cloud computing, EMR software, and MedRec are discussed in the following paragraphs.

Cloud computing is a well-adopted alternative that uses a network of remote servers to store and manage information. It differs from traditional EMR software in that the data are not tangibly stored within the facility [20]. In theory, cloud-based EMRs can be streamlined and centralized to enable interoperability. However, it is very challenging to do so if there is no additional software integration across all communication facilities, especially if they use different cloud-based EMR systems [20]. The primary benefit of cloud computing is that there is access to reliable network backups and there is a contingency plan for data loss to local hardware failures. Currently, most EMR systems used are cloud based given that these are less rigid and easier to update with new information

in comparison with a local host system [20]. According to the Capterra EMR Software Directory, the top 5 cloud-based systems by provider utilization are eClinicalWorks, Allscripts, CureMD, Epic, and Cerner [20]. Therefore, unlike host EMR software, interoperability could be an actual possibility through cloud-based EMR, but it would still be a major challenge [20].

There are hundreds of vendors offering EMR software that relies on an EMR host system. In this solution, all data are stored on one computer, designated as a server that houses a database file [20]. This file contains all the desired data. Every computer in a facility has a workstation installed that is linked to the server and is continuously sending and receiving data from the server. It is important to note with EMR software that although workstations require installation, data from the server can still be accessed outside the facility [20]. One of the main benefits of IT solutions that entail a hosted EMR system is that it allows a facility to take ownership of the data [20]. However, because of the wide selection and range of host EMRs that suit different environments, health care facilities often choose different software that are built very differently and therefore cannot communicate with each other. This makes interoperability and data sharing across points of care nearly impossible.

The blockchain-based EMR platform MedRec, an MIT project, is currently being tested and is thought to be a viable solution for health care data management. It is designed for a patient-centric health care model that focuses on managing authentication, confidentiality, and data sharing [28]. The system is completely decentralized and is based on access rights to a private blockchain. Patients and providers are given control over the access and retrieval of relevant medical records [28]. The system network operates on the internet, but there is no central repository of permissions or private data, which decreases the incentive to hack the system as such an attempt yields a low yield. The system works by establishing a relationship between the patients and the medical record originator on the blockchain. This then creates the foundation of a smart contract that other members of the network must use to request access to the medical record [28]. The medical records themselves are held by the patient in a wallet, accessible on their mobile devices. Smart contracts are executed automatically when the required conditions are met. These smart contracts remove the necessity to trust any one party with the storage and sharing of congregated personal health data.

It is important to note that for nonblockchain solutions, interoperability is the underlying constraint with existing alternatives for health data management. There are administrative and bureaucratic power dynamics at play in selecting EMR software or cloud-based EMR systems that perpetuate the misalignment of systems across entities (personal communication, data consultant). The sunk costs associated with upheaving employed systems and software are reasons for pushing back on the topic of standardization. Although it is not costly to update existing alternative systems and software, it is not cost-effective to replace current systems for a marginally similar alternative (personal communication, data consultant).

## Blockchain for Health Data Management

The current major stumbling block for health care data is interoperability, as health care facilities across Canada use a wide range of unique information systems [29]. However, if blockchain is used to create the overarching health care information system architecture, this issue could be significantly improved. Blockchain can retrieve and make use of any digitized information contained in each system via an application programming interface (API), which is a package of code that teaches the blockchain how to access the data in a different information system [29]. The API method can be used as a transition toward a fully blockchain-based EMR. In this manner, replacing the EMR or data system in each facility can be implemented over a longer period, as each system simply requires an API that connects it to the blockchain. Each facility can continue to use their system that serves their needs, and physicians' resistance to change, another major issue in health care data management, can be tackled over a much longer grace period of 3 to 4 years. The strain on health systems and the resource requirement needed for the proposed blockchain solution are far less when compared with simply switching EMRs abruptly. Abrupt switches often require a complete change in systems over 6 months to 1 year and often have negative effects on clinic workflows and efficiencies [29]. In addition, immediate costs will be greatly reduced, as installation and training will be more spread out.

Cryptocurrencies are often mentioned in the same breath as blockchain, which may lead to some confusion over why a cryptocurrency does not exist in this study's application of blockchain. First, it must be understood that blockchain is not the same thing as cryptocurrency [30]. Blockchain is the underlying technology that enables the creation of cryptocurrencies; thus, cryptocurrencies are simply an application of blockchain [30]. In the same vein, the patient-centric, verifiable, and secure health record management system this study suggests is simply another application of blockchain technology. In some applications of blockchain, cryptocurrency can be added as an incentive program [31-33]. For example, in a blockchain-based electric vehicle charging station scheduling program, users are rewarded with coins if they use the charging station during an off-peak period but receive no coins if they use it in a peak period [31]. Therefore, there is a potential to add cryptocurrency to the health care data blockchain if there is a need to incentivize some users to do certain things. For example, you could reward patients who renew their prescriptions on time or those who follow their vaccination schedules properly. However, this incentive program is not vital to the operation of the health care record blockchain; rather, it is more of a bonus than an integral component. As such, a cryptocurrency will not be included in first-generation health care blockchains but may be an option in the future as health care blockchains develop.

One of the biggest features of using a blockchain to store health care data is the possibility of patient-centric health care [34]. Patient-centric health care increases patients' understanding of their condition and allows the patient to play a more active role in their health care. This has been proven to increase concordance and adherence, improving health outcomes. [35].

Poor adherence to treatment plans costs Canada more than Can \$9 billion (US \$6.8 billion) per year [35]. In addition, because medical records are attached to the patient and not the provider, it makes seeking care in different provinces or even different countries much easier.

Blockchain also allows a user to have varying degrees of anonymity and privacy from each node in the network [34]. These nodes can represent health care facilities, government agencies, or even individual patients. A critical issue in health care is that different entities do not fully trust each other, which impedes data sharing between entities. This is linked to poorer health outcomes, as mentioned above [10-13]. The nature of blockchain allows for *trustless disintermediation*, which enables parties who do not trust each other to share certain digital information when protecting their private data [36]. In a private blockchain, each node is only allowed to join after their identity is verified, and the majority of the network approves it. For nodes in the network that own medical records, blockchain allows fine control of the specific information that they choose to share with other nodes in the network. If one wishes, relevant records could be shared without revealing any personally identifiable information (PII). This is because PII is never stored on the blockchain itself; it is always on *state channels*, which are interactions that are conducted off the blockchain [34]. In summary, each node will have its personal identity key, an alphanumeric code masking their identity, which is linked to all the medical records they own. This key can be used to retrieve information whenever they want. If patients want to share their medical records, they can conduct a state channel interaction with a health care professional or facility on a state channel [34]. This interaction allows them to transmit the particular records they choose to share with the health care provider, and afterward, a record of this transaction will be published onto the blockchain.

The decentralized manner in which information is stored on a blockchain discourages hacking efforts, as mentioned above when describing MedRec. Currently, personal data are stored in centralized databases. Hacking of this centralized system will leak every single file the database contains, along with all the personal information attached to each record [34]. If Canada decides to build a single, centralized database with the health records of all Canadians on it to enable interoperability, it becomes an immense target for hackers [37,38]. This is a bad idea as demonstrated by examples from around the world: in 2016, 15.5 million EMRs were breached in the United States and the global health care industry spent US \$6.2 billion in that year alone to deal with security breaches [37]. The value of a hacked EMR in the black market is high, estimated to be approximately 10 times the value of a credit card number, increasing the incentive to attack EMRs [39].

However, with a decentralized blockchain model, keys can only be hacked individually. With 1 hacked key, the hacker gains access to only a single person's files and information. There is no large database of information freely available to steal from [34]. If there were health records of 35 million people on the *Canadian Healthcare Data Blockchain*, a hacker would have to hack the blockchain 35 million times to gain access to the entire database. In addition, to be able to manipulate the

network, the hacker must attack 51% of all computers, which is currently infeasible without prohibitively large computer resources [3].

### Centralization Versus Decentralization in a Blockchain Context

To further explore whether centralization or decentralization is more beneficial for Canada's health data management, definitions are required. Current solutions for interoperability lead to a centralized solution. This means creating a centralized repository of all the data from different health care bodies and allowing certain entities access as required. This allows for a certain degree of interoperability, as multiple entities can access data in that central repository. A few of these central repositories already exist, as presented in Figure 1, such as PLIS, CDI, PHIP, and PharmaNet. However, this has not solved interoperability issues at the point-of-care level, which negatively affects patient care and therefore affects health outcomes, as discussed above [10-13]. Critical issues such as data security of a giant central repository [3,34], determining access rights for different clinics or institutions, and a general lack of trust between most health care organizations and clinics severely limit the centralization solution (these issues are further discussed in the Results section). In addition, ownership of the data also becomes an issue. Who will be the owner of all this information? How will the owner or controllers of the repository be chosen? What institution will be held responsible if there is a data leak or if data are lost? This lack of trust among different health care bodies hinders willingness to hand over patient data and willingness to allow others to access their patient data. It also makes it unlikely that organizations will be able to cooperate sufficiently to maintain a centralized repository together.

However, blockchain solves the issue of interoperability without introducing centralization. Blockchain allows each body to store their data in their respective *node*, and data are then accessed upon request from a connecting node. There is neither a centralized node that stores all the information nor a central body that owns or controls the whole network [4]. Therefore, the blockchain acts as a network that allows independent bodies to communicate when retaining their desired level of privacy and without requiring trust between the bodies. Limitations to this decentralized model include reduced control of hospitals or clinics over their patient data because this control is now handed over to the patients. However, this level of control over data can also be customized in a blockchain solution such that hospitals can still retain a level of control that is workable and efficient for their workflows. In addition, patients will have simpler, more efficient, and timely access to their data, thereby enabling better health outcomes because of larger data access for treatments.

## Discussion

### Lack of Flow of Information Back to the Primary Health Care System

The overview of health information flow in British Columbia illustrates that hospital EMRs only send information in one way. There is no direct, timely, and complete flow of information

back to the EMR in hospitals. Although the diagnostic imaging and laboratory test result databases work well by allowing different physicians with access to view their patients' test results, it is often only their past results and not their latest test results. These results could take days to weeks to become accessible in the database. Therefore, patients may still be required to repeat expensive, sometimes even painful, tests when treated by different physicians working in different facilities. In addition, additional information such as consultation notes, diagnoses, or discharge summaries are sent to these provincial databases for record keeping with no intention of sharing it with other hospitals or patients in a retrievable format. Indeed, the sending hospital must keep its hard copy of the notes, as the copy sent to the provincial database is not accessible by frontline health care staff. As a result, patients who are transferred between hospitals receive extremely inefficient care as their file must be physically transferred with them. These manual processes often result in man-made mistakes, misplacement, or missing pages, or even a swap of pages between patients and mistakes with a huge impact when it comes to providing care for sick patients [40]. This issue leads directly to the next issue, the lack of focus on improving frontline care.

### Lack of Focus on Direct Improvement of Patient Care

The entire British Columbia health care data system infrastructure was not built to make frontline patient care better immediately. It was built to collect information and analyze outcomes over time, to provide raw data for research, and to create new standards of care for future patients. Although this is an important goal, it does not help current patients receive better, timely, and evidence-based care. In short, there is no focus and no budget to help current patients.

Given the manner in which the current system is built, hospitals are encouraged to continue using physical notes; digitization of these notes only serves to fulfill the hospital's legal obligation to send information to government databases. As the complete set of information never comes back in quick-enough turnaround time to inform health care decisions for the same patient, health care professionals would rather keep a complete set of physical notes on hand at the hospital than spend time and resources to change the *status quo*. However, if there is a system that allows health care staff to enter patient notes, diagnoses, and collect test results in real time, with easy accessibility from different institutions when strictly guarding privacy and permissions, there would be a much larger incentive for physicians to change their habits. This would assist health care staff in conforming to a digital health data storage solution, which allows better patient data transfers and can build a more complete patient health profile over time.

This calls for a significant change in how health care data networks should be managed, as both current and future care are equally important and should be taken into consideration in a health care data storage and transmission network.

### Network Latency

Network latency refers to the delay between an order to transfer and the actual transmission of the data. This is a critical issue



in health care data networks as information is often needed immediately, and consequences can be dire if there are information mix-ups or delays. Due to the build of the current system, health care staff and personnel cannot transfer real-time, updated digital information to any health care professional who needs it. Doctors write physical notes, and these papers are given to coders who digitize the information month by month. Although any information entered by these coders is seen immediately in the provincial database and the data warehouse, no party will begin processing and analyzing these data until the month cutoff date is reached and the coders declare that the block of monthly information entered is complete and correct. Therefore, as a patient progresses in the hospital, their eHealth records can get very disjointed and information can easily be lost or incorrectly linked over time. This means that there is a significant, unbridgeable gap between finalized digital health information passed onto the government and the most current unprocessed information collected in the hospital. This leads to the conclusion that current EMR systems themselves, and the way they are used by health care professionals, are not optimized for frontline care and should be re-evaluated with the active patient as the focus rather than the focus on data collection.

### Information Silos

Many information silos are shown in Figure 1. Individual hospitals (depicted by dark blue) that do not use the same EMR as other hospitals are only able to send information to each other via fax or physical transfers of papers, and private clinics (depicted by green).

### Barriers to Health Data Digitization in British Columbia

The PITO reports that British Columbia has one of the highest rates of EMR adoption, where 80% of physicians in the province had an operating EMR system [27]. However, this statement does not reflect the optimization of EMR software within a clinic or a hospital. On consultation with the industry experts described above, it became evident that EMR systems in hospitals across the PHSA were using highly outdated methods for information sharing, although EMR systems were deployed and adopted [41]. This implies that systems integration is still severely lacking, and there is inadequate will and incentive to optimize use. Although there is a massive digitization backlog, an adequate integration system that allows interoperability between health care facilities has yet to be identified.

There are notable cost implications with effective integration given that for an EMR system to be useful from the onset, it must be introduced and integrated at every level of care [27]. To provide context, handwritten notes are transcribed to charts that are then transcribed to an EMR. After this point, when a patient is seen at a neighboring hospital within PHSA, sensitive patient information is transferred via hospital courier, irrespective of whether that information has been digitized or not. Therefore, redundancy and inefficiency are blatant, but the financial implications to eliminate paper notes are tremendous [19]. This is a huge barrier to effective digitization efforts.

Furthermore, there is a steep learning curve when adopting a new system entirely [29]. Optimized health data digitization takes time and patience; doctors and health care staff must be in agreement to take on the endeavor to reroute the current *status quo* [19]. However, getting all parties on board and launching such a massive training program takes time. There is a scale-up process that may take years to implement even within a single hospital. Furthermore, given the nature of the private industry in which EMR vendors operate, there are limited regulations and standards that companies comply with across the board [42]. This means that when different facilities take on different vendors as their IT solution proponents, standardization of how the system is used and how it communicates is undetermined until initial adoption [42]. The process by which vendors are selected is a serious factor with political influence that affects interoperability and data digitization.

### Limitations of Blockchain

The personal identity key setup discussed in the Results section begs an important question: what if a user loses their identity key? In traditional blockchain systems, once a user loses their key, their data become completely irretrievable. Essentially, their data are lost forever. This is not acceptable for medical records [15]. Therefore, any future blockchain health care data management system will need to have built-in key retrieval processes that are as secure as the blockchain itself. These protocols have already been invented and are actively in use in other industries, such as payment systems [43]. Possible solutions to handling key retrieval after death can follow current protocols: allowing next of kin or the person with power of attorney the ability to retrieve the key if it is lost or if the original owner is incapacitated.

Another issue that may come to mind is throughput. Health data are seemingly endless, and more data are being generated at a dizzying pace each day, which requires sophisticated software and enormous amounts of computing power to sort through. Low throughput has also been a popular criticism of early blockchains, such as Bitcoin and Ethereum [44-46]. However, poor throughput is a common phenomenon in early and badly constructed blockchains [44]. Many inexperienced coders will put certain processes and data on the chain that should not be on the chain, for example, the actual protocol and private user data. This is one of the major reasons why throughput is assumed to be a common problem (personal communication, blockchain consultant) in addition to old consensus fabrics that simply could not handle high volumes efficiently [44]. In addition, large public chains also have poor throughput, as the number of users often increases exponentially without sufficient server and engineering support. However, a well-constructed private chain has been demonstrated to be able to handle high volumes of transactions. Examples include the University Health Network, which launched a blockchain-based patient consent gateway in 2018, in partnership with eHealth Ontario and IBM. It enables patients to share specific digital medical records with trusted providers and monitor entities' access to different records [47-50]. VeChain has active solutions for incentive programs, supply chains, logistics solutions, copyright tracking, document management, and smart agriculture [51]. There have also been significant leaps in research in the past few years in the

scalability and efficiency of blockchains [44,45,52], which has paved the way for the possibility of a high-throughput health care blockchain.

## Proposed Implementation Plan of Blockchain Solution

### Viability of Blockchain

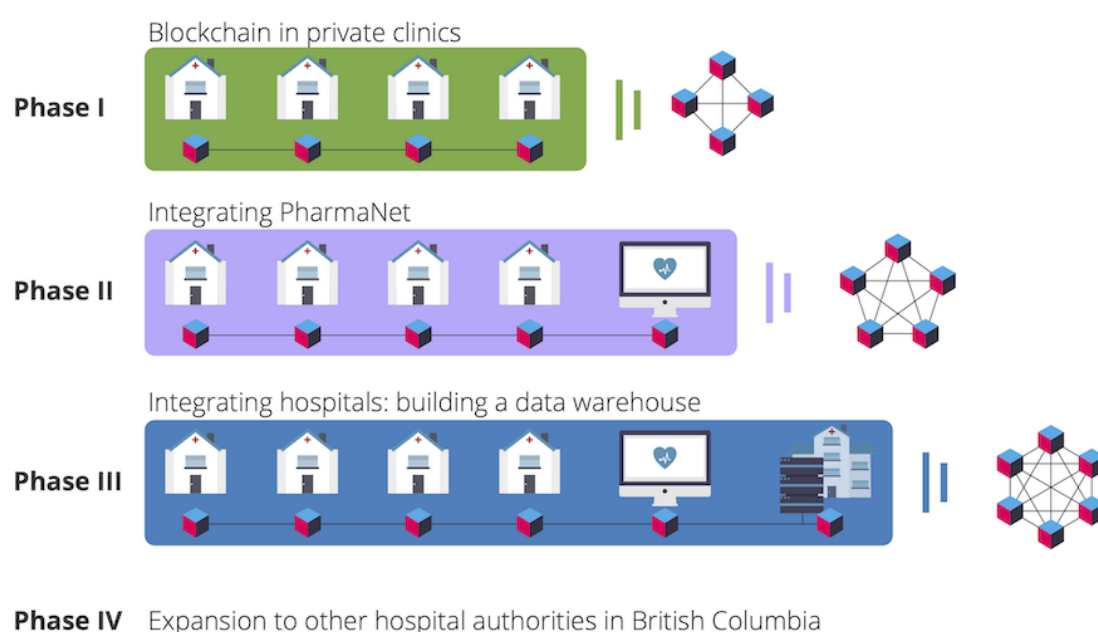
The success of a blockchain solution is directly underpinned by the rate of health record digitization. Issues encompassing a lack of flow within the health system, the lack of focus on the primary health system, and network latency are directly related to how quickly patient records are transcribed, coded, digitized, and made available in EMRs. The proposed solution, as outlined below, is built upon the existing cloud-based EMRs that are dependent on health data digitization. Therefore, although

blockchain resolves the issue of information silos, it does not bridge the gap between physical and electronic records. The viability of blockchain as a solution for interoperability is contingent on health data digitization.

### Phase 1: Blockchain in Private Clinics

First, a suitable blockchain EMR system needs to be built (Figure 2) that will connect all private clinics. The necessary components of a blockchain-based EMR are as follows: the use of ZKPs to execute contracts and transactions on the blockchain, self-sovereign identity framework, a decentralized cloud for data storage, and the permissioned blockchain itself for storage of proofs of interactions. These are described in the following paragraphs.

**Figure 2.** The 4 proposed phases of implementation of blockchain in British Columbia's health care data management.



ZKPs mean that as the blockchain is executing commands or transactions, it does so without knowing about anything that is contained within the files being stored, uploaded, or transferred between nodes [6-8]. In an oversimplified example of a ZKP, an imaginary young person goes to the bar. The imaginary bouncer needs to check their ID to ensure that they are of age before allowing entrance. In this scenario, instead of showing a piece of ID that tells the bouncer exactly how old this person is (and contains a lot of other sensitive information), a ZKP could present a simple, trusted statement to the bouncer that the person is old enough to enter. The main point is that the bouncer never needs to know how old the bar goer is, just that their age satisfies the requirement. This is a special feature of blockchains and it enhances security and privacy because there is minimum revelation and sharing of data between third parties [6-8]. However, attention must be paid to the type of ZKP used in the health care data blockchain. There are some types of ZKPs in use currently that are very inefficient and would not be viable in a health care setting, for example, the Zcash public chain, which uses a zero knowledge protocol called zkSNARK (zero

knowledge succinct noninteractive argument of knowledge) [53,54]. However, other types of ZKPs exist and are currently in use in different industries, such as in financial services, risk management, insurance, and supply chain management, demonstrating that there are ZKPs efficient enough to handle large-scale data requests [55,56]. Running times of specific ZKPs are published in academic papers, but the methods of creating efficient ZKPs are not published as they are considered proprietary techniques in nearly all leading blockchain companies. In addition, techniques involving combining hardware with the protocols, such as designing chips that can run certain cryptographic algorithms, are other methods of speeding up running times immensely. These techniques are already in use in many industries, and their viability is widely accepted among blockchain technology experts ([55,56]; personal communication, blockchain consultant).

A private self-sovereign identity framework means that each party in the network has been verified and has a way of confirming their identity each time they access the blockchain,



ensuring that other identities are not allowed to access or make any transactions in that network. This allows adherence to Canadian privacy laws regarding health data but also complies with the stricter General Data Protection Regulation from the European Union (EU), which applies whenever an EU citizen's information is involved. It also allows accurate, private, and efficient retrieval of data from storage.

Regarding cloud data storage, as each node stores its records, it creates decentralized storage of data. Although the nation or province's health care records will all be retrievable from any node given the correct permissions, there will be no large central database. This is critical as large databases often become an inviting target for hackers; decentralization discourages information breaches.

A permissioned blockchain is a feature that gives the EMR shared trust and immutability functions. A permissioned chain means that parties cannot freely join the network and become a node. They must obtain permission from the other nodes in the network before gaining access to any information shared on the network. Only proofs are stored on the chains, never personal data. However, it can facilitate transfers and keep a record of these transfers on the chain without ever knowing what was in the transferred files, with the help of ZKPs. As it must be a permissioned chain, the type of platform that can be used to build this chain is limited. Ethereum is a permissionless chain and therefore cannot be used; Corda is an option, but its peer-to-peer design will impose significant restrictions on the building of the platform. Hyperledger fabric would be an ideal platform to use, as it is a private chain and its channels are more flexible and allow for more options to be built in (personal communication, blockchain consultant).

In this phase, the nodes in the blockchain will be the doctors or health care staff of the individual clinics, insurers, and regulators.

### Phase 2: Integrating PharmaNet

Once the individual clinics have been linked in a blockchain EMR network, the next step is to integrate PharmaNet [23]. Individual clinics, especially family doctors, depend on pharmacists to fill prescriptions. Increased communication between pharmacies and clinics would allow monitoring over prescriptions to be filled legally and properly, which could reduce prescription drug fraud.

PharmaNet is an existing network [23]; therefore, further studies are needed to determine how this platform can feed information to the blockchain. Ideally, PharmaNet will transition into a fully blockchain-based network in phases, phasing out the use of their existing network. As this is not always possible, other solutions can include building a data warehouse to store and reformat all information collected from PharmaNet and injecting this into the blockchain via a single node for PharmaNet. However, if PharmaNet wants reciprocal information from clinics, they will need to access this information via the blockchain EMR system. As they use it for their own needs, that is, to retrieve the information they want, they will slowly get accustomed to the blockchain EMR interface, which will help smoothen the transition between PharmaNet and the

blockchain PharmaNet. More details on building a data warehouse are presented in the upcoming section on integrating hospitals, as hospitals are likely more resistant to switching EMR systems. This will require a phase-by-phase transition into a blockchain EMR, similar to PharmaNet's transition, but at a slower pace.

### Phase 3: Integrating Hospitals—Building a Data Warehouse

In order for a hospital to be added to the blockchain network where the blockchain acts as an external body, a separate data warehouse would need to be constructed to tether the network to the hospital. On the basis of the information flow presented in Figure 1, there are restrictions and privacy limitations on the flow of information from EMR systems to the existing PHSA data warehouse. The technical components of creating an API between existing hospital EMRs or the regional data warehouse that system feeds information into requires permission to be accessed on several levels. According to the industry expert consulted for this work, gaining access or permission to connect to existing infrastructure is challenging because of data governance. Given these limitations, when it becomes relevant for hospitals within the region to join the network for the preliminary purpose of accessing patient information from external clinics as opposed to sharing data externally, a separate database must be developed. This proposed database connects to the hospital EMR and enables an API connection to the blockchain node.

The process of building a data warehouse is a complex and sensitive endeavor itself without considering that it would be a stepping stone to the blockchain. Health Catalyst, a health care data analytics company, proposes a late-binding approach to data warehouse implementation as opposed to the enterprise data model or independent data model. A late-binding approach differs from a solely top-bottom or bottom-up method in that it is more pragmatic and equipped to handle the rapidly changing environments within health care [57]. This method uses features of both enterprise data warehouse and independent data mart, where data are taken from the source system in its most atomic form and transferred into course marts that exist within an enterprise data warehouse. Unlike independent data marts, data are not transformed as soon as they are transferred out of their source; rather, the data are kept in these raw marts in their rawest form [57]. Data are then transferred from the source marts to a data mart where data transformation and binding can occur. This is an incremental model that allows for data binding only when necessary [57]. It eliminates the disadvantages associated with the enterprise data model, which forces one to hammer out the system at the beginning before there is an understanding of what the data can be used for. This model can be thought of as a just-in-time data-binding approach. Decisions are made only about data transformation and binding as needed [57].

### Phase 4: Expansion to Other Hospital Authorities in British Columbia

Once the blockchain EMR is fully established and functional within all hospitals and the smaller health care facilities in PHSA, this blockchain can be pushed toward the other 6 hospital authorities in British Columbia. PHSA will act as a catalyst and demonstrate the possibility of a controlled, private, yet free flow

of health care information for those who require it and have permission to view it. Implementation in other hospital authorities will likely be able to follow the same general steps as implementation in PHSA, but, of course, each authority will have unique challenges and requirements. Therefore, a detailed and full-scale consultation with each stakeholder should be done before making changes to ensure that the transition is smooth and to safeguard the quality and availability of patient care.

### Conclusions

The exploration of the *status quo* in British Columbia's health care data management has exposed information silos even within small communities, expensive mistakes, and issues with implementation, and a general unwillingness of parties within a health care system to trust and share data. Other challenges include the potential for security breaches and operational issues in the current EMR infrastructure, although existing EMR systems follow legal security requirements stipulated by the government.

Compared with alternative technologies such as cloud-based solutions, IT solutions, and EMR systems, the blockchain-based solution has the highest potential for solving most of the common challenges in managing health care data. Blockchain offers (1) a plausible system to unite different groups that do

not trust each other, (2) decentralized storage to increase security, (3) sovereign identities to give patients and health care facilities secure access to medical records, and (4) most importantly, interoperability to allow the transfer of medical records to anybody who has permission and needs to access them. A blockchain-based solution will change the current infrastructure to a more universal and patient-centric system. The implementation plan will include targeting independent clinics in the PHSA of British Columbia first, then including PharmaNet, which will be followed by local hospitals. After establishing a strong and secure network to demonstrate the benefits that blockchain can bring to health care data management systems, the solution can be expanded to other hospital authorities in British Columbia and in the other provinces in Canada.

Further studies are required to address the limitations of using blockchain in health data management. More work is needed specifically in designing built-in key retrieval processes that are as secure as the blockchain itself as part of the health data management process. The future direction also includes exploring solutions to address the lack of health records digitization, which prevails and underpins health records management as a whole.

### Acknowledgments

The authors are grateful for the assistance from industry experts (who wish to remain anonymous) familiar with the health systems IT network of British Columbia's PHSA and blockchain technology. They also wish to thank Takeru Yasunaga for his time and expertise in designing the figures in this paper.

### Authors' Contributions

DC and TK worked on the manuscript and proposed the implementation plan for blockchain in British Columbia. OI, PV, and PM contributed to the conceptualization and design of the manuscript. All authors contributed to writing and revising the manuscript. All authors provided final approval of the manuscript and agree to be accountable for the manuscript.

### Conflicts of Interest

DC is a cofounder of Immunodex, a company that offers a blockchain-based, verifiable health record system that specializes in immunizations and self-declarations.

### References

1. Bublitz FM, Oetomo A, Sahu KS, Kuang A, Fadrique LX, Velmovitsky PE, et al. Disruptive technologies for environment and health research: an overview of artificial intelligence, blockchain, and internet of things. *Int J Environ Res Public Health* 2019 Oct 11;16(20) [FREE Full text] [doi: [10.3390/ijerph16203847](https://doi.org/10.3390/ijerph16203847)] [Medline: [31614632](https://pubmed.ncbi.nlm.nih.gov/31614632/)]
2. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing blockchains for efficient health care: systematic review. *J Med Internet Res* 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: [10.2196/12439](https://doi.org/10.2196/12439)] [Medline: [30747714](https://pubmed.ncbi.nlm.nih.gov/30747714/)]
3. Pineda D, Urban M. Inside The Black Blocks. Munk School of Global Affairs and Public Policy. URL: <https://munkschool.utoronto.ca/mowatcentre/inside-the-black-blocks/> [accessed 2020-10-01]
4. Funk E, Riddell J, Ankel F, Cabrera D. Blockchain technology: a data framework to improve validity, trust, and accountability of information exchange in health professions education. *Acad Med* 2018 Dec;93(12):1791-1794. [doi: [10.1097/ACM.0000000000002326](https://doi.org/10.1097/ACM.0000000000002326)] [Medline: [29901658](https://pubmed.ncbi.nlm.nih.gov/29901658/)]
5. dos Santos D, Baeßler B. Big data, artificial intelligence, and structured reporting. *Eur Radiol Exp* 2018 Dec 5;2(1):42 [FREE Full text] [doi: [10.1186/s41747-018-0071-4](https://doi.org/10.1186/s41747-018-0071-4)] [Medline: [30515717](https://pubmed.ncbi.nlm.nih.gov/30515717/)]
6. Goldwasser S, Micali S, Rackoff C. The knowledge complexity of interactive proof systems. *SIAM J Comput* 1989 Feb;18(1):186-208. [doi: [10.1137/0218012](https://doi.org/10.1137/0218012)]
7. Goldreich O, Krawczyk H. On the composition of zero-knowledge proof systems. *SIAM J Comput* 1996 Feb;25(1):169-192. [doi: [10.1137/s0097539791220688](https://doi.org/10.1137/s0097539791220688)]

8. Goldreich O, Oren Y. Definitions and properties of zero-knowledge proof systems. *J Cryptology* 1994 Dec;7(1):1-32. [doi: [10.1007/bf00195207](https://doi.org/10.1007/bf00195207)]
9. Archer N, Fevrier-Thomas U, Lokker C, McKibbin KA, Straus SE. Personal health records: a scoping review. *J Am Med Inform Assoc* 2011;18(4):515-522 [FREE Full text] [doi: [10.1136/amiainl-2011-000105](https://doi.org/10.1136/amiainl-2011-000105)] [Medline: [21672914](https://pubmed.ncbi.nlm.nih.gov/21672914/)]
10. Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *J Clin Bioinforma* 2015;5:4 [FREE Full text] [doi: [10.1186/s13336-015-0019-3](https://doi.org/10.1186/s13336-015-0019-3)] [Medline: [25834725](https://pubmed.ncbi.nlm.nih.gov/25834725/)]
11. Singh H, Giardina TD, Meyer AN, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med* 2013 Mar 25;173(6):418-425 [FREE Full text] [doi: [10.1001/jamainternmed.2013.2777](https://doi.org/10.1001/jamainternmed.2013.2777)] [Medline: [23440149](https://pubmed.ncbi.nlm.nih.gov/23440149/)]
12. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Arch Intern Med* 2003 Jun 23;163(12):1409-1416. [doi: [10.1001/archinte.163.12.1409](https://doi.org/10.1001/archinte.163.12.1409)] [Medline: [12824090](https://pubmed.ncbi.nlm.nih.gov/12824090/)]
13. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009 Nov 9;169(20):1881-1887. [doi: [10.1001/archinternmed.2009.333](https://doi.org/10.1001/archinternmed.2009.333)] [Medline: [19901140](https://pubmed.ncbi.nlm.nih.gov/19901140/)]
14. Berman M, Fenaughty A. Technology and managed care: patient benefits of telemedicine in a rural health care network. *Health Econ* 2005 Jun;14(6):559-573. [doi: [10.1002/hec.952](https://doi.org/10.1002/hec.952)] [Medline: [15497196](https://pubmed.ncbi.nlm.nih.gov/15497196/)]
15. McGhin T, Choo K, Liu C, He D. Blockchain in healthcare applications: research challenges and opportunities. *J Netw Comput Appl* 2019 Jun;135:62-75 [FREE Full text] [doi: [10.1016/j.jnca.2019.02.027](https://doi.org/10.1016/j.jnca.2019.02.027)]
16. Electronic Health Records in Canada—An Overview of Federal and Provincial Audit Reports. Office of Auditor General of Canada. URL: [https://www.oag-bvg.gc.ca/internet/English/parl\\_oag\\_201004\\_07\\_e\\_33720.html](https://www.oag-bvg.gc.ca/internet/English/parl_oag_201004_07_e_33720.html) [accessed 2020-06-01]
17. Draft Pan-Canadian Primary Healthcare Medical Record Content Standard, Version 2. Canadian Institute for Health Information. URL: [https://secure.cihi.ca/free\\_products/PHC\\_EMR\\_ContentStandards\\_E.pdf](https://secure.cihi.ca/free_products/PHC_EMR_ContentStandards_E.pdf) [accessed 2020-06-01]
18. Health Indicators. Canadian Institute for Health Information. 2013. URL: [https://secure.cihi.ca/free\\_products/HI2013\\_EN.pdf](https://secure.cihi.ca/free_products/HI2013_EN.pdf) [accessed 2020-06-01]
19. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. *BMC Health Serv Res* 2010 Aug 6;10:231 [FREE Full text] [doi: [10.1186/1472-6963-10-231](https://doi.org/10.1186/1472-6963-10-231)] [Medline: [20691097](https://pubmed.ncbi.nlm.nih.gov/20691097/)]
20. Ratushny V. The Benefits of Cloud-based EMR Systems. DermEngine: The Most Intelligent Dermatology Platform. 2018. URL: <https://www.dermengine.com/blog/dermatology-emr-software-cloud-interoperability> [accessed 2020-06-01]
21. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: Using Blockchain for Medical Data Access and Permission Management. In: 2nd International Conference on Open and Big Data. 2016 Presented at: OBD; 22-24 August, 2016; Vienna, Australia p. 25-30 URL: <https://ieeexplore.ieee.org/document/7573685> [doi: [10.1109/OBD.2016.11](https://doi.org/10.1109/OBD.2016.11)]
22. Ekblaw A. MedRec: Blockchain for Medical Data Access, Permission Management and Trend Analysis. MIT Libraries. 2017. URL: <https://dspace.mit.edu/handle/1721.1/109658> [accessed 2020-10-06]
23. What is eHealth? eHealth British Columbia. URL: <http://www.himconnect.ca/meet-him/faqs/what-is-ehealth> [accessed 2020-05-31]
24. Intrahealth. URL: <https://www.intrahealth.com/> [accessed 2020-06-01]
25. Med Access EMR. Telus Health. 2019. URL: <https://www.telushealth.co/products/med-access-emr/> [accessed 2020-05-31]
26. Telus Health. 2019. URL: <https://www.telushealth.co/health-solutions/electronic-medical-records/electronic-medical-records-emr/> [accessed 2020-05-31]
27. Hobson B. EMR use in BC: the future is now (part 2). *BC Med J* 2012 Dec;54(10):468 [FREE Full text]
28. MedRec. URL: <https://medrec.media.mit.edu/technical/> [accessed 2020-05-31]
29. Chang F, Gupta N. Progress in electronic medical record adoption in Canada. *Can Fam Physician* 2015 Dec;61(12):1076-1084. [Medline: [27035020](https://pubmed.ncbi.nlm.nih.gov/27035020/)]
30. Tasatanattakool P, Techapanupreeda C. Blockchain: Challenges and Applications. *IEEE Xplore (Institute of Electrical and Electronics Engineers)*. 2018 Jan 10. URL: <https://doi.org/10.1109/ICOIN.2018.8343163> [accessed 2020-05-31]
31. Zhang T, Pota H, Chu C, Gadh R. Real-time renewable energy incentive system for electric vehicles using prioritization and cryptocurrency. *Appl Energy* 2018 Sep;226:582-594 [FREE Full text] [doi: [10.1016/j.apenergy.2018.06.025](https://doi.org/10.1016/j.apenergy.2018.06.025)]
32. Chatzopoulos D, Ahmadi M, Kosta S, Hui P. Floppcoin: a cryptocurrency for computation offloading. *IEEE Trans Mobile Comput* 2018 May 1;17(5):1062-1075 [FREE Full text] [doi: [10.1109/tmc.2017.2748133](https://doi.org/10.1109/tmc.2017.2748133)]
33. Wang J, Li M, He Y, Li H, Xiao K, Wang C. A blockchain based privacy-preserving incentive mechanism in crowdsensing applications. *IEEE Access* 2018;6:17545-17556 [FREE Full text] [doi: [10.1109/access.2018.2805837](https://doi.org/10.1109/access.2018.2805837)]
34. Esposito C, De Santis A, Tortora G, Chang H, Choo KR. Blockchain: a panacea for healthcare cloud-based data security and privacy? *IEEE Cloud Comput* 2018 Jan;5(1):31-37. [doi: [10.1109/MCC.2018.011791712](https://doi.org/10.1109/MCC.2018.011791712)]
35. Wahl C, Gregoire J, Teo K, Beaulieu M, Labelle S, Leduc B, et al. Concordance, compliance and adherence in healthcare: closing gaps and improving outcomes. *Healthc Q* 2005;8(1):65-70. [doi: [10.12927/hcq.16941](https://doi.org/10.12927/hcq.16941)] [Medline: [15715337](https://pubmed.ncbi.nlm.nih.gov/15715337/)]

36. Zhang P, White J, Schmidt D, Lenz G, Rosenbloom S. FHIRchain: applying blockchain to securely and scalably share clinical data. *Comput Struct Biotechnol J* 2018;16:267-278 [FREE Full text] [doi: [10.1016/j.csbj.2018.07.004](https://doi.org/10.1016/j.csbj.2018.07.004)] [Medline: [30108685](https://pubmed.ncbi.nlm.nih.gov/30108685/)]
37. McMillan E. Hundreds of People to Be Notified After Patient Records Breached. CBC. 2017 Jul 17. URL: <https://www.cbc.ca/news/canada/nova-scotia/health-authority-patient-records-privacy-breach-1.4208723> [accessed 2020-05-31]
38. Rose B. Pharmanet Breach Compromises Personal Information of 7,500 B. CBC. 2017. URL: <https://www.cbc.ca/news/canada/british-columbia/bc-pharmanet-breach-1.3985173> [accessed 2020-05-31]
39. Line Blockchain Designed for Everyone. BlockChain. 2020. URL: [https://blockchain-org.line-apps.com/wp-content/uploads/2020/08/LINE%20Blockchain%20Whitepaper%20v2.1\\_EN.pdf](https://blockchain-org.line-apps.com/wp-content/uploads/2020/08/LINE%20Blockchain%20Whitepaper%20v2.1_EN.pdf) [accessed 2020-10-02]
40. Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing paper-based with electronic patient records: lessons learned during a study on diagnosis and procedure codes. *J Am Med Inform Assoc* 2003;10(5):470-477 [FREE Full text] [doi: [10.1197/jamia.M1290](https://doi.org/10.1197/jamia.M1290)] [Medline: [12807808](https://pubmed.ncbi.nlm.nih.gov/12807808/)]
41. Richardson D. Mistaken. *BC Med J* 2019;61(4):154 [FREE Full text]
42. Affleck E. Interoperability of electronic medical records requires more than just technical understanding. *Can Med Assoc J* 2019 May 13;191(19):E541. [doi: [10.1503/cmaj.71818](https://doi.org/10.1503/cmaj.71818)] [Medline: [31085567](https://pubmed.ncbi.nlm.nih.gov/31085567/)]
43. Milkau U, Bott J. Digitalisation in payments: from interoperability to centralised models? *J Payments Strat* 2015;9(3):321-340 [FREE Full text]
44. Vukolić M. In: Camenisch J, Kesdoğan D, editors. *The Quest for Scalable Blockchain Fabric: Proof-of-Work vs BFT Replication*. Cham, UK: Springer; May 1, 2016:112-125.
45. Zamani M, Movahedi M, Raykova M. Rapidchain: Scaling Blockchain via Full Sharding. In: *ACM SIGSAC Conference on Computer and Communications Security*. 2018 Jan Presented at: CCS'18; January 15, 2018; Toronto, Canada p. 931-948 URL: <https://doi.org/10.1145/3243734.3243853> [doi: [10.1145/3243734.3243853](https://doi.org/10.1145/3243734.3243853)]
46. Zhou Q, Huang H, Zheng Z, Bian J. Solutions to scalability of blockchain: a survey. *IEEE Access* 2020;8:16440-16455. [doi: [10.1109/access.2020.2967218](https://doi.org/10.1109/access.2020.2967218)]
47. Smith K. Clinician Engagement, Local Impact Awards, Budget Risk Meetings and Beyond. Toronto, Canada: University Health Network; 2019.
48. Carter D. Accessing Patient Health Records via Blockchain. Blockchain Research Institute. URL: <https://www.blockchainresearchinstitute.org/project/accessing-patient-health-records-via-blockchain/> [accessed 2020-05-31]
49. Rubin J. Toronto Hospital Trial Allows Patients to See Their Medical Records - and Control Who Else Sees Them Too. Thestar. URL: <https://www.thestar.com/business/2020/03/04/toronto-hospital-trial-looks-at-allowing-patients-to-see-all-their-records-and-control-who-else-sees-them-too.html> [accessed 2020-05-31]
50. Wiljer D, Brudnicki S. Bringing Blockchain to Healthcare for a New View on Data. IBM. 2019 Aug 19. URL: <https://www.ibm.com/blogs/think/2019/08/bringing-blockchain-to-healthcare-for-a-new-view-on-data/> [accessed 2020-06-01]
51. VeChain. VeChain Solutions. URL: <https://www.vechain.com/#solutions> [accessed 2020-05-31]
52. Dong M, Liang Q, Li X, Liu J. Celer network: bring internet scale to every blockchain. *ArXiv* 2018 Jun 15:1-68 [FREE Full text]
53. Wahby R, Tzialla I, Shelat A, Thaler J, Walfish M. Doubly-Efficient zkSNARKs Without Trusted Setup. In: *IEEE Symposium on Security and Privacy*. 2018 Presented at: SP'18; May 20-24, 2018; San Francisco, CA. [doi: [10.1109/sp.2018.00060](https://doi.org/10.1109/sp.2018.00060)]
54. Yuen T, Sun S, Liu J, Au M, Esgin M, Zhang Q, et al. RingCT 3.0 For Blockchain Confidential Transaction: Shorter Size and Stronger Security. In: *International Conference on Financial Cryptography and Data Security*. 2020 Feb Presented at: FCDS'20; February 14, 2020; Kota Kinabalu, Malaysia p. 464-483 URL: [https://doi.org/10.1007/978-3-030-51280-4\\_25](https://doi.org/10.1007/978-3-030-51280-4_25) [doi: [10.1007/978-3-030-51280-4\\_25](https://doi.org/10.1007/978-3-030-51280-4_25)]
55. Meng W, Wang J, Wang X, Liu J, Yu Z, Li J, et al. Position Paper on Blockchain Technology: Smart Contract and Applications. In: *International Conference on Network and System Security*. 2018 Presented at: NSS'18; August 27, 2018; Hong Kong, China URL: [https://doi.org/10.1007/978-3-030-02744-5\\_35](https://doi.org/10.1007/978-3-030-02744-5_35) [doi: [10.1007/978-3-030-02744-5\\_35](https://doi.org/10.1007/978-3-030-02744-5_35)]
56. Engberg S, Harning M, Jensen C. Second Annual Conference on Privacy, Security and Trust. 2004 Jan Presented at: PST'04; October 13-15, 2004; Fredericton p. 89-101.
57. Barlow S. What is the Best Healthcare Data Warehouse Model? Comparing Enterprise Data Models, Independent Data Marts, and Late-binding Solutions. *Health Catalyst*. 2017. URL: <https://www.healthcatalyst.com/> [accessed 2020-06-01]

## Abbreviations

**API:** application programming interface  
**BCMOH:** Ministry of Health in British Columbia  
**BCPHSA:** British Columbia's Provincial Health Services Authorities  
**CDI:** Clinical Diagnostic Imaging  
**CIHI:** Canadian Institute for Health Information  
**EMR:** electronic medical record



**IAR:** integrated assessment record  
**IT:** information technology  
**MIT:** Massachusetts Institute of Technology  
**PCSA:** point-of-care service application  
**PHIP:** Public Health Improvement Plan  
**PHSA:** Provincial Health Services Authority  
**PII:** personally identifiable information  
**PITO:** Physician Information Technology Office  
**PLIS:** Provincial Laboratory Information Solution  
**ZKP:** zero knowledge proof

*Edited by G Eysenbach; submitted 01.06.20; peer-reviewed by D He, J Liu; comments to author 22.06.20; revised version received 15.08.20; accepted 18.08.20; published 23.10.20.*

*Please cite as:*

Cadoret D, Kailas T, Velmovitsky P, Morita P, Igboeli O  
*Proposed Implementation of Blockchain in British Columbia's Health Care Data Management*  
*J Med Internet Res* 2020;22(10):e20897  
URL: <http://www.jmir.org/2020/10/e20897/>  
doi: [10.2196/20897](https://doi.org/10.2196/20897)  
PMID: [33095183](https://pubmed.ncbi.nlm.nih.gov/33095183/)

©Danielle Cadoret, Tamara Kailas, Pedro Velmovitsky, Plinio Morita, Okechukwu Igboeli. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Notes From the Field

# Selecting Mobile Health Technologies for Electronic Health Record Integration: Case Study

Ryan Shaw<sup>1,2</sup>, PhD, RN; Marissa Stroo<sup>1</sup>, MMCI; Christopher Fiander<sup>3</sup>, MMCI; Katlyn McMillan<sup>1,3</sup>, MPH

<sup>1</sup>Mobile App Gateway, Clinical & Translational Science Institute, Duke University, Durham, NC, United States

<sup>2</sup>School of Nursing, Duke University, Durham, NC, United States

<sup>3</sup>Duke Health Technology Solutions, Duke University Health System, Durham, NC, United States

**Corresponding Author:**

Ryan Shaw, PhD, RN

Mobile App Gateway

Clinical & Translational Science Institute

Duke University

307 Trent Dr.

DUMC 3322

Durham, NC, 27710

United States

Phone: 1 9196849434

Email: [ryan.shaw@duke.edu](mailto:ryan.shaw@duke.edu)

## Abstract

Mobile health (mHealth) technologies, such as wearable devices and sensors that can be placed in the home, allow for the capture of physiologic, behavioral, and environmental data from patients between clinic visits. The inclusion of these data in the medical record may benefit patients and providers. Most health systems now have electronic health records (EHRs), and the ability to pull and send data to and from mobile devices via smartphones and other methods is increasing; however, many challenges exist in the evaluation and selection of devices to integrate to meet the needs of diverse patients with a range of clinical needs. We present a case report that describes a method that our health system uses, guided by a telehealth model to evaluate the selection of devices for EHR integration.

(*J Med Internet Res* 2020;22(10):e23314) doi:[10.2196/23314](https://doi.org/10.2196/23314)

**KEYWORDS**

mobile health; mHealth; electronic health record; health technology; mobile phone

## Introduction

Mobile health (mHealth) technologies, such as wearable devices and sensors that can be placed in the home, allow for the capture of physiologic, behavioral, and environmental data from patients between clinic visits. This patient-generated health data (PGHD) can help reveal underlying mechanisms of health by filling in gaps in the information, providing insights into the day-to-day activities of an individual, and allowing for better strategies to prevent and manage acute and chronic illnesses. Moreover, with the proliferation of smartphones rising to over 81% of the US population [1] and over 73% of households gaining in-home broadband internet [2], the ability to collect these data from diverse socioeconomic and geographic populations is growing. According to a 2018 survey conducted by Accenture, 75% of US consumers felt that technology was an important part of managing their health [3]. Rapid growth in the global digital

health market, estimated to be over US \$423 billion by 2024 [4], supports that sentiment. Because mHealth technologies tether to smartphones and Wi-Fi or have cellular-embedded chips, health data can be collected in near real time from patients in their daily environments.

In the United States, over 96% of all nonfederal acute care hospitals now possess a certified electronic health record (EHR) system [5], and over 9 in 10 office-based physician offices have adopted an EHR system [6]. As health care facilities move beyond EHR implementation, the integration of data from connected devices, including mHealth technologies, is gaining speed. Companies such as Apple Inc, for example, have enabled the ability for patients to aggregate their health records on an iPhone from multiple hospitals via authentication by health system patient portals, such as Epic's MyChart [7]. It is also possible to integrate third-party data, such as patient-generated blood glucose levels, for example, into the EHR system via

Apple HealthKit [8]. This capability is possible with many of the major EHR vendors, including Epic, Cerner, and Athena Health, among others. Furthermore, this capability is expanding to Android platforms as well with the use of Google Fit.

While these technologies afford much promise, many challenges exist for health systems and others in the selection of devices to integrate and recommend for equitable patient care. The Office of the National Coordinator for Health Information Technology published a white paper in 2018 highlighting some of the challenges of collecting and using PGHD [9]. These include the technical challenges related to accuracy of measurements, data provenance, and privacy and security concerns. They also explored the patients' challenges and opportunities, which included the lack of internet or smartphone access as well as health and technology literacy deficits. A 2018 review by Reading and Merrill examined the needs of patients and providers around the use of PGHD in health care. Their review highlighted common needs for technology, including data quality, electronic integration, simple-to-understand actionable insights, and security.

The challenges and opportunities for PGHD are clear, but the path to moving forward remains undefined. One large obstacle is selecting the right devices from the ever-increasing number of consumer digital health devices on the market. Technology selection depends on the data of interest and the technology the patient, clinician, and health care system have ready access to and can use for clinical decision making or population health management. In this use case, we describe a method that Duke University Health System uses to evaluate and select devices for EHR integration.

## Methods

Our team of researchers, clinicians, and informatics technology professionals met to identify factors involved with the selection of devices to integrate with the EHR system. These factors evolve based on feedback from stakeholders and the

ever-growing digital health market. Key considerations included clinical validity of devices, patient satisfaction, and usability of both the connected device and the app interface associated with each smart, remote, monitoring device.

We use the Model for ASsessment of Telemedicine applications (MAST) [10] as a guide for device selection. This validated model is used by decision makers to aid in the choosing of the most appropriate telehealth technologies. We have modified the model to reflect variables needed in the selection of mHealth technologies for EHR integration. This includes, for example, details on US Food and Drug Administration (FDA) medical class and technological aspects, such as Bluetooth or Wi-Fi connection. The model includes three steps: Step 1: Preceding Considerations, Step 2: Multidisciplinary Assessment, and Step 3: Assessment of Transferability (see Figure 1) [10].

Step 1 involves determining the purpose of the connected device and relevant alternatives. The goal is to understand the primary outcomes and whether the device involves an upgraded or new technology. Next, several conditions are considered, including the following: legislation (ie, regulations, accreditations, and liability), reimbursement (ie, insurance vs hospital paid), maturity (ie, development time and resources needed over time to support the tool and how safe the tool is), and the number of patients involved to inform an economic analysis. Step 2 then involves a multidisciplinary assessment across eight domains. We added an eighth domain on technological aspects to reflect specific aspects of connected devices. The domains include the following: (1) health problem and description of the application, (2) safety, (3) clinical effectiveness, (4) patient perspectives, (5) economic aspects, (6) organizational aspects, (7) health equity [11], sociocultural, ethical, and legal aspects, and (8) technological aspects. Finally, in Step 3 an assessment is made as to the transferability of connected devices including interoperability (ie, Fast Healthcare Interoperability Resources) and the number of patients who will use the tool to determine costs per patient.

**Figure 1.** Process for evaluation of connected devices.**Step 1. Preceding Considerations**

- Aim of the connected device
- Relevant alternatives
- Legislation
- Reimbursement
- Maturity of the application
- Number of patients

**Step 2. Multidisciplinary Assessment**

- (1) Health problem and description of the application
  - Assess literature for evidence of clinical need
  - Describe the proposed tool and how it is used
  - Describe data flow
- (2) Safety
  - Assess US Food and Drug Administration medical class: Class I (47%). These devices present minimal potential for harm to the user; 95% are exempt Class II (43%). Higher risk than Class I; must undergo special controls Class III (10%). Usually sustain or support life, are implanted, or present risk of illness or injury
  - Assess published, grey literature for safety concerns
- (3) Clinical Effectiveness
  - Assess literature for evidence of clinical need
  - Assess evidence for device validity and accuracy
- (4) Patient perspectives
  - Assess literature for device recommendations based upon consumer preference
  - Usability testing with targeted patient groups
  - Accessibility for targeted population
- (5) Economic Aspects
  - Determine cost of product
  - Assess for costs to patient (insurance coverage vs out of pocket)
  - Assess for costs to organization to purchase
- (6) Organizational aspects
  - Assess security and privacy features and requirements
  - Assess for development and maintenance support needs and costs
  - Assess for other relevant metrics such as technical support needs
- (7) Health equity, sociocultural, ethical, and legal aspects
  - Consequences of implementation or not
  - Technology access needs (i.e., broadband, computer, tablet, and smartphone)
  - Socioeconomic implications and affordability
  - Digital health literacy
  - Inclusive design
  - Legal obligations and legal barriers that may exist
- (8) Technological aspects
 

Describe technical elements such as use with iOS, Apple HealthKit, Android, Google Fit, Bluetooth, Wi-Fi, cellular, third-party websites, and application programming interface (API) integration

**Step 3. Assessment of Transferability**

- Interoperability: Assess for integration into electronic health record infrastructure (ie, patient portal, Fast Healthcare Interoperability Resources, APIs, and others)
- Number of patients and economies of scale

## Results

Figure 1 illustrates the three-step process for evaluating and selecting connected devices based upon a modified version of the MAST model. As we evaluate devices, our team maintains an internal working document of a table of devices. This document is refined and expanded as we make decisions and approach device integration. In order to create Figure 1, we began by going through the process of selecting glucometers

to recommend be integrated into our Epic-based EHR system (see Figure 2). This exercise allowed us to refine the process and add variables to Figure 1. For example, because the evidence for many devices is limited we expanded to grey literature, including Consumer Reports and Amazon reviews, to gain perspective on patient usability and utility. Other examples include discovering the need to list technical requirements, such as Apple or Android capabilities, connection to Apple Health and Google Fit, and how data are collected and transmitted (ie,

Bluetooth, Wi-Fi, and cellular). Of note, this case report focuses on evaluating device selection. Future work will evaluate clinical and institutional outcomes, as these tools are used as part of patient care delivery and research endeavors.

**Figure 2.** Example evaluation of noncontinuous glucometers: CONTOUR NEXT ONE.

### Step 1. Preceding Considerations

Glucose monitoring is important for patients to test their blood sugar level and understand how foods, medications, and activities affect their diabetes. Keeping track of glucose helps patients and clinicians manage the condition. Glucose monitoring is well established, reimbursable, and US Food and Drug Administration (FDA) regulated. Thousands of patients use it in our health system, thus enabling economy of scale.

### Step 2. Multidisciplinary Assessment

- (1) **Health problem and description of the application**
  - Glucose: 2019 American Diabetes Association recommendations
  - Insulin-dependent diabetes mellitus: evidence clearly supports self-monitoring of blood glucose (SMBG) (lower hemoglobin A1C and rates of daytime and nocturnal hypoglycemia)
  - Noninsulin-dependent diabetes mellitus: key consideration is that performing SMBG alone does not lower blood glucose
  - To be useful, information must be integrated into clinical and self-management plans
  - Patients use the glucometer to make decisions on diet, exercise, and medications
  - Data from the glucometer is transferred via Bluetooth to an associated app; these data then go into a patient portal or are pulled into the electronic health record (EHR) via an application programming interface (API)
- (2) **Safety**
  - Assess if glucometers require FDA approvals (yes, Class II)
  - Glucometers must be accurate or poor medication decision making can occur
- (3) **Clinical Effectiveness**
  - In the United States, currently marketed monitors must meet the standard under which they were approved, which may not be the current standard
  - Moreover, the monitoring of current accuracy is left to the manufacturer and not routinely checked by an independent source
  - Klonoff DC, Parkes JL, Kovatchev BP, et al. Investigation of the accuracy of 18 marketed blood glucose monitors. *Diabetes Care*. 2018;41(8):1681-1688
- (4) **Patient perspectives**
  - Assess consumer literature for recommendations. CONTOUR NEXT ONE is the highest recommended glucometer by consumer reports due to ease of use of device and associated app design
- (5) **Economic Aspects**
  - US \$9.99 for the meter plus, on average, US \$20 for 50 test strips
  - Private insurance coverage may vary
  - Medicare Part B coverage with prescription
- (6) **Organizational aspects**
  - Patients agree to the privacy policy from Ascensia when downloading and using the app
  - Integration of glucometer data into EHR flow sheet requires development time, testing, and ongoing maintenance to support software updates
  - Technical support: patients may call the health system or a clinician for technical assistance
- (7) **Health equity, sociocultural, ethical, and legal aspects**
  - CONTOUR DIABETES App compatible with iOS and Android
  - Broadband not required, though a smartphone is needed
  - Relatively affordable with private and government insurance coverage
  - Would be useable across diverse populations
  - Must create a process and consent language for patients that data will not be reviewed by the health system in real time; standard of care is practiced
- (8) **Technological aspects**
  - iOS, Android, Apple Health, Google Fit, and connects via Bluetooth
  - No current health system technical support
  - Not EHR enabled yet

### Step 3. Assessment of Transferability

- Epic-based EHR patient portal can query Apple Health for glucose data via Fast Healthcare Interoperability Resources
- MyChart can access Glooko via public API
- Scalable to thousands of patients at our health system

We selected noncontinuous glucometers as our case study (see [Figure 2](#)) because of requests from clinician groups to retrieve glucometer data from patients and our experience integrating glucose data into our EHR system via Apple HealthKit [8]. As presented in [Figure 2](#), the exercise revealed that glucose is a data point of value for clinical care. Further, glucometers are

considered FDA Class II medical devices and must demonstrate substantial equivalence to a predicate device. A review article by Klonoff et al investigated the accuracy of 18 marketed blood glucose monitors [12]. We searched the consumer-facing literature, such as Consumer Reports, to compare recommendations. The next steps in the evaluation process



involved documenting the ability for each glucometer to be used on iOS and/or Android devices, integration with Apple HealthKit and Google Fit, costs, additional technical features, current integration with our EHR infrastructure and how data are retrieved, and if technical support is available. Results showed congruence across these measures and the CONTOUR NEXT ONE glucometer came out as the top contender.

## Discussion

### Principal Findings

The proliferation of wireless and mobile technologies provides opportunities to connect information in real-world settings via wearable sensors and, when coupled with fixed sensors embedded in the environment, to produce continuous streams of data on an individual's biology, psychology, behavior, and daily environment. These collected data have the potential to be analyzed and used in real time to prompt individuals to change behavior or their environmental exposures that can reduce health risks or to optimize health outcomes.

Selecting devices for integration requires many factors to be evaluated. These factors are technical, clinical, organizational, economic, and patient focused. Popular and currently well-known devices, such as the Fitbit and Apple Watch, are easier to identify due to their accessibility and widespread adoption. Evidence needed for activity trackers like these to be on the consumer market is also less stringent compared to evidence needed for a portable electrocardiogram or glucometer, which require FDA clearance. Devices that require FDA clearance provide an additional layer of evidence for safety and utility. This is in contrast to devices that do not require FDA clearance, such as sleep monitors.

Continued discussion with clinical and operational leaders suggests how broad the idea of technical support could be. Technical support can include such steps as configuring the device for the patient, providing support in person or remotely, and having staff available for ongoing troubleshooting. Other levels of technical support include patient support for managing the clinical data landing in the EHR system, with or without notification, along with addressing support related to the notifications specifically. Lastly, technical support should implicitly include the presentation of the data to providers so that they are actionable and accessible. Actionable and accessible data are essential for the provider or care manager to be able to intervene, while also not exacerbating provider burnout, which is more frequently reported since the large-scale implementation of EHRs. While these concepts are fundamental, they are also often frequent attributes referenced as potential barriers to inclusion by clinical and operational leaders.

A variety of devices should be selected for integration so that access to, and accessibility of, these tools is more equitable

across patient populations. Patients have both iOS and Android devices and choosing one platform to focus on limits patient accessibility. Further, for devices that connect via a web portal or consume significant data through video, for example, patients may or may not have in-home broadband internet and are limited to internet access via their phone. This could be a limiting factor for patients based upon their geographic location or socioeconomic status. This is also important because the literature shows that devices are not always designed to be accurate across diverse populations. It was reported that the light sensor in some wearable devices was not usable on patients with darker skin tones due to the color of the optical sensor selected. While this has been addressed by many device manufacturers [13], it is a lesson in the importance of ensuring devices are usable across a variety of patient populations.

Future evaluation will also expand to include software platforms, such as those from Livongo, for example, which incorporate a variety of devices and provide personalized guidance to patients and clinicians with chronic illnesses. A third scenario is also necessary to consider regarding integration: there are applications and devices that offer their own portal for viewing data, but do not offer compatibility to iOS or Android, nor can these devices integrate into an aggregator inherently. Given this scenario, it becomes necessary to evaluate a device's or platform's capabilities through application programming interfaces (APIs) so that data can be aggregated productively and used in a clinical environment.

While mHealth technologies, specifically connected devices, hold promise to benefit patient care delivery and patient self-management, many challenges exist with their integration into health care. There is limited regulation, and rigorous scientific evaluation of many devices is lacking. There are many devices on the market, and every device must be tested for data quality, interoperability, and usefulness by patients and clinicians. Further, the rapid evolution of the connected device market requires frequent re-evaluation and system software updates. Finally, use of these tools in formal care delivery models is relatively new and, thus, understanding how to support patients and how to integrate and present the wealth of data from devices into actionable insights for clinical decision making continues to advance.

### Conclusions

We present an example on how we recommend which mHealth devices should be integrated into a health system's EHR system to collect PGHD. Many factors are involved, and it is important to conduct a thorough assessment to assess for clinical requirements, technical features, and patient-level factors such as usability and costs. Figures 1 and 2 can be used as templates for others to expand upon.

### Acknowledgments

Support was provided, in part, by the Duke Clinical and Translational Science Institute, which is supported, in part, by a US National Institutes of Health Clinical and Translational Science Award (grant number UL1TR002553).



## Authors' Contributions

RS led the team and was responsible for all aspects of the project. MS, CF, and KM substantially contributed to the methods, data acquisition, results, and interpretation, and also participated in all aspects of writing the manuscript.

## Conflicts of Interest

None declared.

## References

1. Mobile fact sheet. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2019 Jun 12. URL: <https://www.pewresearch.org/internet/fact-sheet/mobile/> [accessed 2020-10-13]
2. Internet/broadband fact sheet. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2019 Jun 12. URL: <https://www.pewinternet.org/fact-sheet/internet-broadband/> [accessed 2020-10-13]
3. Safavi K, Kalis B. 2018 Consumer Survey on Digital Health: US Results. Dublin, Ireland: Accenture; 2018. URL: [https://www.accenture.com/t20180306t103559z\\_w\\_us-en/acnmedia/pdf-71/accenture-health-2018-consumer-survey-digital-health.pdf](https://www.accenture.com/t20180306t103559z_w_us-en/acnmedia/pdf-71/accenture-health-2018-consumer-survey-digital-health.pdf) [accessed 2020-10-13]
4. Digital Health Market by Product Type, by Component and by End User: Global Industry Perspective, Comprehensive Analysis and Forecast, 2017 - 2024. Maharashtra, India: Zion Market Research; 2018 Jul 12. URL: <https://www.zionmarketresearch.com/report/digital-health-market> [accessed 2020-10-13]
5. Percent of hospitals, by type, that possess certified health IT: Health IT quick-stat #52. Office of the National Coordinator for Health Information Technology. 2018 Sep. URL: <https://dashboard.healthit.gov/quickstats/pages/certified-electronic-health-record-technology-in-hospitals.php> [accessed 2020-10-13]
6. Office-based physician electronic health record adoption: Health IT quick-stat #50. Office of the National Coordinator for Health Information Technology. 2019 Jan. URL: <https://dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php> [accessed 2020-10-13]
7. Empower your patients with Health Records on iPhone. Apple. 2019. URL: <https://www.apple.com/healthcare/health-records/> [accessed 2020-10-13]
8. Lewinski AA, Drake C, Shaw RJ, Jackson GL, Bosworth HB, Oakes M, et al. Bridging the integration gap between patient-generated blood glucose data and electronic health records. *J Am Med Inform Assoc* 2019 Jul 01;26(7):667-672. [doi: [10.1093/jamia/ocz039](https://doi.org/10.1093/jamia/ocz039)] [Medline: [31192360](https://pubmed.ncbi.nlm.nih.gov/31192360/)]
9. Cortez A, Hsui P, Mitchell E, Riehl V, Smith P. Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024. Washington, DC: Office of the National Coordinator for Health Information Technology; 2018 Jan. URL: [https://www.healthit.gov/sites/default/files/one\\_pghd\\_practical\\_guide.pdf](https://www.healthit.gov/sites/default/files/one_pghd_practical_guide.pdf) [accessed 2020-10-13]
10. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: MAST. *Int J Technol Assess Health Care* 2012 Jan 23;28(1):44-51. [doi: [10.1017/s0266462311000638](https://doi.org/10.1017/s0266462311000638)]
11. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st century Cures Act era. *JAMA* 2020 Jun 16;323(23):2381-2382. [doi: [10.1001/jama.2020.7858](https://doi.org/10.1001/jama.2020.7858)] [Medline: [32463421](https://pubmed.ncbi.nlm.nih.gov/32463421/)]
12. Klonoff DC, Parkes JL, Kovatchev BP, Kerr D, Bevier WC, Brazg RL, et al. Investigation of the accuracy of 18 marketed blood glucose monitors. *Diabetes Care* 2018 Aug;41(8):1681-1688. [doi: [10.2337/dc17-1960](https://doi.org/10.2337/dc17-1960)] [Medline: [29898901](https://pubmed.ncbi.nlm.nih.gov/29898901/)]
13. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digit Med* 2020;3:18 [FREE Full text] [doi: [10.1038/s41746-020-0226-6](https://doi.org/10.1038/s41746-020-0226-6)] [Medline: [32047863](https://pubmed.ncbi.nlm.nih.gov/32047863/)]

## Abbreviations

**API:** application programming interface  
**EHR:** electronic health record  
**FDA:** US Food and Drug Administration  
**MAST:** Model for ASsessment of Telemedicine applications  
**mHealth:** mobile health  
**PGHD:** patient-generated health data

*Edited by G Eysenbach; submitted 07.08.20; peer-reviewed by Z Ni, J Bolton; comments to author 22.09.20; revised version received 25.09.20; accepted 30.09.20; published 28.10.20.*

*Please cite as:*

Shaw R, Stroo M, Fiander C, McMillan K

*Selecting Mobile Health Technologies for Electronic Health Record Integration: Case Study*

*J Med Internet Res 2020;22(10):e23314*

URL: <http://www.jmir.org/2020/10/e23314/>

doi: [10.2196/23314](https://doi.org/10.2196/23314)

PMID: [33112248](https://pubmed.ncbi.nlm.nih.gov/33112248/)

©Ryan Shaw, Marissa Stroo, Christopher Fiander, Katlyn McMillan. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Tutorial

# Maximizing the Potential of Patient-Reported Assessments by Using the Open-Source Concerto Platform With Computerized Adaptive Testing and Machine Learning

Conrad Harrison<sup>1</sup>, BSc, MBBS, MRCS; Bao Sheng Loe<sup>2</sup>, PhD; Przemysław Lis<sup>2</sup>, MSc; Chris Sidey-Gibbons<sup>3</sup>, PhD

<sup>1</sup>Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

<sup>2</sup>The Psychometrics Centre, University of Cambridge, Cambridge, United Kingdom

<sup>3</sup>MD Anderson Center for INSPIRED Cancer Care, University of Texas, Houston, TX, United States

**Corresponding Author:**

Conrad Harrison, BSc, MBBS, MRCS

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences

University of Oxford

The Botnar Research Centre

Windmill Road

Oxford, OX3 7LD

United Kingdom

Phone: 44 1865227374

Email: [conrad.harrison@medsci.ox.ac.uk](mailto:conrad.harrison@medsci.ox.ac.uk)

## Abstract

Patient-reported assessments are transforming many facets of health care, but there is scope to modernize their delivery. Contemporary assessment techniques like computerized adaptive testing (CAT) and machine learning can be applied to patient-reported assessments to reduce burden on both patients and health care professionals; improve test accuracy; and provide individualized, actionable feedback. The Concerto platform is a highly adaptable, secure, and easy-to-use console that can harness the power of CAT and machine learning for developing and administering advanced patient-reported assessments. This paper introduces readers to contemporary assessment techniques and the Concerto platform. It reviews advances in the field of patient-reported assessment that have been driven by the Concerto platform and explains how to create an advanced, adaptive assessment, for free, with minimal prior experience with CAT or programming.

(*J Med Internet Res* 2020;22(10):e20950) doi:[10.2196/20950](https://doi.org/10.2196/20950)

**KEYWORDS**

computerized adaptive testing; computerized adaptive test; CAT; machine learning; patient reported outcome measures; outcome assessment; Concerto

## Introduction

**Patient-Reported Assessment**

Patient-reported outcome measures (PROMs) measure the outcomes of health care that are most meaningful to patients. The importance and validity of a patient-centered approach to outcome assessment is gaining widespread acceptance across a diverse range of stakeholders including clinicians, researchers, clinical commissioners, health care business strategists, and patients themselves [1-4].

Despite growing interest in the use of patient-reported assessments, there are considerable barriers to their use, especially in time-pressured clinical environments. Two barriers, in particular, are both problematic and addressable. First,

questionnaires can be burdensome to complete, especially when multiple domains of patient health are assessed at the same time [5]. Second, it can be unclear to health care professionals what actions should be taken based on the information received. In this paper, we will discuss strategies to reduce the burden of patient-reported assessment and improve the actionability and relevance of feedback. Specifically, it introduces modern psychometric theories, which can be used to create individualized assessments and reduce the burden of completion, as well as techniques for providing individualized feedback.

**Modern Psychometrics and Item-Response Theory**

The accuracy, reliability, and validity of patient-reported assessments are underpinned by complex psychometric statistical theory. Applying psychometric methods ensures that

scores generated from PROMs can be used with high confidence in clinical practice and research [6].

Psychometrics can be divided into two broad domains. The first, referred to as classical test theory, uses correlational statistics to assess questionnaire properties, for example, how well the responses to certain items (questions) correlate with each other [7]. The second, known as modern test theory, uses probabilistic models to determine the properties of individual items [6].

The benefit of using modern test theory over classical test theory is that modern test theory allows researchers to evaluate the psychometric properties of individual items in relation to a targeted trait, whereas classical test theory correlational statistics mainly focuses on test-level performance. A more detailed comparison between the two approaches can be found elsewhere [6-8]. One of the biggest advantages of using modern test theory is that item properties can be used in a computerized adaptive testing (CAT) environment.

## CAT

CAT refers to a process of selecting the most informative items for people responding to questionnaires. In contrast to fixed-length assessments, where a standard set of items are presented to every respondent all at once, CAT employs a psychometric algorithm to select items one at a time based on the amount of information they will provide about the individual assessment taker [9]. Each item is calibrated using statistical models described by modern test theory, and this process provides parameters for each item which are then used by the CAT algorithm to both calculate a respondent's score and select items [5].

After each response, CAT algorithms calculate a respondent's score based on the information available and select the next most informative item to administer. As more items are answered, the person's score is calculated with increasing accuracy. The CAT will eventually terminate when a *stopping rule* has been met. Stopping rules are typically based on a prespecified time limit, the number of items, the minimum standard error of measurement (SEM), or a combination thereof. Adapting assessments in this way can make them either briefer, more accurate, or in certain cases, both [10].

Many studies have assessed the impact of CAT on the length and accuracy of patient-reported outcome assessments. Experiments conducted both in silico and using human participants have robustly demonstrated that CAT can reduce the length of assessments by more than 50% while keeping excellent agreement between fixed-length assessment scores and CAT [5,11-14].

Despite their impressive performance, the uptake of currently available CAT platforms has been limited, including the Patient-Reported Outcome Measurement Information System (PROMIS) CAT which is accessible in the United States within the Epic electronic health record system [15]. In order to improve the uptake of this transformative technology and move toward truly patient-centered care, we must provide a more accessible way to implement CAT platforms in clinical practice and research. Research has demonstrated that PROM interventions are likely to have the greatest positive impact on

patient outcomes when they are closely aligned with clinical care [16], and to achieve this we must administer such interventions through a versatile platform that meets the needs of clinicians, researchers, and patients.

## Machine Learning

Machine learning refers to the process of developing or *training* algorithms to recognize patterns in existing data and to use this knowledge to make successful predictions with new data [17]. A great deal of enthusiasm has been shown for machine learning, as it has demonstrated exceptional performance in a variety of tasks including predicting the outcome of individuals following certain medical interventions, interpreting diagnostic images, and assessing the meaning of open-text passages [17-20].

There are a number of ways in which the collection and analysis of patient-reported data could be improved using machine learning. For instance, a branch of machine learning known as natural language processing may be used to generate quantifiable information from unstructured passages of open text [20]. Patient-reported assessments with integrated machine learning functions could include written (or spoken) patient responses, quantify said responses in a meaningful way, and use them to make recommendations for patient care or service improvement.

## Concerto

Concerto was developed with the intention of providing a secure, versatile, and easy-to-use platform for creating patient-reported assessments that can incorporate CAT and machine learning. It is free to use and features a point-and-click interface that can be used to build advanced assessments with minimal prior programming experience [21].

Assessments created in Concerto are administered through highly adaptable front-end user interfaces that can be accessed from computers, smartphones, and electronic tablets. These interfaces are built similarly to websites, using HTML, JavaScript and CSS. The Concerto platform comes with inbuilt stock templates for users that do not wish to write their own code.

This user interface interacts with back-end functions, which can include scoring, CAT, and/or machine learning algorithms, using the R programming language. R programming has become popular among statisticians and data scientists for its breadth and accessibility [22]. There are currently over 15,000 available R packages, which can be used free of charge for statistical computing tasks including psychometric analyses, adaptive testing, and machine learning [23]. Concerto incorporates prewritten R code that can administer non-adaptive assessments or computerized adaptive tests, using item parameter tables that are uploaded by the user. The code is fully customizable for developers wishing to create more specialist assessments.

Patient data are stored securely using the MySQL database management system. The Concerto platform itself can be installed on Amazon cloud-based servers that comply with rigorous security demands. Alternatively, it can be installed on local servers (eg, those belonging to a health care provider) to comply with institutional security protocols. This has enabled

the platform to be used successfully in clinical trials and for routine clinical care in the British National Health Service [24].

The Concerto platform can present assessment results immediately. Assessments can be presented in many forms depending on the needs of the end user. For example, radial plots can capture multiple dimensions of a person's health state (eg, different PROM subscale scores) at discrete times, and trend plots can show how a person's scores have changed over time or following major clinical events. Scores can be compared to normative values or other interpretability estimates, and SEMs can be presented alongside CAT scores. Respondents can even receive personalized written feedback to contextualize results (eg, "Your result is... This means..."). Providing immediate graphical and text-based feedback in this way has been shown to improve the experience of assessment when compared with traditional administration [2]. Results can be directly imported into a person's electronic health record through application programming interfaces.

In the following section, we demonstrate how a new Concerto user can create a computerized adaptive test for the Centre for Epidemiologic Studies Depression (CES-D) PROM.

## Concerto: a Worked Example

### Installation

Up to date installation guidance for personal Concerto use can be found at the Concerto GitHub webpage [25]. Readers should be aware that if they choose to use the Amazon Web Service (AWS) for installation, they will need to submit credit or debit card details as part of the registration process. Provided the

default instance type (t2.micro) is selected, the following exercise should fall under Amazon's Free Tier. Readers are solely responsible for any costs they incur, and we would recommend that inexperienced AWS users take care when using the service.

### Download CES-D Items

To complete this exercise, readers will need to download a CSV file that contains the item wordings, item response theory parameters and response options for the CES-D. This is available to download from the Open Science Framework [26].

We will refer to this table as a "flat" item table because all the data are stored in one layer (ie, there are no sub-tables within it).

Download the flat item table, install Concerto, and log in.

### Create a New Assessment

On the *Tests* tab, click *Add new*.

Enter the name of your assessment in the *Name* box (eg, CESD\_adaptive). The *Type* dropdown box should be set to *flowchart*. Click *Save*. Your test should appear under the *Tests* tab.

### Create a Table to Store Item Responses

Click on the *Data Tables* tab and select *Starter content*. Click *Edit* next to the *assessmentResponses* table (see Figure 1). Click *Copy* and change the name of the table (eg, CESDResponses). Click *Save*. This action saves a new table in which to store your responses.

**Figure 1.** Concerto screenshot: creating a data table to store item responses.

Info	ID	Name	Last update by	Last update on	
✓	1	users	-	2020-01-27 15:35:18	Edit Export Delete
✓	2	translationDictionary	-	2020-01-27 15:35:18	Edit Export Delete
✓	3	sessions	-	2020-01-27 15:35:18	Edit Export Delete
✓	4	linearTestResponses	-	2019-12-02 14:53:31	Edit Export Delete
✓	5	linearTestItems	-	2019-12-02 14:53:31	Edit Export Delete
✓	6	CATResponses	-	2019-12-02 14:53:31	Edit Export Delete
✓	7	CATItems	-	2019-12-02 14:53:31	Edit Export Delete
✓	8	assessmentResponses	-	2020-01-27 15:35:18	Edit Export Delete
✓	9	assessmentItems	-	2020-01-27 15:35:18	Edit Export Delete
✓	10	assessmentFlatItems	-	2020-01-27 15:35:18	Edit Export Delete
✓	11	feedbackRanges	-	2020-01-27 15:35:19	Edit Export Delete
✓	12	percentileRanges	-	2020-01-27 15:35:19	Edit Export Delete
✓	13	data	-	2020-01-27 15:35:19	Edit Export Delete

Total Items: 74 (Showing Items: 13)

### Upload Your Items

Click on *All tables* and select *User made*. Click *Add new* and type the name of your item table (eg, CESDFlatItems). Click *Save*. Your item table should appear in the list of user-made tables.

Click *Edit* next to your new item table. Click *Upload CSV*. Check the *Restructure* and *Header row* boxes. Use the *Choose File* button to select your flat item table and click *Save*. Alternatively, when the column header names in the CSV file are identical to those in the default flat item table, this can be copied over from the starter content in the same way as the item



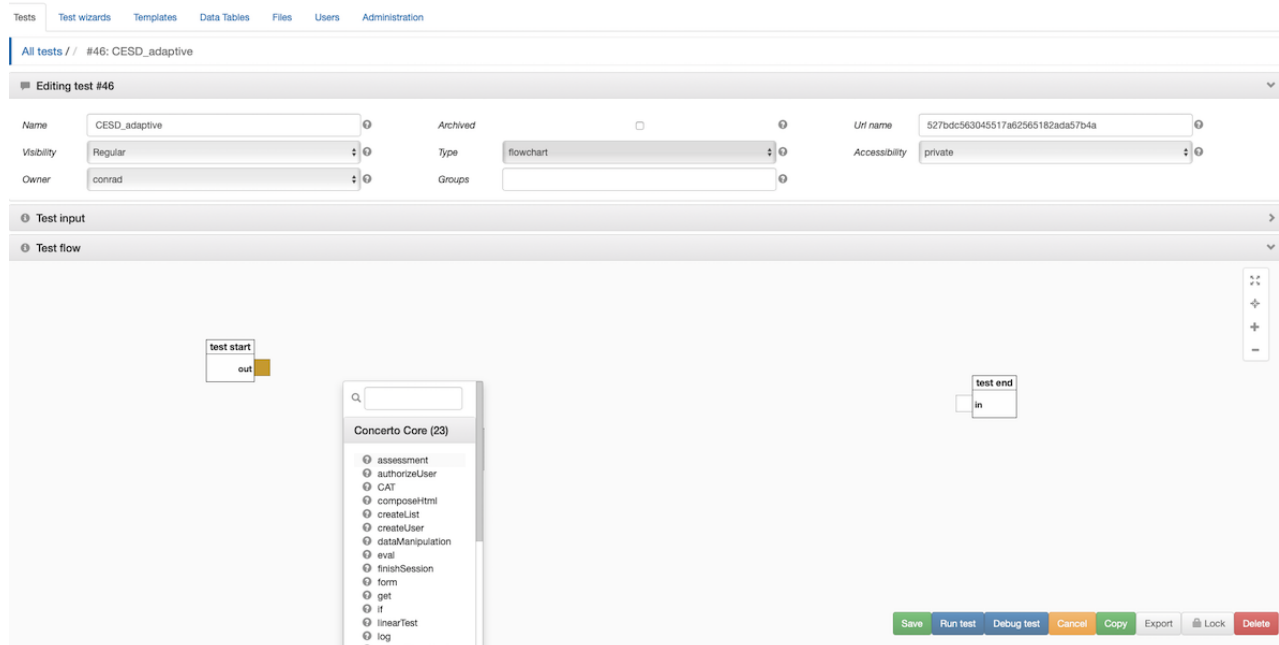
response table. The CSV file can then be uploaded with the *Restructure* box unchecked. This may improve system performance when collecting a large volume of responses by preserving certain database column types.

## Open the Flow Chart

Under the *Tests* tab, click *Edit* next to your assessment. In the *Test flow* window, you will see your assessment displayed as a

flow chart. It will have two nodes: *test start* and *test end*, which have a yellow output port and a white input port, respectively. Drag the *test end* node towards the right-hand side of the window to create some empty space between your nodes. Right-click the space between the nodes to create a third node (see [Figure 2](#)). Select *assessment*. The assessment node wizard should open automatically, with the *Items* tab preselected.

**Figure 2.** Concerto screenshot: opening the flowchart.



## Customize Your CAT

Under *Type*, select *Flat Table* from the drop-down menu. Click the *launch setter dialog* icon under *Flat Table*. Select your item table from the *Table* drop-down menu and click *Save*. Select *CAT* from the *Order* drop-down menu.

Under the *Stopping Rules* tab, you can set stopping rules for your CAT. Try setting *Minimum Accuracy* to 0.5. This number represents the SEM that your assessment will achieve before terminating.

Under the *CAT Options* tab, select *GRM* from the *Model* drop-down menu. This relates to the psychometric model parameters which we will use for the CAT. In this example, our item parameters relate to a model known as the graded response model (GRM) [27].

Under the *Responses* tab, click the *launch setter dialog* icon next to *Response Bank*. Select your responses table from the *Table* drop-down menu. Click *Save*. Check the boxes next to *Calculate Theta* and *Calculate SEM*.

Under the *Templates* tab, enter a name for your assessment in the *Title* box (eg, CESD). Click the *launch setter dialog* icon under *Instructions*. Type some instructions for your assessment (eg, "Below is a list of the ways you might have felt or behaved - please tell me how often you have felt this way during the past week"). Click *Save*. Uncheck the *Show Page Info* box. Click *Save*.

## Add the CAT to Your Assessment

Connect the *test start* node to the *assessment* node by dragging the yellow output port on the *test start* node to the white input port on the *assessment* node. Click the red plus sign on the *assessment* node to create new return ports. Check the boxes next to *theta* and *sem* and click *Save*. This enables the *assessment* node to pass on a person's score (theta) and the SEM associated with that score.

Right click the empty space between your *assessment* node and *test end* node to create a fourth node. Select *scoring*. Choose *Percentile (normal distribution)* from the *Score Type* drop-down menu. Enter a mean of 0 and a standard deviation of 1. Click *Save*.

On the *scoring* node click the blue plus sign to add an input port. Check the box next to *rawScore* and click *Save*. Click the red plus sign on the *scoring* node to add a return port. Check the box next to *score* and click *Save*. Connect the yellow output port from the *assessment* node to the white input port of the *scoring* node. Connect the *theta* return port to the *rawScore* input port.

## Create a Results Page With Contextual Feedback

Right click the empty space between your *scoring* node and your *test end* node to create a fifth node; you may need to reposition the nodes to make sufficient space. Select *showPage*. Enter a title (eg, CESD), then click the *launch setter dialog* icon under *Content* and copy the following:

Theta is {{theta}}. SEM is {{sem}}.

Your score is higher than {{percent}}% of the general population.

The double braces (curly brackets) specify which values for our feedback page to take from our scoring node. Click *Save*. Clear the *Button Label* and click *Save* again.

Click the blue plus sign on the *showPage* node to create a new input port. Type theta into the text box, noting the lowercase. Click *Add*.

Click the blue plus sign on the *showPage* node to create a second input port. Type sem into the text box, again using lowercase. Click *Add*.

Click the blue plus sign on the *showPage* node to create a third input port. Type percent into the text box, again using lowercase. Click *Add*.

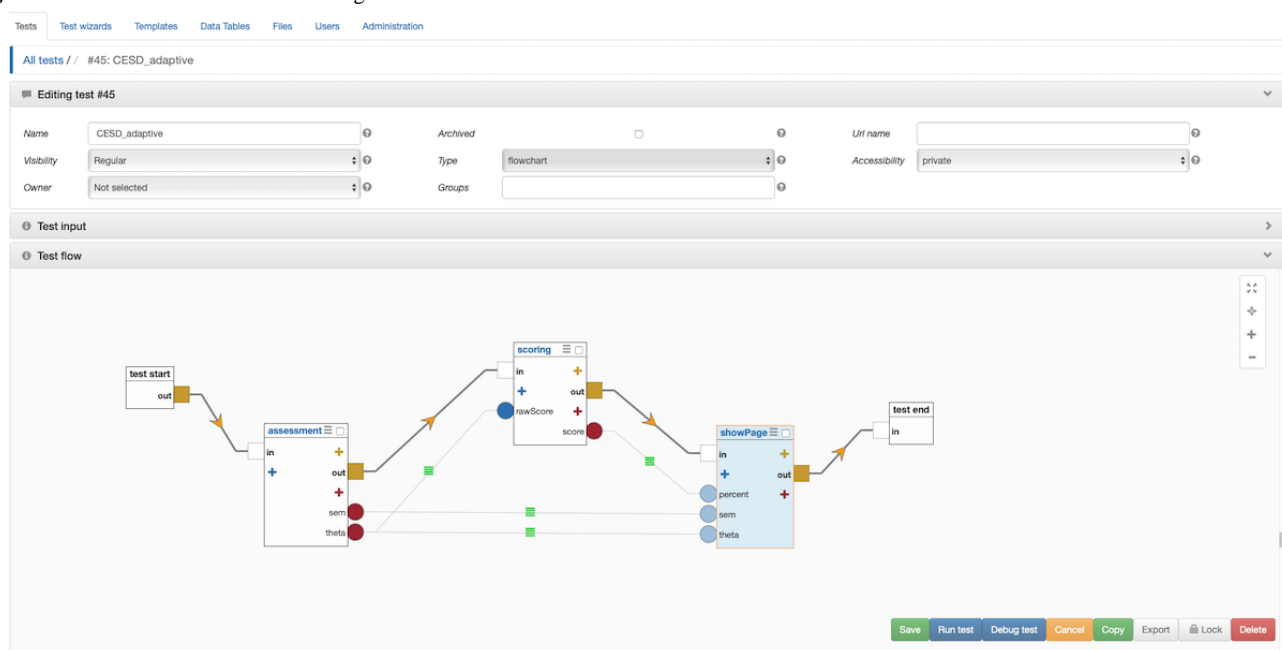
Connect the *sem* return port on the *assessment* node to the *sem* input port on the *showPage* node.

Connect the *theta* return port on the *assessment* node to the *theta* input port on the *showPage* node.

Connect the *score* return port on the *scoring* node to the *percent* input port on the *showPage* node.

Connect the yellow output port on the *scoring* node to the white input port on the *showPage* node. Connect the yellow output port on the *showPage* node to the white input port on the *test end* node (see [Figure 3](#)). Click *Save*. You are now ready to run your assessment. Responses will be stored in the CESDResponses table.

**Figure 3.** Concerto screenshot: connecting nodes.



## Concerto Case Studies

### Quality of Life Assessment

The World Health Organization Quality of Life (WHOQOL)-100 questionnaire was developed in the 1990s by the WHOQOL group as part of an international collaborative effort to produce a generic, cross-cultural, and widely accepted tool for measuring quality of life. The 100-item questionnaire assesses quality of life across six domains (physical, psychological, social, emotional, independence and spiritual), and is also available as a 26 item (WHOQOL-BREF) version with four domains (physical, psychological, social and environmental) [28].

In 2016, our team calibrated item banks from the 100-item questionnaire using modern test theory and trialed a unidimensional CAT version of the questionnaire using simulated data across four domains (physical, psychological, social and environmental). The CAT version of the questionnaire

used 43% and 75% fewer items than the WHOQOL-BREF and WHOQOL-100 questionnaires respectively, at comparable levels of reliability [5]. The WHOQOL CAT, when administered through Concerto, takes a mean of 121 seconds to complete, approximately 10 minutes faster than the WHOQOL-100 [29].

### Analyzing Open-Text Feedback of Doctors' Performance With Machine Learning

Multisource feedback has become a routine part of UK doctors' training and appraisal. Often, these feedback assessments contain open-text comments from colleagues about a doctor's performance [30]. Automating the analysis of these comments could provide real-time, objective insights into both an individual doctor's performance and the interpersonal dynamics of a team or department.

In 2017, our team demonstrated the ability for machine learning algorithms to classify open-text feedback from the General Medical Council Colleague Questionnaire (GMC-CQ) into five themes, with human-level accuracy. These themes were

innovation, interpersonal skills, popularity, professionalism and respect. Doctors classified as professional, respected or with good interpersonal skills achieved higher GMC-CQ scores than those who were not classified as such [31].

Interested readers can freely apply these algorithms to their own open text comments using the Concerto-based platform [32].

### Improving Assessments of Patient Experience With Machine Learning

Patient-reported experience measures (PREMs) that take the form of questionnaires are often limited by a *ceiling effect*. This describes a skew towards positive reporting, which can limit the discriminative ability of a PREM and mask poor service performance [33,34].

We have shown that in the context of UK primary care, spoken feedback from patients can provide more accurate, more detailed, and more actionable insights into the consultation experience than questionnaire results alone. In one study, we found a tendency for patients to rate consultation experiences positively when answering items from the interpersonal skills domain of the national GP Patient Survey, although nearly 60% of respondents who rated their consultation as “good” provided contradictory feedback when interviewed about their experience [35].

Using Concerto, we have developed a patient satisfaction assessment called INSPiRES (Innovative Systems for Patient Reported Experience in Surgery) that combines multiple choice responses with open-text analysis. During the assessment, respondents first select 1 of 12 emotions that describe how they feel about the care they have received. Next, respondents provide an open-text description indicating why they feel that way. Finally, respondents explain which part of their experience led to that feeling (eg, waiting times, cleanliness, care providers) by either selecting a prespecified option or entering free text. The tool is being trialed at the Brigham and Women’s Hospital, Boston, MA, USA, and is expected to provide specific, actionable feedback that will drive service improvement [36]. Although currently used during surgical outpatient clinics, the assessment is also available as a smartphone app that patients can complete from home.

### Discussion

In this article, we have introduced Concerto and demonstrated how to create an advanced, adaptive assessment, for free, with

minimal prior experience of CAT or programming. Concerto assessments can incorporate other features, including those that use machine learning, although this is less straightforward at present. The platform has been used internationally to improve the performance of PROMs in research and clinical practice, to classify open-text assessments of health care providers, and to provide meaningful insights into the experience of health care delivery [2,24,31,32,36].

In future, Concerto could be used to develop and deploy advanced clinical decision support systems (CDSSs) that harness the power of CAT and machine learning to assist clinicians in making evidence-based decisions during daily practice. These systems, which can use patient-reported assessments to predict the outcomes of an individual following a health care intervention, are already being trialed to streamline UK GP referrals and support shared decision making in surgery [37]. Existing CDSSs, most notably the NHS Pathways CDSS, which is used by NHS 111 to triage over 14 million telephone calls a year [38], could be trained to automatically interpret spoken word or open-text through natural language processing.

Concerto-based assessments can be deployed on mobile devices as a tool for remote symptom monitoring. Besides the survival advantage this can bring patients with cancer [16], it has quite obvious implications for a broad range of domiciliary disciplines (eg, out-of-hospital palliative care, general practice, and psychiatry).

Patient-reported assessments can transform clinical practice, research, commissioning, and health care management strategies by measuring the impact of an intervention from the patient’s perspective. To deliver the full potential of these assessments, they should be short, accurate, and acceptable to both respondents and those administering the assessment. They should be personalized, ask only the most relevant questions to an individual, and not be limited to multiple-choice responses. Results should be analyzed in real-time and presented to assessment users in an engaging and meaningful way. Where appropriate, data should be easily available for use in secondary analyses including predictive models. These assessments must integrate easily with health care services, including interoperating with electronic health records. Patient data must be stored and processed securely and ethically.

The Concerto platform bridges the implementation gap between the assessments of today and those of tomorrow.

### Conflicts of Interest

BSL, PS, and CS-GS are employed by the Psychometrics Centre, University of Cambridge. The Concerto platform is open source and free to use, but in some circumstances the Psychometrics Centre will receive revenue from organizations requiring technical support with the platform.

### References

1. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights* 2013 Aug;6:61-68 [FREE Full text] [doi: [10.4137/HSLS11093](https://doi.org/10.4137/HSLS11093)] [Medline: [25114561](https://pubmed.ncbi.nlm.nih.gov/25114561/)]

2. Geerards D, Pusic A, Hoogbergen M, van der Hulst R, Sidey-Gibbons C. Computerized Quality of Life Assessment: A Randomized Experiment to Determine the Impact of Individualized Feedback on Assessment Experience. *J Med Internet Res* 2019 Jul 11;21(7):e12212. [doi: [10.2196/12212](https://doi.org/10.2196/12212)]
3. Porter ME. What Is Value in Health Care? *N Engl J Med* 2010 Dec 23;363(26):2477-2481. [doi: [10.1056/NEJMp1011024](https://doi.org/10.1056/NEJMp1011024)]
4. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013 Jan 28;346(jan28 1):f167-f167. [doi: [10.1136/bmj.f167](https://doi.org/10.1136/bmj.f167)]
5. Gibbons C, Bower P, Lovell K, Valderas J, Skevington S. Electronic Quality of Life Assessment Using Computer-Adaptive Testing. *J Med Internet Res* 2016 Sep 30;18(9):e240. [doi: [10.2196/jmir.6053](https://doi.org/10.2196/jmir.6053)]
6. Nguyen TH, Han H, Kim MT, Chan KS. An Introduction to Item Response Theory for Patient-Reported Outcome Measurement. *Patient* 2014 Jan 9;7(1):23-35. [doi: [10.1007/s40271-013-0041-0](https://doi.org/10.1007/s40271-013-0041-0)]
7. DeVellis RF. Classical test theory. *Med Care* 2006 Nov;44(11 Suppl 3):S50-S59. [doi: [10.1097/01.mlr.0000245426.10853.30](https://doi.org/10.1097/01.mlr.0000245426.10853.30)] [Medline: [17060836](https://pubmed.ncbi.nlm.nih.gov/17060836/)]
8. Cai L, Choi K, Hansen M, Harrell L. Item Response Theory. *Annu. Rev. Stat. Appl* 2016 Jun;3(1):297-321. [doi: [10.1146/annurev-statistics-041715-033702](https://doi.org/10.1146/annurev-statistics-041715-033702)]
9. Weiss DJ, Vale CD. Adaptive Testing. *Applied Psychology* 1987 Jul;36(3-4):249-262. [doi: [10.1111/j.1464-0597.1987.tb01190.x](https://doi.org/10.1111/j.1464-0597.1987.tb01190.x)]
10. Loe BS, Stillwell D, Gibbons C. Computerized Adaptive Testing Provides Reliable and Efficient Depression Measurement Using the CES-D Scale. *J Med Internet Res* 2017 Dec 20;19(9):e302 [FREE Full text] [doi: [10.2196/jmir.7453](https://doi.org/10.2196/jmir.7453)] [Medline: [28931496](https://pubmed.ncbi.nlm.nih.gov/28931496/)]
11. Young-Afat D, Gibbons C, Klassen A, Vickers A, Cano S, Pusic A. Introducing BREAST-Q Computerized Adaptive Testing: Short and Individualized Patient-Reported Outcome Assessment following Reconstructive Breast Surgery. In: *Plastic and Reconstructive Surgery*; 2019.
12. Geerards D, Klassen AF, Hoogbergen MM, van der Hulst RR, van den Berg L, Pusic AL, et al. Streamlining the Assessment of Patient-Reported Outcomes in Weight Loss and Body Contouring Patients. *Plastic and Reconstructive Surgery* 2019;143(5):946e-955e. [doi: [10.1097/prs.0000000000005587](https://doi.org/10.1097/prs.0000000000005587)]
13. Harrison CJ, Geerards D, Ottenhof MJ, Klassen AF, Riff KWW, Swan MC, et al. Computerised adaptive testing accurately predicts CLEFT-Q scores by selecting fewer, more patient-focused questions. 2019 Nov;72(11):1819-1824. [doi: [10.1016/j.bjps.2019.05.039](https://doi.org/10.1016/j.bjps.2019.05.039)]
14. Bass M, Morris S, Neapolitan R. Utilizing Multidimensional Computer Adaptive Testing to Mitigate Burden With Patient Reported Outcomes. 2015 Nov Presented at: AMIA. Annu Symp proceedings AMIA Symp; 2015; San Fransico.
15. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care* 2007;45(Suppl 1):S3-S11. [doi: [10.1097/01.mlr.0000258615.42478.55](https://doi.org/10.1097/01.mlr.0000258615.42478.55)]
16. Denis F, Basch E, Septans A, Bennouna J, Urban T, Dueck AC, et al. Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer. *JAMA* 2019 Jan 22;321(3):306. [doi: [10.1001/jama.2018.18085](https://doi.org/10.1001/jama.2018.18085)]
17. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* 2019 Mar 19;19(1):64 [FREE Full text] [doi: [10.1186/s12874-019-0681-4](https://doi.org/10.1186/s12874-019-0681-4)] [Medline: [30890124](https://pubmed.ncbi.nlm.nih.gov/30890124/)]
18. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman ML, et al. Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review. *World Neurosurgery* 2018 Jan;109:476-486.e1. [doi: [10.1016/j.wneu.2017.09.149](https://doi.org/10.1016/j.wneu.2017.09.149)]
19. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017 Jan 25;542(7639):115-118. [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)]
20. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. *J Mach Learn Res* 2011;2493-2537.
21. The Psychometrics Centre. Concerto Adaptive Testing Platform. University of Cambridge. URL: <https://www.psychometrics.cam.ac.uk/newconcerto> [accessed 2020-10-02]
22. R: A language and environment for statistical computing. R Project. URL: <http://www.r-project.org/> [accessed 2020-10-02]
23. Contributed Packages. CRAN. URL: <https://cran.r-project.org/web/packages/> [accessed 2020-10-02]
24. Gibbons C. Introducing Concerto, an open-source platform designed to realise the potential of modern measurement theories. *Eur Heal Psychol. (S)* 2016;18:421.
25. Lis P. Concerto platform quickstart guide. GitHub. 2019. URL: <https://github.com/campsych/concerto-platform/wiki/quickstart-guide> [accessed 2020-10-02]
26. Delivering the full potential of patient-reported assessments: computerized adaptive testing and machine learning using the open source Concerto platform. OSFHOME. URL: <https://osf.io/8jypd/> [accessed 2020-10-14]
27. Samejima F. Erratum Estimation of latent ability using a response pattern of graded scores. *Psychometrika* 1970 Mar 1;35(1):139-139. [doi: [10.1007/bf02290599](https://doi.org/10.1007/bf02290599)]
28. The Whoqol Group. Development of the World Health Organization WHOQOL-BREF Quality of Life Assessment. *Psychol. Med* 1998 May 01;28(3):551-558. [doi: [10.1017/S0033291798006667](https://doi.org/10.1017/S0033291798006667)]



29. Velikova G, Valderas JM, Potter C, et al. Advances in Patient Reported Outcomes Research. In: Health Qual Life Outcomes. 2017 Oct 05 Presented at: Proceedings of Patient Reported Outcome Measure's (PROMs) Conference; 2017; Oxford p. 185 URL: <https://hqlo.biomedcentral.com/articles/10.1186/s12955-017-0757-y> [doi: [10.1186/s12955-017-0757-y](https://doi.org/10.1186/s12955-017-0757-y)]
30. Whitehouse A, Hassell A, Wood L, Wall D, Walzman M, Campbell I. Development and reliability testing of TAB a form for 360° assessment of Senior House Officers' professional behaviour, as specified by the General Medical Council. Medical Teacher 2009 Jul 03;27(3):252-258. [doi: [10.1080/01421590500047062](https://doi.org/10.1080/01421590500047062)]
31. Gibbons C, Richards S, Valderas JM, Campbell J. Supervised Machine Learning Algorithms Can Classify Open-Text Feedback of Doctor Performance With Human-Level Accuracy. J Med Internet Res 2017 Mar 15;19(3):e65. [doi: [10.2196/jmir.6533](https://doi.org/10.2196/jmir.6533)]
32. The Psychometrics Centre. Colleague feedback tool. Open Text Analysis. 2017. URL: [www.OpenTextAnalysis.com](http://www.OpenTextAnalysis.com) [accessed 2020-10-02]
33. Hodson M, Andrew S, Michael Roberts C. Towards an understanding of PREMS and PROMS in COPD. Breathe 2013 Sep 01;9(5):358-364. [doi: [10.1183/20734735.006813](https://doi.org/10.1183/20734735.006813)]
34. GP Patient Survey 2015. NHS England. 2016. URL: <https://www.england.nhs.uk/statistics/2016/01/07/gp-patient-survey-2015/> [accessed 2020-10-02]
35. Burt J, Newbould J, Abel G, Elliott MN, Beckwith J, Llanwarne N, et al. Investigating the meaning of 'good' or 'very good' patient evaluations of care in English general practice: a mixed methods study. BMJ Open 2017 Mar 02;7(3):e014718. [doi: [10.1136/bmjopen-2016-014718](https://doi.org/10.1136/bmjopen-2016-014718)]
36. Cambridge Network. Converting clicks into care. Cambridge Network. URL: <https://www.cambridgenetwork.co.uk/news/600685> [accessed 2020-10-02]
37. Price A, Smith J, Dakin H, et al. The Arthroplasty Candidacy Help Engine tool to select candidates for hip and knee replacement surgery: development and economic modelling. Health Technol Assess 2019 Jun;23(32):1-216. [doi: [10.3310/hta23320](https://doi.org/10.3310/hta23320)]
38. NHS Pathways. NHS Digital. URL: <https://digital.nhs.uk/services/nhs-pathways> [accessed 2020-04-16]

## Abbreviations

**CAT:** computerized adaptive test  
**CDSS:** clinical decision support system  
**CES-D:** Centre for Epidemiologic Studies Depression Scale  
**GMC-CQ:** General Medical Council colleague questionnaire  
**GRM:** graded response model  
**INSPIRES:** Innovative Systems for Patient Reported Experience in Surgery  
**NLP:** natural language processing  
**PREM:** patient-reported experience measure  
**PRO:** patient reported outcome  
**PROM:** patient-reported outcome measure  
**PROMIS:** Patient-Reported Outcome Measurement Information System  
**SEM:** standard error of measurement  
**WHOQOL:** World Health Organization quality of life

*Edited by G Eysenbach; submitted 02.06.20; peer-reviewed by J Chaplin, M Goldberg; comments to author 21.07.20; revised version received 28.07.20; accepted 18.08.20; published 29.10.20.*

### *Please cite as:*

Harrison C, Loe BS, Lis P, Sidey-Gibbons C  
 Maximizing the Potential of Patient-Reported Assessments by Using the Open-Source Concerto Platform With Computerized Adaptive Testing and Machine Learning  
 J Med Internet Res 2020;22(10):e20950  
 URL: <http://www.jmir.org/2020/10/e20950/>  
 doi: [10.2196/20950](https://doi.org/10.2196/20950)  
 PMID: [33118937](https://pubmed.ncbi.nlm.nih.gov/33118937/)

©Conrad Harrison, Bao Sheng Loe, Przemysław Lis, Chris Sidey-Gibbons. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The



complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Assessment of a Personal Interactive Carbon Monoxide Breath Sensor in People Who Smoke Cigarettes: Single-Arm Cohort Study

Jennifer D Marler<sup>1</sup>, MD; Craig A Fujii<sup>1</sup>, BS, MIDS; Kristine S Wong<sup>1</sup>, BS; Joseph A Galanko<sup>2</sup>, PhD; Daniel J Balbierz<sup>1</sup>, BS; David S Utley<sup>1</sup>, MD

<sup>1</sup>Carrot Inc., Redwood City, CA, United States

<sup>2</sup>Biostatistics Core for the Center for Gastrointestinal Biology and Disease and the biostatistician for the Clinical Nutrition Research Center, Department of Medicine, Division of Gastroenterology and Hepatology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

**Corresponding Author:**

Jennifer D Marler, MD

Carrot Inc.

1400A Seaport Blvd.

Ste. 501

Redwood City, CA, 94063

United States

Phone: 1 4157577696

Email: [marler@carrot.co](mailto:marler@carrot.co)

## Abstract

**Background:** Tobacco use is the leading cause of preventable morbidity and mortality. Existing evidence-based treatments are underutilized and have seen little recent innovation. The success of personal biofeedback interventions in other disease states portends a similar opportunity in smoking cessation. The Pivot Breath Sensor is a personal interactive FDA-cleared (over-the-counter) device that measures carbon monoxide (CO) in exhaled breath, enabling users to link their smoking behavior and CO values, and track their progress in reducing or quitting smoking.

**Objective:** The objective of this study is to assess the Pivot Breath Sensor in people who smoke cigarettes, evaluating changes in attitudes toward quitting smoking, changes in smoking behavior, and use experience.

**Methods:** US adults (18-80 years of age,  $\geq 10$  cigarettes per day [CPD]) were recruited online for this remote 12-week study. Participants completed a screening call, informed consent, and baseline questionnaire, and then were mailed their sensor. Participants were asked to submit 4 or more breath samples per day and complete questionnaires at 1-4, 8, and 12 weeks. Outcomes included attitudes toward quitting smoking (Stage of Change, success to quit, and perceived difficulty of quitting), smoking behavior (quit attempts, CPD reduction, and 7-, 30-day point prevalence abstinence [PPA]), and use experience (impact and learning).

**Results:** Participants comprised 234 smokers, mean age 39.9 (SD 11.3) years, 52.6% (123/234) female, mean CPD 20.3 (SD 8.0). The 4- and 12-week questionnaires were completed by 92.3% (216/234) and 91.9% (215/234) of participants, respectively. Concerning attitude outcomes, at baseline, 15.4% (36/234) were seriously thinking of quitting in the next 30 days, increasing to 38.9% (84/216) at 4 weeks and 47.9% (103/215) at 12 weeks (both  $P < .001$ ). At 12 weeks, motivation to quit was increased in 39.1% (84/215), unchanged in 54.9% (118/215), and decreased in 6.0% (13/215;  $P < .001$ ). Additional attitudes toward quitting improved from baseline to 12 weeks: success to quit 3.3 versus 5.0 ( $P < .001$ ) and difficulty of quitting 2.8 versus 4.3 ( $P < .001$ ). Regarding smoking behavior, at 4 weeks, 28.2% (66/234) had made 1 or more quit attempts ( $\geq 1$  day of abstinence), increasing to 48.3% (113/234) at 12 weeks. At 4 weeks, 23.1% (54/234) had reduced CPD by 50% or more, increasing to 38.5% (90/234) at 12 weeks. At 12 weeks, CPD decreased by 41.1% from baseline ( $P < .001$ ), and 7- and 30-day PPA were 12.0% (28/234) and 6.0% (14/234), respectively. Concerning use experience, 75.3% (171/227) reported the sensor increased their motivation to quit. More than 90% ( $> 196/214$ ) indicated the sensor taught them about their CO levels and smoking behavior, and 73.1% (166/227) reported that seeing their CO values made them want to quit smoking.

**Conclusions:** Use of the Pivot Breath Sensor resulted in a significant increase in motivation to quit, a reduction in CPD, and favorable quit attempt rates. These outcomes confer increased likelihood of quitting smoking. Accordingly, the results support a role for biofeedback via personal CO breath sampling in smoking cessation.

**Trial Registration:** ClinicalTrials.gov NCT04133064; <https://clinicaltrials.gov/ct2/show/NCT04133064>

**KEYWORDS**

smoking cessation; digital health; smartphone; digital sensor; carbon monoxide; breath sensor; biofeedback

## Introduction

Cigarette smoking is the largest preventable cause of morbidity and mortality, accounting for approximately 480,000 annual deaths in the United States, including about 30% of cancer deaths and 30% of cardiovascular disease deaths [1-3]. Smoking remains a pervasive public health problem with a prevalence of 13.7% (34.2 million people) in 2018 [4]. Efforts to advance smoking cessation are a top priority. Launched by the US government in 2010, the 2020 Healthy People initiative identified several related goals including decreasing the prevalence of smoking to 12% or less (tobacco use objective 1.1, or TU-1.1), increasing the proportion of US adults who attempt to quit smoking cigarettes to 80.0% or more (TU-4.1), and increasing recent smoking cessation success to 8.0% or more (TU-5.1) [5]. As a check-in, half way through the 2020 Healthy People agenda in 2015, 68.0% of adult smokers wanted to stop smoking, 55.4% made a past-year quit attempt, and 7.4% recently quit smoking [6,7].

Evidence-based interventions proven to increase quit rates include counseling (individual, group, or phone) and FDA-approved pharmacotherapy. While efficacious, the success of these interventions has been limited by challenges with access, desirability, and convenience, with less than a third of individuals using counseling or medication during quit attempts [6].

Accordingly, smoking cessation is ripe for new technology and approaches. The management of other disease states such as overweight/obesity, hypertension, and diabetes have included the application of personal devices that provide biofeedback, such as wearable activity trackers (eg, Fitbit, Garmin, Jawbone), home-use blood pressure monitors, and continuous glucose monitoring and insulin pumps. These technologies are associated with improved outcomes [8-13], and have a common thread of enabling the user to quantify and monitor personal disease-specific metrics and track progress toward associated goals.

The successes of these novel approaches raise the question of a possible role of personal biofeedback in smoking cessation. One such type of biofeedback is carbon monoxide (CO), a product of the combustion process of smoking. During smoking, CO enters the lungs and crosses pulmonary capillaries to enter the blood stream, where it binds to heme in red blood cells. CO is eliminated from the body by exhalation. Exhaled CO, measured in parts per million (ppm), can be quantified and tracked using a CO breath sensor. With a half-life of approximately 4-5 hours, exhaled CO is well-suited for tracking changes over relatively short periods; once an individual stops smoking, exhaled CO decreases, returning to nonsmoking levels within approximately 24 hours [14].

In “The Tobacco Dependence Treatment Handbook: A Guide to Best Practices,” Abrams et al [15] report that, “providing

individualized feedback about changes in personal levels of carbon monoxide before and after smoking is a powerful message that encourages individuals to make a quit attempt.” Further, Abrams notes that, “In the context of smoking cessation treatment, carbon monoxide levels” and other biomarker feedback “can be utilized to demonstrate the impact of smoking on the smoker and his/her family.”

Beard et al [16] conducted work in this area by providing a personal, mobile CO breath sensor to smokers and asking them to use the monitor regularly throughout the day for 6 weeks, with the goal of maintaining their CO level below 10 ppm. During the first 2 weeks, participants were instructed to record daily their cigarette consumption, usage of the CO monitor and any nicotine replacement therapy, average CO levels, and whether they had attempted to keep their reading below 10 ppm. The participants were not told to quit and were not specifically seeking a quit program. Participants (n=10, 5 males, average age 48.6 years) used the monitor an average of 3 times per day. Average daily cigarette consumption decreased from 14.1 (SD 6.03) at baseline to 9.8 (SD 4.95;  $P=.036$ ) during the 2 weeks of daily CO monitoring and to 9.5 (SD 5.50;  $P=.127$ ) at 6-week follow-up. At follow-up, 50% (5/10) of participants had attempted to quit smoking and one of these participants successfully quit. The majority of smokers reported that they found the CO monitor helpful (79.3%, n=111/140 responses) and that they felt as though the monitors had reduced their cigarette consumption (70%, 7/10 participants). The study investigators concluded that the use of the CO monitors was found to be acceptable and to increase motivation to consider a quit attempt.

In 2018, Patrick et al [17] reported results of a 9-day study of 41 participants using an FDA 510k-cleared mobile CO breath sensor, the Carbon Monoxide Breath Sensor System (COBSS) [17]. This study evaluated the first phase of the multiphase Pivot Smoking Cessation Program, designed to deliver the US Clinical Practice Guidelines for Treating Tobacco Use and Dependence. The focus of the evaluated program phase was to encourage the participants to explore their smoking behavior. Participants completed activities and had the opportunity to log cigarettes within the Pivot app, and interact with a coach via SMS text message-based interactions. More than 80% of participant (34-39 of 41) completed 1 or more CO breath samples each day, and more than 56% (23-27 of 41) completed 5 or more samples each day. In matched pair analyses, significant positive changes in mean attitudes toward quitting (scale 1-10) were evident from baseline (T1) to study exit (T2), including increased readiness to quit (T1 mean 6.1, T2 mean 7.4,  $P=.005$ ), lower perceived difficulty (T1 mean 3.7, T2 mean 5.6,  $P=.001$ ), and greater expectations of success (T1 mean 4.5, T2 mean 6.5,  $P<.001$ ). At exit, 78% (32/41) of participants reported decreasing the number of cigarettes smoked per day during the study.

Marler et al [18] followed the aforementioned study with results in 319 smokers who underwent the complete Pivot program,

which included a personal CO breath sensor, smartphone app, and in-app SMS text messaging–based human coaching. There were significant positive changes in attitudes during the prequit portion of the program, including increased confidence to quit ( $P<.001$ ) and decreased expected difficulty maintaining quit ( $P<.001$ ). Among the participants who completed the final questionnaire and reported the program increased their motivation to stop smoking (85.7%, 233/272), using the breath sensor was the most common reason for the increased motivation. At the end of the program, 7- and 30-day point prevalence abstinence (PPA) rates were 32.0% (102/319, intention to treat [ITT]) and 27.6% (88/319, ITT), respectively. Of those who did not achieve PPA, 25.9% (44/170) had reduced their cigarettes per day (CPD) by 50% or more.

Additional studies have explored the use of exhaled CO as a tool to add to quit programs to bolster motivation and support quit attempts and cessation. Results from these studies are mixed. Foulds et al [19] assessed outcomes 28 days after target quit date in 225 smokers randomized to receive motivational “Lung Age” feedback (exhaled CO values and forced expiratory volume over 1 second) versus minimal feedback. All participants were offered 6 weekly group coaching sessions and nicotine patches. Lung Age feedback did not improve quit rates or compliance with the program. Hajek et al [20] assessed outcomes in pregnant smokers who were randomized to receive midwife-delivered smoking cessation intervention (brief counseling, written materials, arrangements for self-help support, and feedback on exhaled CO levels) versus usual care. A significant difference was reported in those who had quit smoking in the 3 months prior to study start; they had a higher postdelivery PPA rate compared with those who were not recent quitters at the start of the study (65% vs 53%,  $P<.05$ ). However, significant differences were not reported in other outcomes, such as continuous abstinence for at least 3 months prior to delivery or continuous abstinence from 3 months predelivery to 6 months postdelivery. The authors concluded the midwife-delivered intervention did not seem to be an effective method of helping pregnant smokers stop smoking.

Some studies showed favorable early outcomes that did not translate to longer-term results. In 160 smokers randomized to receive cessation leaflets and quit advice (usual care) versus usual care plus exhaled CO level feedback (intervention), Shahab et al [21] reported favorable short-term effects on the cognitive antecedents of smoking behavior and cessation in those who received the intervention. While the investigators reported a greater likelihood of cessation in the intervention group among those with higher self-efficacy, there were no differences in quit attempts or abstinence between the 2 groups at 6 months. McClure et al [22] assessed 536 smokers randomized to receive personally tailored feedback based on lung function, CO exposure, and smoking-related symptoms (experimental group) versus generic information about the risks of smoking and personalized counseling focused on diet, BMI, and physical activity (control group). All participants were advised to quit smoking and offered access to a free telephone smoking cessation counseling program. Immediately post-treatment, the experimental group rated themselves as more likely to try to quit ( $P=.02$ ) and reported a greater mean increase

in their motivation to quit than controls ( $P=.04$ ). These group differences in motivation did not persist at 1-month follow-up. At 6- and 12-month follow-up, there was no greater motivation to quit, use of treatment services, or abstinence in the experimental group compared with controls [23]. Indeed, the control group had greater motivation to quit at 12 months, use of pharmacotherapy at 6 months, and 30-day PPA at 6 months.

And some studies report favorable longer-term results as well. Choi et al [24] randomized 95 adult male smokers to receive 5–10 minutes of smoking cessation education, undergo exhaled CO measurement, and complete questionnaires (intervention) versus receive self-help materials (control). At 4 weeks, motivation to quit was significantly improved in the intervention group ( $P=.03$ ). In another randomized control trial including 98 smokers, home health nurses provided motivation enhancement (motivational interviewing and exhaled CO feedback) or standard care (AHCPR [Agency for Health Care Policy and Research] guidelines for smoking cessation) [25]. Individuals in the motivation enhancement group had more quit attempts and a greater reduction in CPD at follow ups through 12 months.

With the exception of Beard et al [16], the aforementioned studies incorporated exhaled CO as part of multicomponent smoking cessation programs. As a result, it is difficult to specifically identify the role of exhaled CO in reported motivation and smoking behavior outcomes. Moreover, with few exceptions [16–18], these exhaled CO measurements were obtained through health care providers or study personnel during study visits, with very few CO breath samples collected during these studies. It is, however, unclear how outcomes might differ when exhaled CO is regularly measured and tracked by smokers as personal biofeedback. Accordingly, this study sought to more directly focus on the potential role of exhaled CO when breath was sampled by smokers themselves using a personal mobile breath sensor, assessing changes in attitudes toward quitting smoking, changes in smoking behavior, and use experience.

## Methods

### Study Design

This was a prospective, open-label single-arm cohort study conducted with Institutional Review Board approval. The study was performed remotely on an ambulatory basis. Study participants were asked to set up the Pivot Breath Sensor and participate for 12 weeks with an emphasis on providing daily breath samples and completing online study questionnaires periodically throughout the study.

### Consent and Ethical Approval

All participants provided electronic informed consent before participation. The study was reviewed and approved by Solutions IRB (protocol number 2019/09/3), and registered with ClinicalTrials.gov (NCT04133064).

### Study Device

The Pivot Breath Sensor is a component of Pivot’s comprehensive evidence-based digital tobacco cessation solution, which also includes an interactive mobile Pivot app,

lessons based in cognitive behavioral therapy and self-determination theory, nicotine replacement therapy, dedicated human coaching by tobacco treatment specialists via SMS text messaging, and a moderated online community. In keeping with this study's focus on the impact and use experience of the Pivot Breath Sensor, participants were not provided access to any of the other aspects of the Pivot cessation program.

The Pivot Breath Sensor ([Figure 1](#)) comprises a personal interactive breath sensor that measures the level of CO in

exhaled breath and displays the CO value (ppm) to the user directly on the device screen. The CO log is accessed from the sensor screen and shows the most recent exhaled breath CO value at the top of the screen. The user can view previous values by scrolling within the log. The sensor is portable, battery-powered, and rechargeable using a micro-USB cable. The user submits a breath sample by exhaling (blowing) into the breath sensor mouthpiece. CO values are color coded with the color levels (red:  $\geq 10$  ppm, orange: 7-9 ppm, green: 0-6 ppm) detailed in the labeling ([Figure 1](#)).

**Figure 1.** Pivot Breath Sensor and color coding of carbon monoxide values.



## Eligibility

The anticipated user population for the Pivot Breath Sensor are lay users who are smokers, aged between 18 and 80, and capable of using a smartphone and basic smartphone apps. As such, study participant inclusion criteria included all of the following: 18-80 years of age, current daily cigarette smokers ( $\geq 10$  CPD), resident of the United States, able to read and comprehend English, owns and uses a smartphone compatible with the study

app (iPhone 5 and above with iOS 11 and above, or an Android smartphone with Android 5.0 and above), and willing to sign the informed consent form. Exclusion criteria included pregnancy (self-reported) or participation in a previous study sponsored by Carrot Inc.

The study employed nonproportional quota sampling ([Table 1](#)) with the aim of enrolling a study population that reflects the expected initial intended user population.



**Table 1.** Nonproportional quota sampling enrollment: targeted proportions.

Category and subcategory	Targeted %
<b>Age (years)</b>	
18-29	≤20
30-60	≥70
61-80	≤10
<b>Cigarettes per day (CPD)</b>	
10-19	40-60
≥20	40-60
<b>Stage of Change<sup>a</sup></b>	
Intend to quit within 30 days	≥20
Intend to quit within 6 months	≥20
Not thinking of quitting	<20
<b>Gender</b>	
Female	40-60
<b>Employment</b>	
Unemployed	4-8 <sup>b</sup>
Employed <20 hours/week	Remainder of sample
Employed ≥20 hours/week	Remainder of sample

<sup>a</sup>Stage of Change question and answer choices: Are you seriously thinking of quitting smoking? (1) Yes, within the next 30 days; (2) Yes, within the next 6 months; (3) No, not thinking of quitting.

<sup>b</sup>The employment rate among study participants was sought to align with the employment rate among the general US population at the time of protocol submission, which was 3.7% [26].

## Recruitment

Participants were recruited in the United States from September 2019 through November 2019 using web media (Facebook). Potential participants were asked to provide contact information, and answer questions on demographics, smartphone ownership, and smoking behavior using the online screening form. Study staff reviewed each potential participant's responses. All study participants underwent a screening phone call where study eligibility was confirmed. Potential participants were called on a first-come-first-served basis with nonproportional quota sampling enrollment guidelines applied. During this call, study personnel informed the potential participant of the study details and answered any questions. Eligible potential participants were offered the opportunity to participate in the study. Potential participants interested in proceeding were emailed the electronic informed consent form. Upon completion, participants completed a baseline questionnaire and were mailed the Pivot Breath Sensor. Participants were considered enrolled after electronically completing the informed consent form and baseline questionnaire, pairing their breath sensor to the study app on their smartphone, and completing their first breath sample.

## Study Procedure

Participants self-trained on the Pivot Breath Sensor using the device labeling, which included product packaging, a Quick Start Guide, and package insert. In addition, participants were asked to load the study app on their smartphone. This app served

as a means to sync breath sample data from the breath sensor and transmit these data to the study team. Participants had access to a technical support phone line and online user manual. A member of the study staff or customer support optionally called, texted, or emailed the participant to assist in device setup if needed. The participant initiated using the breath sensor. The enrollment date was considered study day 1.

Study participants were instructed to use the breath sensor daily for the duration of the study, with a recommendation of completing 4 or more breath samples a day, spread over the course of the day. This suggested use pattern was provided in the labeling materials and during the screening phone call but breath sensor use was ultimately at the discretion of the participant. Participants received up to twice weekly SMS text message instructions to sync their breath sensor data using the study app. No recommendations were made to participants regarding smoking behavior, and the device and study were not positioned as a smoking cessation program with participants.

Participants received periodic electronic questionnaires via email (SurveyMonkey) that focused on attitudes toward quitting, smoking behavior, and use experience with focus on impact and associated learning. There were 7 questionnaires in total, emailed at baseline, and study days 7 (1 week), 14 (2 weeks), 21 (3 weeks), 28 (4 weeks), 56 (8 weeks), and 84 (12 weeks). Participants received periodic reminders from study staff to complete the questionnaires via email, SMS text messages, or phone, as needed. On study day 84, participants received the

final questionnaire and were asked to send the Pivot Breath Sensor back using a provided prepaid mailer.

Participants were compensated for collecting breath samples (US \$5/day for every day in which  $\geq 4$  breath samples were collected during the first 28 days of the study and thereafter US \$10/week for up to 8 weeks in which  $\geq 20$  breath samples/week were collected; up to US \$220 in total for 84 days of breath sampling), completing the online questionnaires (US \$10-50/questionnaire; up to US \$220 in total for 7 questionnaires), and for returning the Pivot Breath Sensor (US \$50). Participants could earn up to US \$490 in total. Compensation was in the form of Visa gift cards that were mailed to participants. Payments were bundled over 5 payments and took 2-3 weeks to arrive to the participant after being ordered.

**Table 2.** Measures assessing attitudes toward quitting smoking.

Question	Answer Options/Scale
Are you seriously thinking of quitting smoking? (Stage of Change)	“Yes, within the next 30 days” or “Yes, within the next 6 months” or “No, not thinking of quitting”
Would you like to completely stop smoking cigarettes?	“Yes” or “No”
How ready are you to quit smoking?	Scale 1-10 (1=Not at all ready, 10=Completely ready)
If you were to quit smoking right now, how successful would you be?	Scale 1 to 10 (1=Not at all successful, 10=Completely successful)
If you were to quit smoking right now, how difficult do you think it would be to stay smoke free?	Scale 1 to 10 (1=Really hard to stay quit, 10=Really easy to stay quit)
What is your goal when it comes to smoking?	“I don’t have a clear goal in mind” or “I want to quit smoking for good, even though I might slip up” or “I want to quit smoking for good” or “I want to reduce my smoking (like smoking less, or quitting for a while and deciding later if I want to quit)” or “Other goal _____”

Smoking behavior outcomes comprised quit attempts, change in CPD, proportion who reduced CPD by 50% or more, and smoking cessation via 7- and 30-day PPA. A quit attempt was defined as going at least one day without smoking cigarettes, even a single puff. Participants were considered to have achieved 7-day (30-day) PPA if they answered *no* to the following question: “In the last 7 (30) days have you smoked any cigarettes, even a single puff?” As the sensor is designed and was implemented here as a tool to be used independently, without requiring face-to-face contact, and data collection was achieved through remote means using the study app and electronic questionnaires, biochemical verification of smoking status was not pursued in accordance with previous recommendations [27]. If participants reported abstinence but indicated they were smoking 1 or more CPD, they were counted as actively smoking in analyses. Finally, participants were asked to expound on their use experience by providing feedback on the impact and learning associated with the Pivot Breath Sensor.

Because previous studies on CO breath sensor use in smoking cessation have reported mixed results, some with changes in attitudes and behaviors documented early that did not persist [21,22] and some with longer-term changes observed [24,25], primary and secondary endpoints were obtained at 4 weeks in this study. These outcomes were also assessed at 12 weeks for longer-term results, which are also reported herein. Overall,

## Data Collection

Data collection took place on electronic case report forms completed by study participants via SurveyMonkey, and through data collected in a study app which was paired to the breath sensor via Bluetooth. Breath sensor usage data and CO results populated the app. The study team periodically synchronized and uploaded logs from the sensors during the study, which was enabled by the study app.

## Outcomes

Study outcomes focused on 3 areas: attitudes toward quitting smoking, smoking behavior, and use experience. Attitudes toward quitting smoking included Stage of Change, desire to quit (yes/no), readiness to quit, (scale 1-10), confidence to quit (scale 1-10), difficulty to quit (scale 1-10), and goals (multiple choice, 5 options; Table 2).

endpoint assessment was designed to capture shorter- and longer-term changes, if present, as informed by previous studies.

The primary endpoint assessed change in motivation to quit smoking via response to Stage of Change at 4 weeks, compared with baseline. A positive outcome was defined as a participant responding as more motivated to quit. For example, a change in response from seriously thinking of quitting smoking “...within the next 6 months” at baseline to seriously thinking of quitting smoking “...within the next 30 days” at 4 weeks was considered a positive outcome [28].

Secondary endpoints included the proportion of participants who reported 1 or more quit attempt by 4 weeks and the proportion of participants who reduced their CPD by 50% or more by 4 weeks.

## Sample Size

Consideration for the sample size included powering the study to observe a clinically meaningful change from baseline for the primary and secondary endpoints using 80% power at a statistical significance of  $P < .05$ . The primary endpoint was informed by preliminary data from a similar 37-participant pilot study (data not shown). Match-paired data from the 35 participants who completed the 14-day timepoint questionnaire showed that motivation to quit, via Stage of Change, had increased in 31% (11/35), remained unchanged in 66% (23/35),

and decreased in 3% (1/35) of participants from baseline. To detect a statistically significant change of these proportions in this study would require enrolling 50 participants.

The estimated proportion of participants achieving 1 or more quit attempt was 25% based on interim results from the aforementioned pilot study. Based on the median prevalence of 65.4% for past-year quit attempts in the general population [29], the average monthly quit attempt rate is approximately 5%. In the context of a 4-week outcome, to show that 25% of participants making a quit attempt by 4 weeks is statistically different from 5% would require enrolling 16 participants.

Finally, previous work indicates approximately 1% of the general population [30-34] and 2%-5% of individuals in cigarette reduction studies [35,36] will reduce their CPD by 50% or more on a monthly basis. Assuming that 10% of participants in this study will reduce their CPD by 50% or more by 4 weeks would require enrolling 185 participants to show a difference from 5%. Taking these analyses into consideration along with expected attrition, this study targeted enrollment of 220 participants.

### Statistical Analysis

Changes in measurements from baseline were assessed at different timepoints in the study. Participants served as their own controls and tests for any change were performed. Analyses were conducted to calculate mean (SD) for normally distributed variables for actual data, or mean (SE) for modeled data. Median (IQR) values were used in instances of non-normally distributed variables. As applicable, paired one-sample *t*-test was used for numeric data. For one-sample change in binary outcomes, compared binomial proportion test was applied. Fisher exact or chi-square tests were used for categorical data. McNemar test was applied for 2-category match-paired data. Stuart–Maxwell test was used for 3-category match-paired data. To evaluate changes in attitudes or changes in CPD over time, repeated

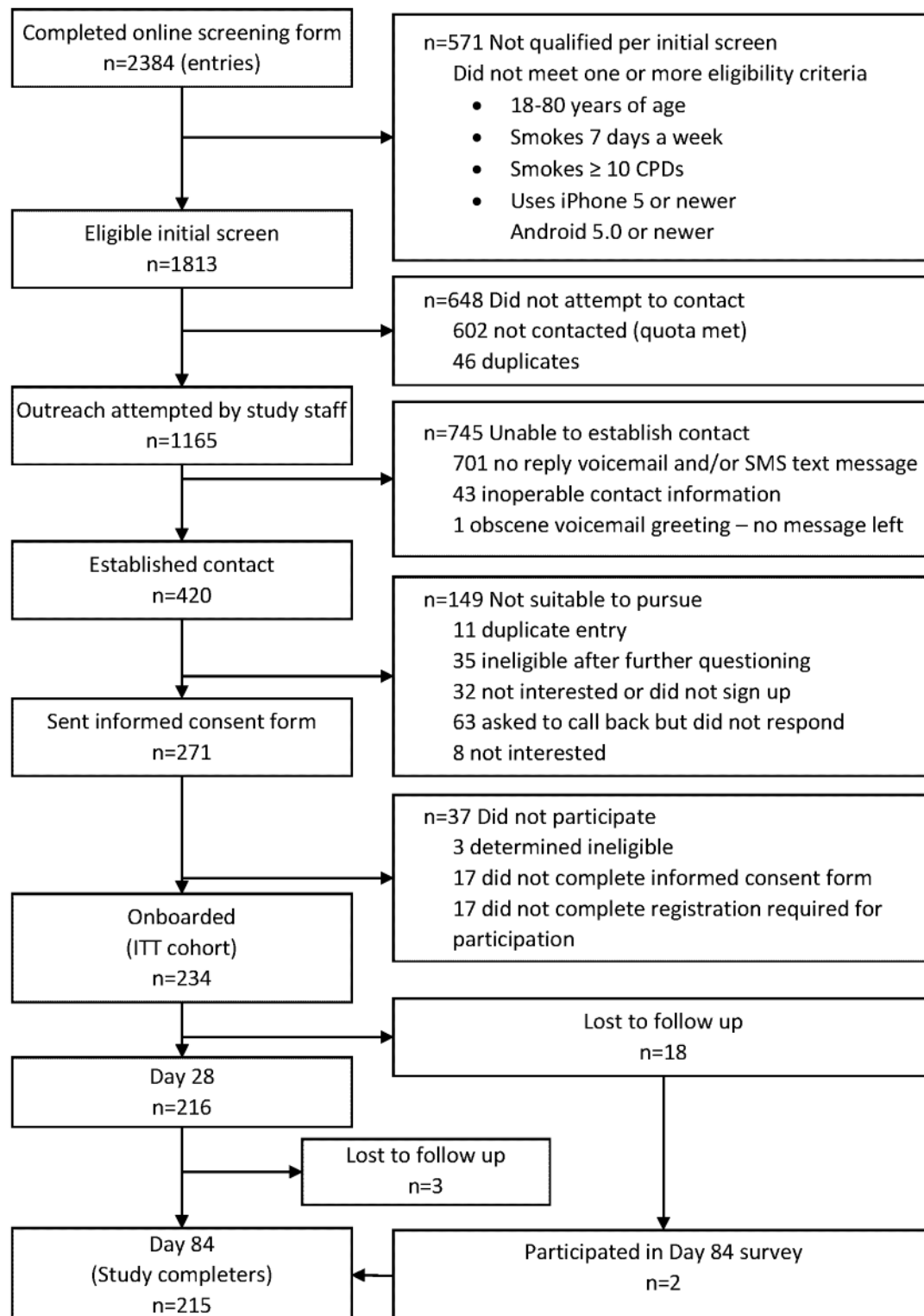
measures linear mixed model analyses were performed using a compound symmetric correlation matrix to model the repeated measures within subjects. To make specific comparisons across time, *F* statistics were computed using the results from the model. Analyses were conducted using SAS version 9.4 (SAS Institute). Statistical significance was set at  $P < .05$ .

Analyses for the outcomes were evaluated for only those who completed the questionnaire (ie, completer) and assessed to take into account missing data. To be included in any completer analysis the participant must have provided a baseline response and the outcome response. For outcomes that were missing data, these were assessed pending the data type. For categorical data, the last response postbaseline and pre-4 weeks was carried forward for the 4-week assessment, and the last response postbaseline and pre-12 weeks was carried forward for the 12-week assessment. If there were no data to carry forward, the data were considered unchanged from baseline. Numeric data used linear mixed modeling.

## Results

### Enrollment

Potential participants responded to an online Facebook ad that included a link to the online screening form. The online screening form was completed by 2384 individuals, of whom 1813 were eligible for the study. Eligible potential participants were called on a first-come-first-served basis with nonproportional quota sampling enrollment guidelines applied; outreach was made to 1165 potential participants. The majority of phone calls went unanswered and unreturned; contact was established with 420 individuals. The electronic informed consent form was sent to 271 individuals, 234 of whom enrolled in the study. Study participant flow is depicted via a Consolidated Standards of Reporting Trials (CONSORT) diagram in Figure 2.

**Figure 2.** Study participant flow: Consolidated Standards of Reporting Trials (CONSORT) diagram.

The study overenrolled by 6.4% (14/220 participants) with a final enrollment of 234 participants. Because of the multistep process of enrollment and expected attrition over the course of the study, we put forth a good faith effort to ensure we obtained data from 185 participants at 4 weeks. The results of these efforts yielded this slight overenrollment. Considering the minimal risk profile of the device and the voluntary, ambulatory nature of the study in which participants gave breath samples

and completed online questionnaires at their discretion, this overenrollment was not felt to be significant.

The study sample consisted of 52.6% women (123/234), had a mean age of 39.9 (SD 11.3) years, smoked a mean of 20.3 (SD 8.0) CPD at baseline, and had been smoking for a mean of 21.7 (SD 11.5) years.

At baseline, 15.4% (36/234) indicated they were seriously thinking of quitting smoking in the next 30 days, 76.9%

(180/234) indicated they were thinking of quitting in the next 6 months, and 7.7% (18/234) indicated they were not seriously thinking of quitting smoking. On average, participants had made 2.1 (SD 6.3) quit attempts over the past 12 months.

All nonproportional quota sampling targets were achieved with the exception of 20% or more ( $\geq 47/234$ ) indicating they intended to quit smoking in the next 30 days. This differential is due to change in participant response to this question between when it was asked on the online screening form and when it was asked on the baseline questionnaire. Specifically, 29.1% (68/234) of study participants indicated they were seriously thinking of quitting smoking in the next 30 days on the online screening

form; however, this decreased to 15.4% (36/234) of participants on the baseline questionnaire. This difference was not felt likely to significantly affect outcomes. All individuals who changed their response went from, “Seriously thinking of quitting in the next 30 days” at online screening to “Seriously thinking of quitting in the next 6 months” at baseline, thereby maintaining some interest in quitting smoking, albeit on a longer timeline. If anything, the effect of enrolling a cohort slightly less motivated to quit than originally anticipated may have made it more challenging to achieve some outcomes, such as quit attempt rates and change in CPD. Study demographic details, including targeted and actual nonproportional quota proportions, are provided in [Table 3](#).



**Table 3.** Baseline demographics including targeted and actual nonproportional quota sampling (N=234, when applicable).

Characteristic	Values <sup>a</sup>	Target %
<b>Gender</b>		
Male	111 (47.4)	40-60
Female	123 (52.6)	40-60
<b>Age (years), mean (SD)</b>	39.9 (11.3)	
18-29 years	37 (15.8)	≤20
30-60 years	185 (79.1)	≥70
61-80 years	12 (5.1)	≤10
<b>Ethnicity</b>		
White	209 (89.3)	
Hispanic, Latinx, or Spanish Origin	5 (2.1)	
Black or African American	9 (3.8)	
Asian	3 (1.3)	
American Indian or Alaska Native	2 (0.9)	
Middle Eastern or North African	1 (0.4)	
Hawaiian or Other Pacific Islander	0 (0)	
Some other race, ethnicity, or origin	5 (2.1)	
Prefer not to answer	0 (0)	
<b>US region</b>		
Northeast	34 (14.5)	
South	112 (47.9)	
Midwest	48 (20.5)	
West	40 (17.1)	
<b>Education</b>		
Less than 8th grade	0 (0)	
Some high school	7 (3.0)	
High school/General educational development	50 (21.4)	
Some college	107 (45.7)	
Associate's (2-year) degree	37 (15.8)	
Bachelor's (4-year) degree	25 (10.7)	
Master's degree	7 (3.0)	
Professional or doctorate degree	1 (0.4)	
<b>Employment</b>		
Unemployed	11 (4.7)	4-8
Employed <20 hours/week	29 (12.4)	Remainder of sample
Employed ≥20 hours/week	194 (82.9)	Remainder of sample
<b>Household income</b>		
<US \$25,000	53 (22.6)	
US \$25,000 to US \$34,999	53 (22.6)	
US \$35,000 to US \$49,999	47 (20.1)	
US \$50,000 to US \$74,999	42 (17.9)	
US \$75,000 to US \$99,999	18 (7.7)	
US \$100,000 to US \$149,999	14 (6.0)	

Characteristic	Values <sup>a</sup>	Target %
≥US \$150,000	4 (1.7)	
Prefer not to answer	3 (1.3)	
<b>Smartphone type</b>		
iPhone	59 (25.2)	
Android	175 (74.8)	
Years smoking, mean (SD)	21.7 (11.5)	
<b>Cigarettes per day (CPD), mean (SD); range (min-max)</b>	20.3 (8.0); 7-50	
<20 CPD	107 (45.7)	40-60
≥20 CPD	127 (54.3)	40-60
<b>How soon after waking up do you typically smoke your first cigarette?</b>		
Within 5 minutes	114 (48.7)	
5 to 30 minutes	99 (42.3)	
31 to 60 minutes	13 (5.6)	
60+ minutes	8 (3.4)	
<b>Would you like to completely stop smoking cigarettes?</b>		
Yes	217 (92.7)	
No	17 (7.3)	
<b>Motivation to Quit (Stage of Change)—Are you seriously thinking of quitting smoking?</b>		
Yes, within the next 30 days	36 (15.4)	≥20
Yes, within the next 6 months	180 (76.9)	≥20
No, not thinking of quitting	18 (7.7)	<20
<b>What is your goal when it comes to smoking?</b>		
I don't have a clear goal in mind	40 (17.1)	
I want to quit smoking for good, even though I might slip up	75 (32.1)	
I want to quit smoking for good	84 (35.9)	
I want to reduce my smoking (like smoking less, or quitting for a while and deciding later if I want to quit)	34 (14.5)	
Other goal	1 (0.4)	
Quit attempts over past 12 months <sup>b</sup> , mean (SD)	2.1 (6.3)	
<b>Use of other tobacco products</b>		
Cigars, cigarillos or little filtered cigars	31 (13.2)	
A regular pipe	0 (0)	
Hookah or water pipe	5 (2.1)	
E-cigarettes or vape	67 (28.6)	
Smokeless tobacco, chew, or snuff	8 (3.4)	
Readiness to quit <sup>c</sup> , mean (SD)	5.6 (2.7)	
Success to quit <sup>d</sup> , mean (SD)	3.3 (2.3)	
Difficulty to quit <sup>e</sup> , mean (SD)	2.8 (2.4)	

Characteristic	Values <sup>a</sup>	Target %
First carbon monoxide measurement (ppm) <sup>f</sup> , mean (SD)	26.7 (18.4)	

<sup>a</sup>All data are presented as n (%) unless otherwise indicated.

<sup>b</sup>Quit attempt="How many times have you tried to quit smoking where you've gone at least one day without smoking a cigarette, even a single puff?"

<sup>c</sup>How ready are you to quit smoking (1=Not at all ready, 10=Completely ready).

<sup>d</sup>If you were to quit smoking right now, how successful would you be? (1=Not at all successful, 10=Completely successful).

<sup>e</sup>If you were to quit smoking right now, how difficult do you think it would be to stay smoke free? (1=Really hard to stay quit, 10=really easy to stay quit).

<sup>f</sup>ppm: parts per million.

## Attitudes Toward Quitting Smoking

For the primary endpoint, motivation to quit smoking at 4 weeks was significantly increased compared with baseline as measured with Stage of Change (Table 4). Motivation to quit smoking improved with 38.9% (84/216) of respondents indicating they

were seriously thinking of quitting in the next 30 days compared with 14.4% (31/216) at baseline ( $P<.001$ ). At 4 weeks, motivation to quit smoking increased in 29.6% (64/216), was unchanged in 66.7% (144/216), and decreased in 3.7% (8/216;  $P<.001$ ).

**Table 4.** Change in motivation to quit smoking (N=216) at baseline (rows) versus 4 weeks (columns).

	Motivation to Quit: 4 weeks, n (%)			Total <sup>b</sup>
	Yes, within the next 30 days	Yes, within the next 6 months	No, not thinking of quitting	
<b>Motivation to Quit <sup>a</sup> : Baseline</b>				
Yes, within the next 30 days, n (%)	25 (11.6)	6 (2.8)	0 (0)	31 (14.4)
Yes, within the next 6 months, n (%)	58 (26.9)	108 (50.0)	2 (0.9)	168 (77.8)
No, not thinking of quitting, n (%)	1 (0.5)	5 (2.3)	11 (5.1)	17 (7.9)
Total, n (%) <sup>b</sup>	84 (38.9)	119 (55.1)	13 (6.0)	216 (100.0) <sup>c</sup>

<sup>a</sup>Motivation to quit smoking assessed via Stage of Change question: "Are you seriously thinking of quitting smoking?" (1) "Yes, within the next 30 days"; (2) "Yes, within the next 6 months"; (3) "No, not thinking of quitting."

<sup>b</sup> $P<.001$ .

<sup>c</sup>A total of 234 participants enrolled in the study; however, only 216 completed the 4-week questionnaire, who are represented here.

Assuming a worst case scenario in which the 18 participants who did not complete the 4-week questionnaire had the lowest possible motivation to quit smoking at 4 weeks (No, not thinking of quitting), 35.9% (84/234) of respondents would have fallen in the category of "seriously thinking of quitting in the next 30 days," compared with 15.4% (36/234) at baseline ( $P<.001$ ). Motivation to quit smoking would have increased in 27.4% (64/234), remain unchanged in 62.0% (145/234), and decreased in 10.7% (25/234;  $P<.001$ ).

There were further increases in motivation at 12 weeks. Among the 215 study participants who completed the 12-week questionnaire, motivation to quit smoking improved, with 47.9% (103/215) of respondents indicating they were seriously thinking of quitting in the next 30 days compared with 14.9% (32/215) at baseline ( $P<.001$ ). Motivation to quit smoking increased in 39.1% (84/215), was unchanged in 54.9% (118/215), and decreased in 6.0% (13/215;  $P<.001$ ).

Similar to the previous worst case analysis, if all 19 participants who did not complete the 12-week questionnaire had the lowest possible motivation to quit smoking (No, not thinking of quitting), 44.0% (103/234) of respondents would have fallen in the category of "seriously thinking of quitting in the next 30 days," compared with 15.4% (36/234) at baseline ( $P<.001$ ). Motivation to quit smoking would have increased in 35.9% (84/234), remain unchanged in 50.9% (119/234), and decreased in 13.2% (31/234;  $P<.001$ ).

Participants were asked if they would like to stop smoking. Matched responses at baseline and at 12 weeks are shown in Table 5. The majority (>90%, >197/215) of participants indicated they would like to stop smoking at baseline and at 12 weeks, with 12/216 people (5.6%) changing their response: 5 from yes to no and 7 from no to yes ( $P=.77$ ).

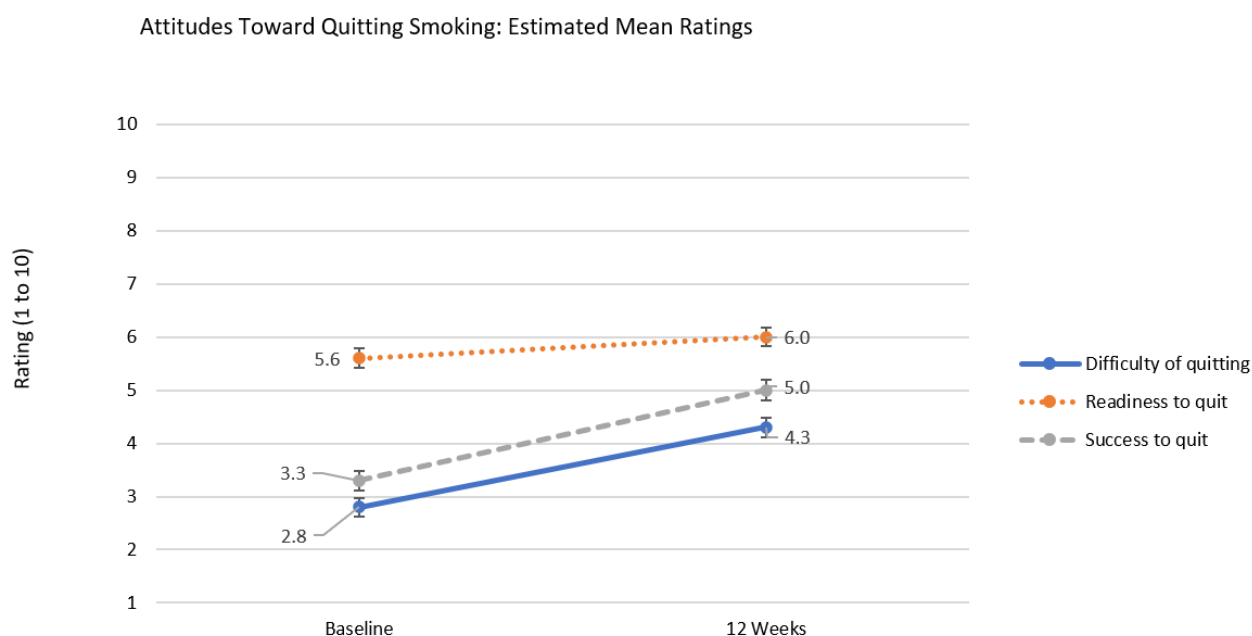
**Table 5.** Would you like to completely stop smoking cigarettes (N=215)? Select one.

	12 weeks, n (%)		
	Yes	No	Total <sup>a</sup>
<b>Baseline</b>			
Yes, n (%)	193 (89.8)	5 (2.3)	198 (92.1)
No, n (%)	7 (3.3)	10 (4.7)	17 (7.9)
Total <sup>a</sup> , n (%)	200 (93.0)	15 (7.0)	215 <sup>b</sup> (100.0)

<sup>a</sup> $P=.77$ .<sup>b</sup>234 participants enrolled in the study; however, only 215 completed the 12-week questionnaire, who are represented here.

Assessment of readiness to quit, success to quit, and perceived difficulty of quitting at baseline versus 12 weeks is depicted in [Figure 3](#). Using linear mixed models to include intervening timepoints, the readiness to quit at 12 weeks was estimated not

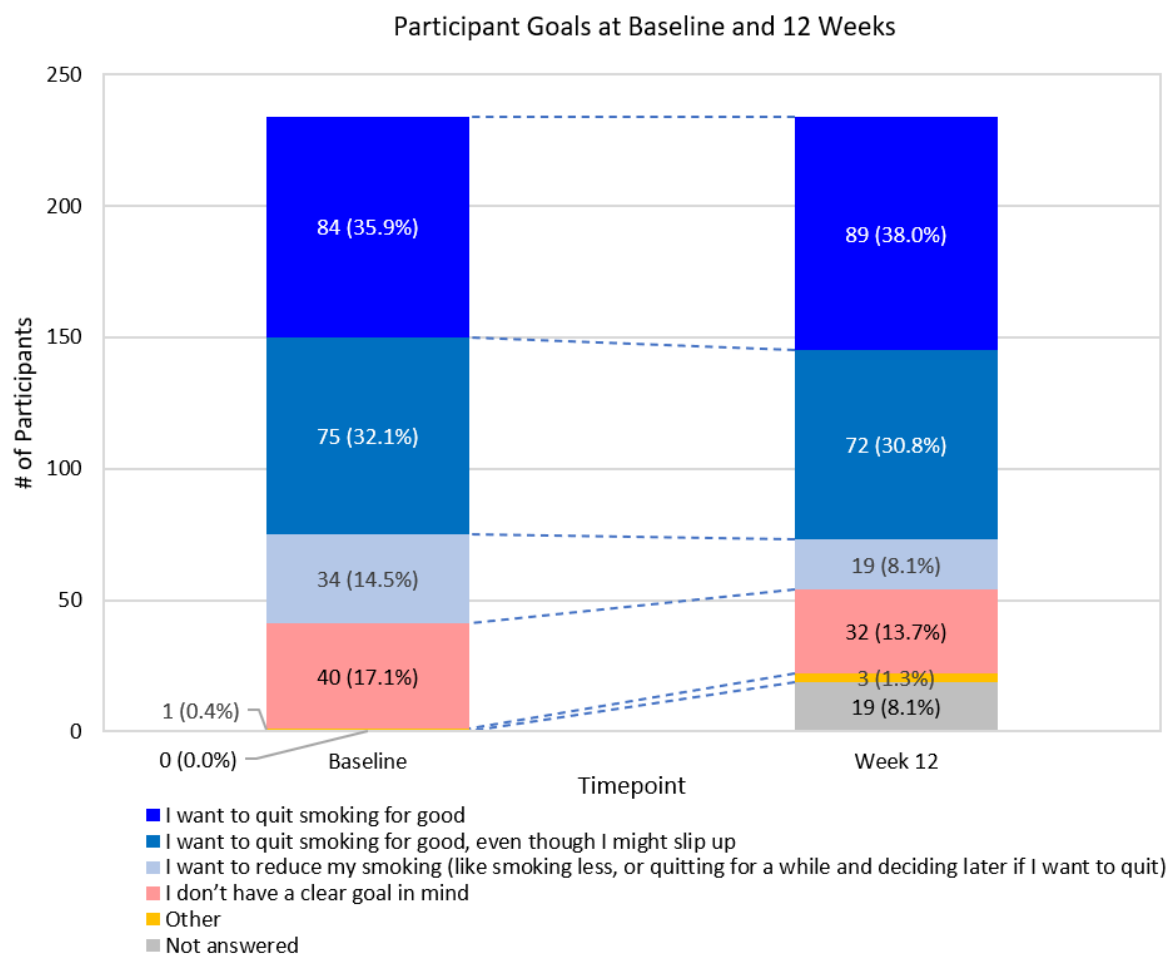
to change (5.6 vs 6.0;  $P=.07$ ). By contrast, the estimated ratings for success to quit and difficulty of quitting were greater than the estimated baseline values at 3.3 versus 5.0 ( $P<.001$ ) and 2.8 versus 4.3 ( $P<.001$ ), respectively.

**Figure 3.** Attitudes towards quitting smoking, ratings (scale 1-10) at baseline vs. 12 weeks. Estimate of means and standard errors based on linear mixed model. Readiness to quit smoking (RTQ), Difficulty to quit smoking (DTQ), Success to quit smoking (STQ).

n=234

Participants were asked their goal as it relates to smoking. Overall, the proportion of the cohort in each goal category was stable at both timepoints, with a slight increase in the proportion indicating they wanted to quit for good and a slight decrease in the proportion indicating they want to reduce their smoking

([Figure 4](#)) at 12 weeks. Matched pair data are detailed in [Table 6](#). Excluding the 4 participants who answered *Other*, 28.0% (59/211) strengthened their goal toward quitting, 55.0% (116/211) maintained their goal, and 17.1% (36/211) weakened their goal.

**Figure 4.** Participant goals at baseline and 12 weeks.**Table 6.** Participant goals at baseline (rows) and 12 weeks (columns), match-paired analysis (N=215).

	Goals: 12 weeks					
	Quit smoking for good	Quit smoking for good, even though I might slip up	Reduce my smoking	No clear goal in mind	Other	Total
Goals: Baseline						
Quit smoking for good, n (%)	55 (25.6)	15 (7.0)	2 (0.9)	6 (2.8)	0 (0.0)	78 (36.3)
Quit smoking for good, even though I might slip up, n (%)	20 (9.3)	38 (17.7)	1 (0.5)	7 (3.3)	2 (0.9)	68 (31.6)
Reduce my smoking, n (%)	5 (2.3)	13 (6.0)	9 (4.2)	5 (2.3)	0 (0.0)	32 (14.9)
No clear goal in mind, n (%)	9 (4.2)	6 (2.8)	6 (2.8)	14 (6.5)	1 (0.5)	36 (16.7)
Other, n (%)	0 (0.0)	0 (0.0)	1 (0.5)	0 (0.0)	0 (0.0)	1 <sup>a</sup> (0.5)
Total, n (%)	89 (41.4)	72 (33.5)	19 (8.8)	32 (14.9)	3 <sup>b</sup> (1.4)	215 (100.0)

<sup>a</sup>Of the 234 participants who completed the baseline questionnaire, 1 person selected *Other* for their goal and wrote, "Smoke less at this time."

<sup>b</sup>Of the 215 participants who completed the 12-week questionnaire, 3 people selected *Other* for their goal and wrote: (1) "Already quit"; (2) "Some of these questions become irrelevant once you quit smoking"; (3) "Already have stopped."

## Smoking Behavior

For secondary endpoints, 28.2% (66/234, ITT; 95% CI 22.5%-34.4%) reported making 1 or more quit attempt, and 23.1% (54/234, ITT; 95% CI 17.8%-29.0%) reduced their CPD by 50% or more by 4 weeks.

Quit attempt and CPD reduction rates increased over time. At 12 weeks, 48.3% (113/234, ITT; 95% CI 41.7%-54.9%) reported making 1 or more quit attempt, with mean 2.4 (SD 9.1; CI 1.2-3.6) quit attempts per participant. Overall, CPD reduction of 50% or more occurred in 38.5% (90/234, ITT; 95% CI 32.2%-45.0%) of participants. Among the study completers

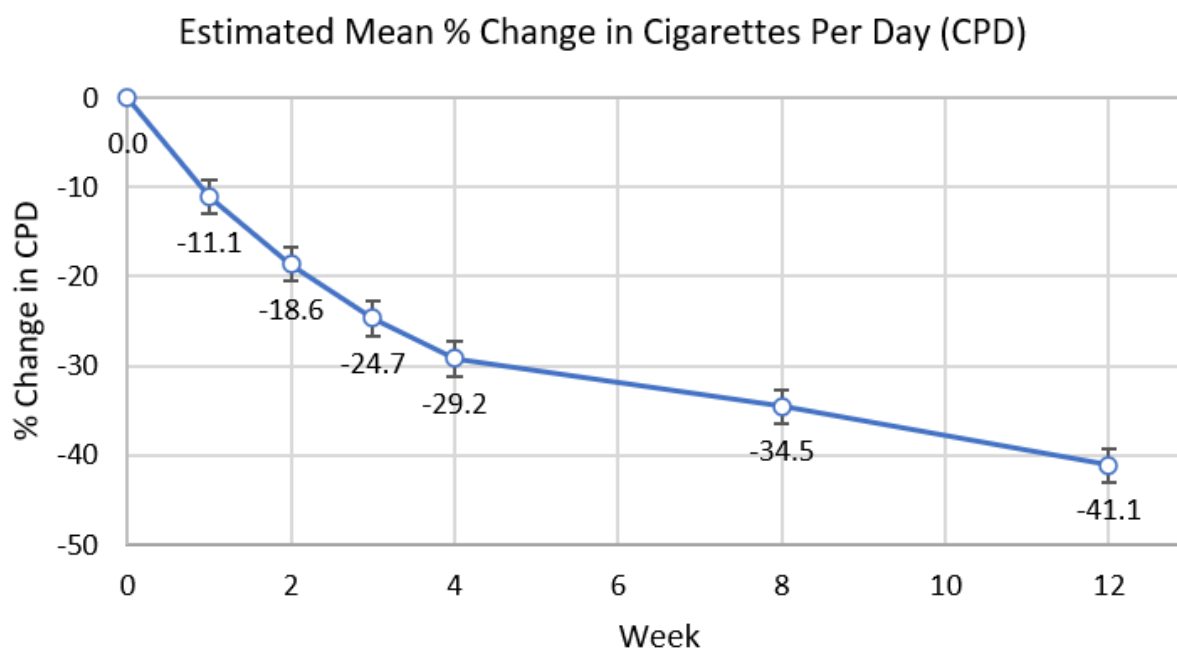


who did not achieve at least seven-day PPA, 33.2% (62/187, completer; 95% CI 26.5%-40.4%) reduced their CPD by 50% or more.

By 12 weeks, 82.8% (178/215) of participants had reduced CPD, 11.6% (25/215) had no change, and 5.6% (12/215) had increased CPD. Linear mixed model analysis was performed, with projected mean CPD values at various timepoints compared

with baseline and each other. Mean CPD decreased steadily over the course of the study, with the most pronounced drop over the first 4 weeks (Figure 5). At 12 weeks, mean CPD was reduced by 41.1% compared with baseline. Decreases in CPD were statistically significant at each timepoint (all  $P < .001$ , but  $P = .001$  for 3 weeks vs 4 weeks). Among the study completers who did not achieve at least seven-day PPA ( $n = 187$ ), CPD were reduced by 32.3% at 12 weeks.

**Figure 5.** Percent change in cigarettes per day (CPD) over time. Estimate of means and standard errors based on linear mixed model.

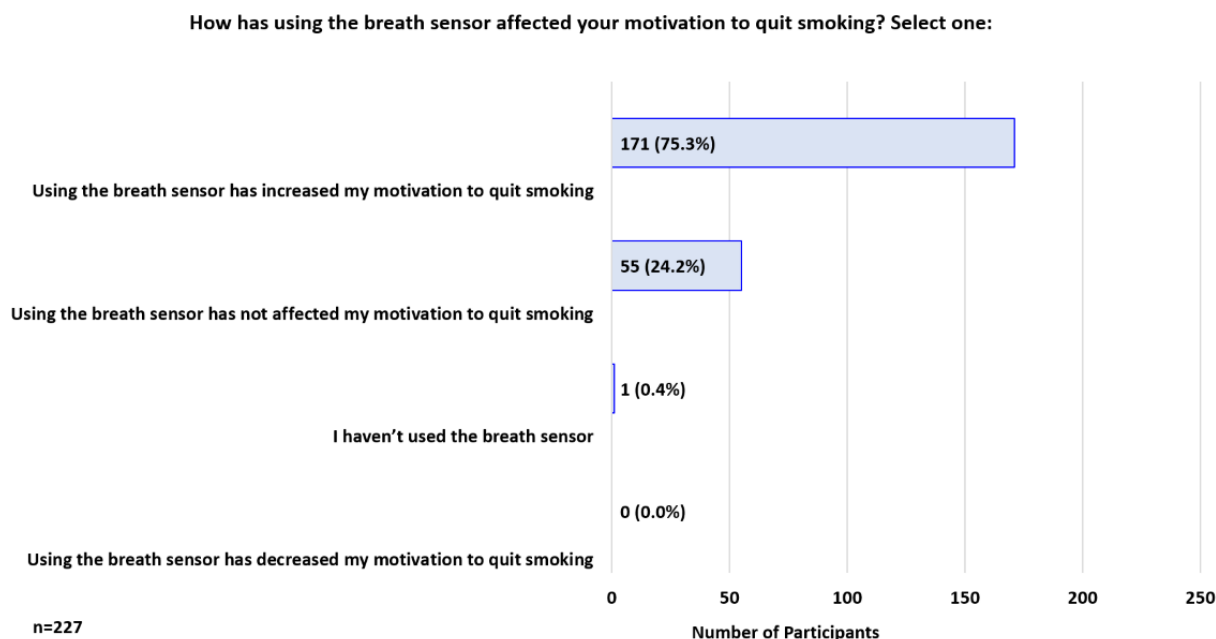
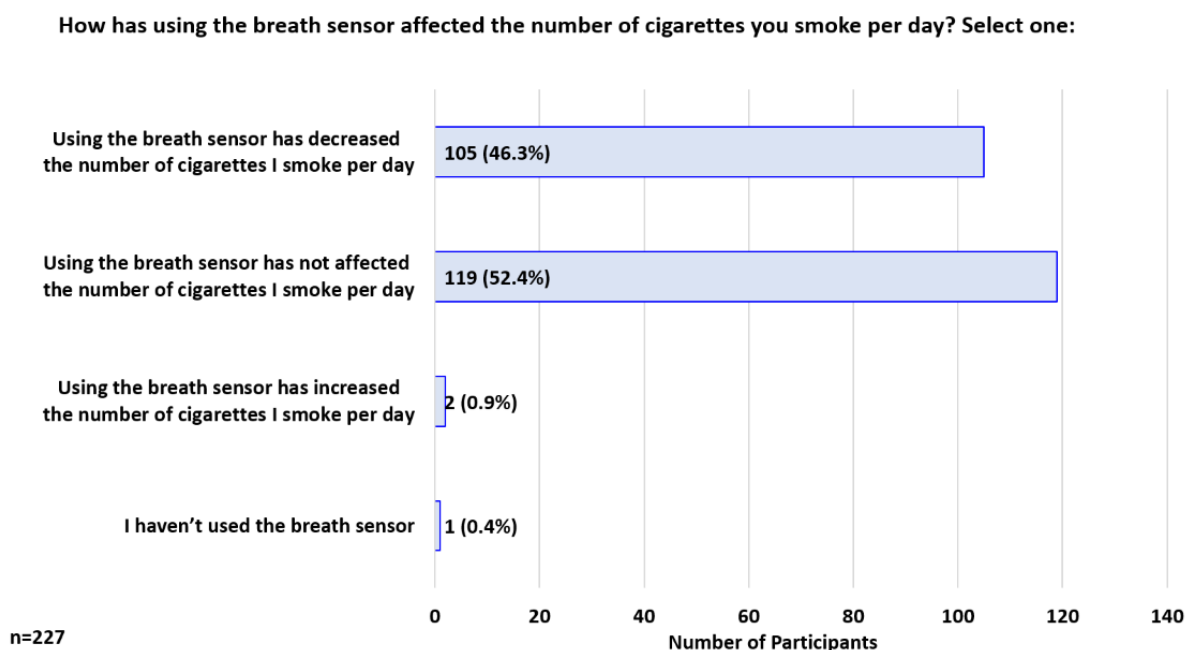


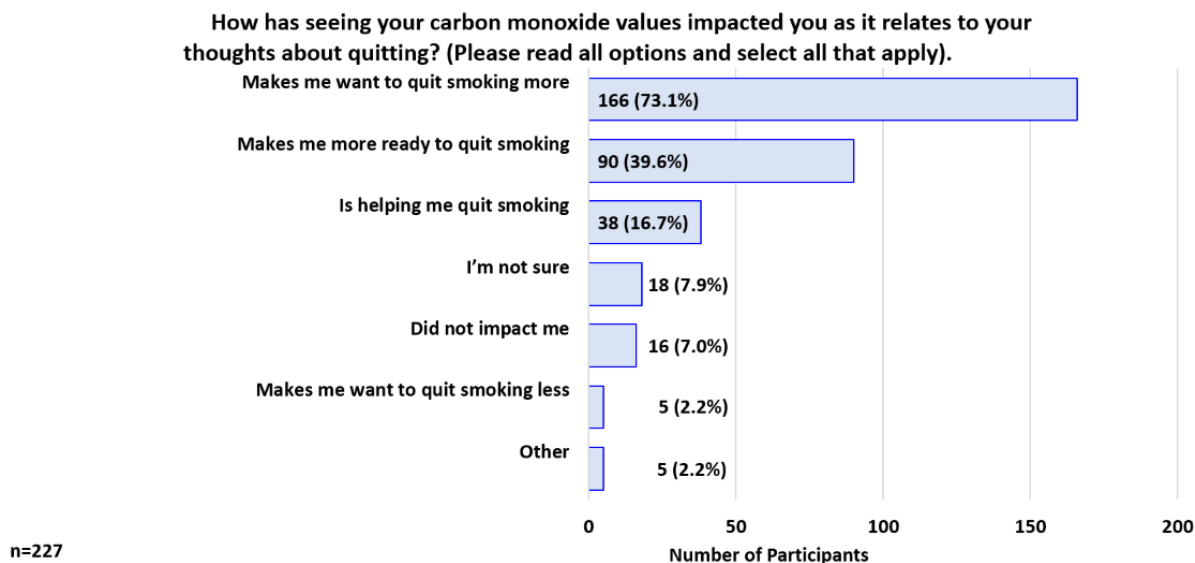
At 12 weeks, 7-day PPA was 12.0% (28/234, ITT) and 30-day PPA was 6.0% (14/234, ITT). Analysis of those who completed the study ( $n = 215$ ) yields a 7-day PPA of 13.0% (28/215, completer) and a 30-day PPA of 6.5% (14/215, completer).

### Use Experience

Over the course of the study, use experience was assessed via participant feedback on the impact of the breath sensor on attitudes toward quitting smoking and smoking behavior, and on learning associated with breath sensor use.

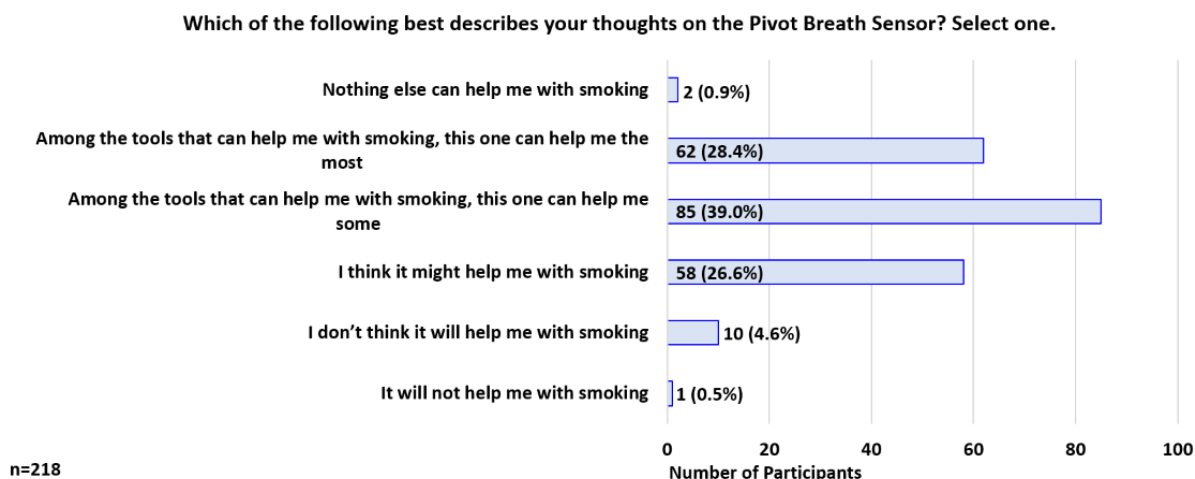
At 1 week, 75.3% (171/227) reported that using the breath sensor increased their motivation to quit smoking (Figure 6). When asked how using the breath sensor had affected the number of cigarettes smoked per day, 52.4% (119/227) responded it had not affected their CPD, while 46.3% (105/227) indicated it had decreased their CPD (Figure 7). Participants were asked how seeing their CO values had impacted their thoughts about quitting smoking; the top 3 responses were: "Makes me want to quit smoking more" 73.1% (166/227), "Makes me more ready to quit smoking" 39.6% (90/227), and "Is helping me quit smoking" 16.7% (38/227; Figure 8).

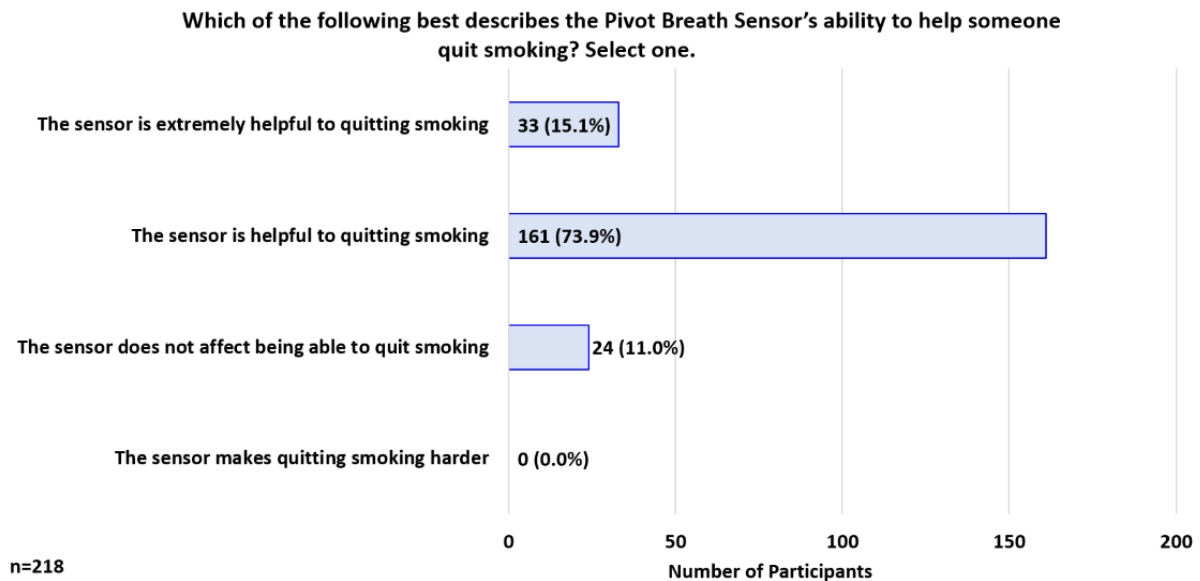
**Figure 6.** Participant Feedback: Effect of breath sensor on motivation to quit smoking (week 1).**Figure 7.** Participant Feedback: Effect of breath sensor on number of cigarettes smoked per day (week 1).

**Figure 8.** Participant Feedback: Impact of carbon monoxide values on thoughts about quitting (week 1).

At 2 weeks, when asked which statement best describes their thoughts on the breath sensor, 39.0% (85/218) indicated, “Among the tools that can help with smoking, this one can help me some” and 28.4% (62/218) indicated, “Among the tools that

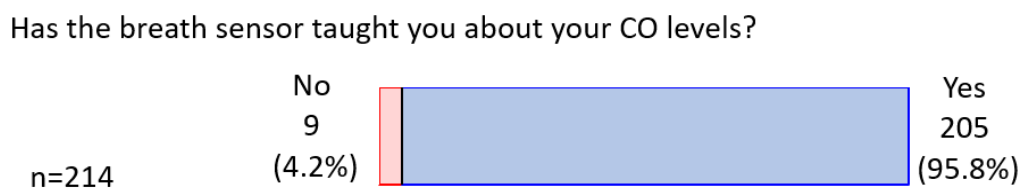
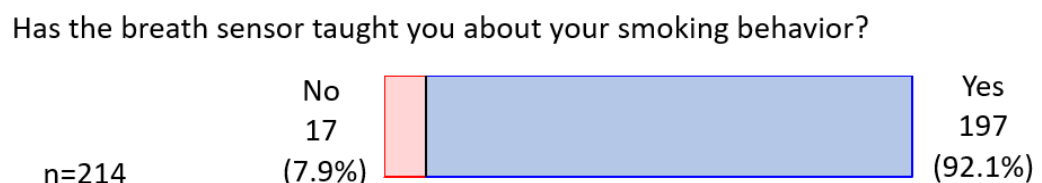
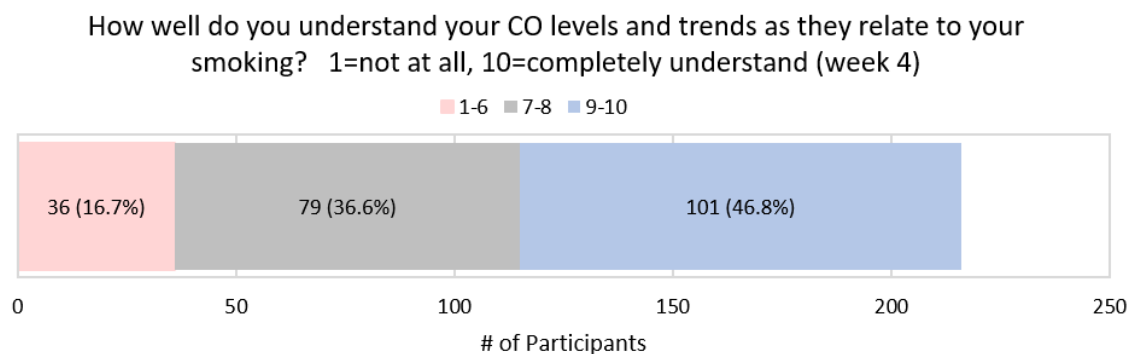
can help me with smoking, this one can help me the most” (Figure 9). In addition, 89.0% (194/218) indicated the sensor is *extremely helpful* or *helpful* in helping someone quit smoking (Figure 10).

**Figure 9.** Participant Feedback: Thoughts on the Pivot Breath Sensor (week 2).

**Figure 10.** Participant Feedback: Pivot Breath Sensor's ability to help someone quit smoking (week 2).

At 3 weeks, the majority (>90%, >196/214) of participants indicated the breath sensor had taught them about their CO levels and smoking behavior (Figures 11 and 12).

At 4 weeks, on a scale of 1-10, the mean score for how well participants understood their CO levels and trends as they relate to their smoking behavior was 8.0 (SD 2.1; Figure 13).

**Figure 11.** Participant Feedback: Has the breath sensor taught you about your carbon monoxide (CO) levels? (week 3).**Figure 12.** Participant Feedback: Has the breath sensor taught you about your smoking behavior? (week 3).**Figure 13.** Participant Feedback: Understanding of CO levels and trends as they relate to smoking behavior (week 4).

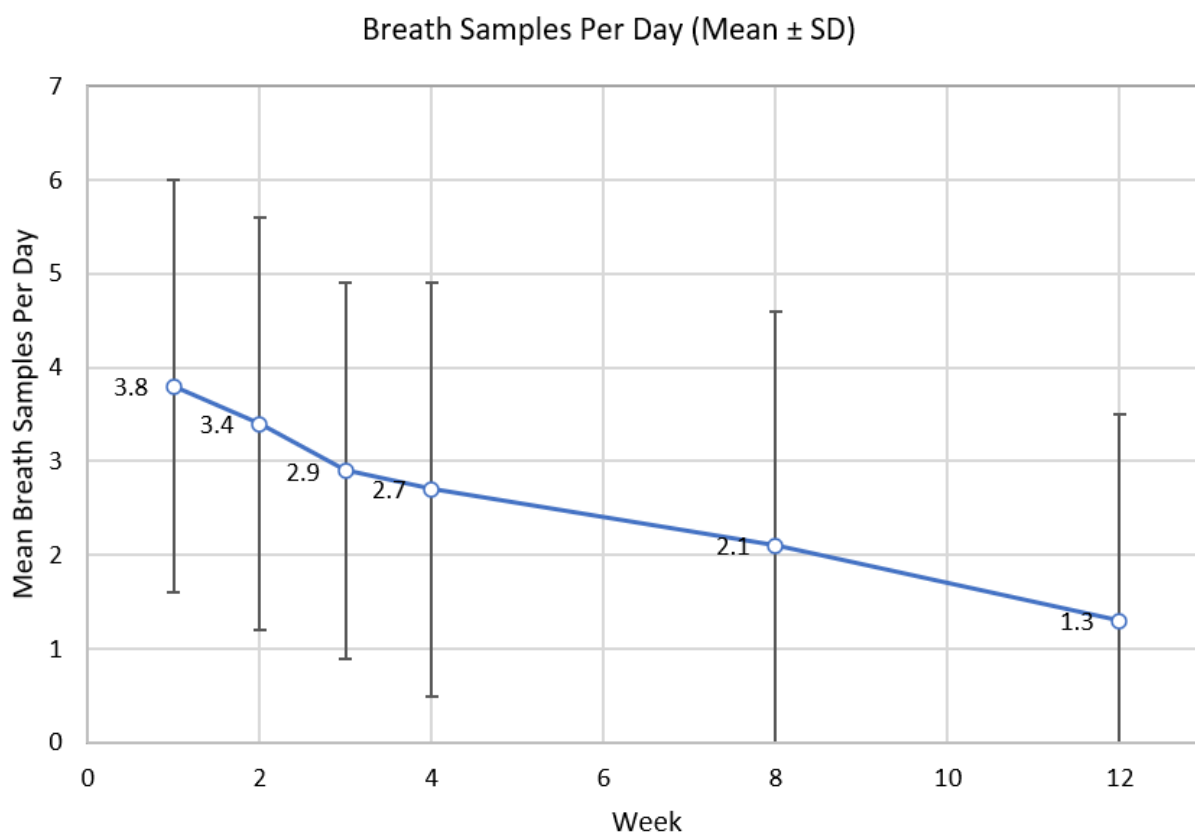
Assessment of participant breath sensor use revealed regular daily use that decreased over time, with a mean of 3.8 (SD 2.2) samples per day at 1 week compared with 1.3 (SD 2.2) at 12

weeks (Figure 14). Overall, a total of 48,747 breath samples were taken over the course of the study with each participant performing a mean of 208.3 (SD 141.8) total breath samples.

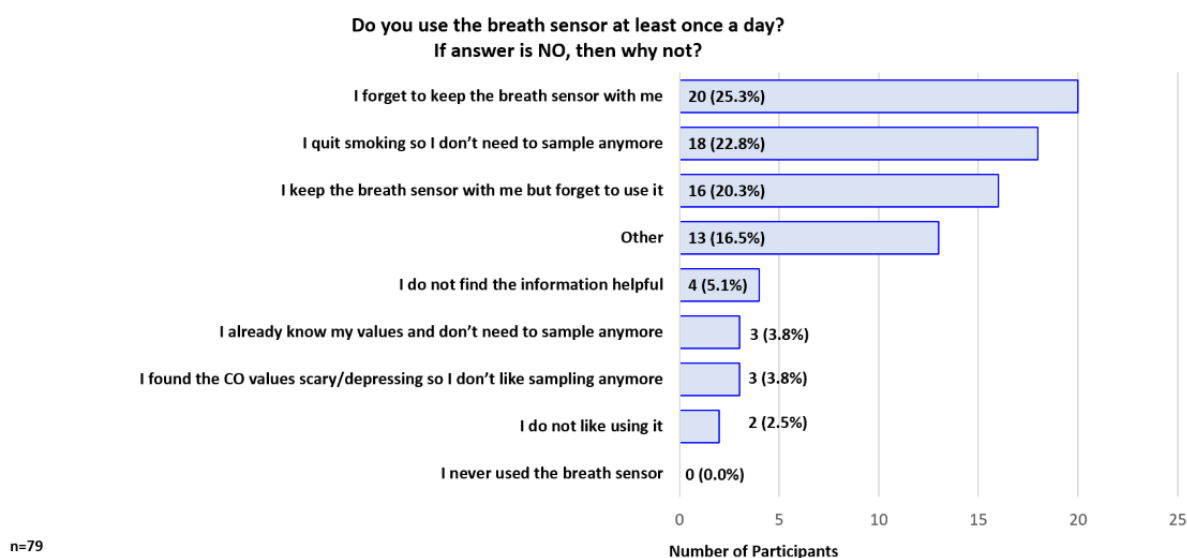
At 12 weeks, among the 79 participants who indicated they were not using the breath sensor at least once per day, the top 3 reasons were “I forget to keep the breath sensor with me”

(25% 20/79), “I quit smoking so I don’t need to sample anymore” (23%, 18/79), and “I keep the breath sensor with me but forget to use it” (20%, 16/79; Figure 15).

**Figure 14.** Breath sensor use.



**Figure 15.** Participant Feedback: Reasons for not using the breath sensor (week 12).



## Adverse Events

There were no adverse events.

## Discussion

### Principal Findings

This study evaluated attitudes toward quitting smoking, smoking behavior, and use experience in 234 adult smokers using the



Pivot Breath Sensor over a 12-week period. Participants had a significant increase in motivation to quit smoking as assessed through Stage of Change ( $P<.001$ ). Specifically, motivation to quit smoking increased in 29.6% (64/216) by week 4, and in 39.1% (84/215) by week 12. Positive changes in smoking behavior occurred as well; 28.2% (66/234) made 1 or more quit attempt and 23.1% (54/234) reduced their CPD by 50% or more at 4 weeks, which increased to 48.3% (113/234) and 38.5% (90/234), respectively, at 12 weeks. Among those who completed the 12-week questionnaire, 82.8% (178/215) had reduced CPD with an average CPD reduction of 41.1%. Moreover, 12.0% (28/234) achieved 7-day PPA and 6.0% (14/234) achieved 30-day PPA. Additional measures of attitudes toward quitting, specifically success to quit and perceived difficulty of quitting, had significant improvement ( $P<.001$ ). In assessing use experience, 75.3% (171/227) reported that using the breath sensor increased their motivation to quit smoking. Most participants (>90%, >196/214) indicated the breath sensor taught them about their CO levels and smoking behavior. A majority of participants (73.1%, 166/227) indicated that seeing their CO values made them want to quit smoking more.

### Evidence-Based Context of Outcomes

Assessing the results with consideration of available published data facilitates contextual interpretation of the study outcomes. Regarding the significance of increased motivation to quit smoking, Prochaska et al [28,37] reported that advancement of one stage in the Stage of Change assessment during the first month of treatment almost doubles the chances that a smoker will take effective action in the next 6 months. In addition, baseline Stage of Change predicts long-term quit rates [38,39]. For example, in an intensive action- and maintenance-oriented smoking cessation program for cardiac patients, validated abstinence from smoking at 6 months was achieved in 11% of those not thinking of quitting smoking at baseline, 27% of those thinking of quitting in the next 6 months at baseline, and 56% of those thinking of quitting in the next 30 days or actively making a quit attempt at baseline [40]. These data underpin the value of having high or recently increased motivation at the outset of a smoking cessation program. In this study, 38.9% (84/216) of study participants were seriously considering quitting in the next 30 days at 4 weeks (up from 14.4%, or 31/216, at study entry), and 29.6% (64/216) had increased motivation to quit over the first 4 weeks of the study; these individuals now have an increased likelihood of quitting smoking.

Concerning quit attempts, based on the median prevalence of 65.4% for past-year quit attempts among adult US smokers [29], the average monthly quit attempt rate is approximately 5%. In this study, 28.2% (66/234) of participants made a quit attempt over a 1-month period, a more than fivefold increase of the general population average. This quit attempt rate is further notable when considering that at baseline, 84.6% (198/234) of the study participants were not particularly motivated to quit, indicating they were seriously thinking about quitting in the next 6 months or not thinking about quitting smoking. Quit attempts are meaningful because increases in smoking cessation are driven in large part by increases in quit attempts [41]. Accordingly, the CDC and Healthy People initiative have

identified increasing quit attempts as an important goal for tobacco control efforts [5,42].

Finally, approximately 1% of the general population of smokers [30-34] and 2%-5% of individuals in cigarette reduction studies [35,36] will reduce their CPD by 50% or more on a monthly basis. In this study, 23.1% (54/234) of participants reduced their CPD by 50% or more over a 1-month period, increasing to 38.5% (90/234) at 12 weeks. Reducing CPD by 50% or more is clinically meaningful as the rates of quit attempts or cessation itself significantly increase among those who achieve this degree of CPD reduction [43].

### Comparison With Prior Work

Comparison of outcomes with previous studies is limited by differences in study design, particularly in the method and frequency of CO breath sampling, a constraint that compelled the undertaking of this study in the first place. In most previous assessments, CO breath sampling was administered by study staff or health professionals at study visits, and participants performed no more than a few breath sampling. In this study, participants used a personal interactive CO breath sensor to sample their breath multiple times per day over a 12-week period, with sampling done at the participant's discretion. Indeed, participants sampled extensively, with each performing an average of 208.3 breath samples.

The comparator study most similar in design to this study is Beard et al's investigation [16], in which participants used a personal CO breath sensor on an outpatient basis over a 6-week period [16]. Acknowledging that the study by Beard et al [16] was small ( $N=10$ ), the results of the 2 studies are in range of each other: CPD was reduced by 32.6% at 6 weeks (Beard et al [16]) and 34.5% at 8 weeks (this study) and quit attempts were made by 50.0% (5/10) of participants at 6 weeks (Beard et al [16]) and 38.5% (90/234) at 8 weeks (this study).

The approach to CO breath sampling in Beard et al's study [16] and this investigation enabled participants to directly link their smoking behavior to their CO values and track their progress over time. The benefit of tracking one's behavior and progress via self-guided biofeedback, evident here in smoking behavior, is also well documented in other disease states [8-10,44-46] lending further support to this approach.

### Limitations

There are a few important limitations of this study. First, while this study reports results from a long-term use period (12 weeks) of a personal CO breath sensor, it does not include outcomes following the period where the breath sensor was used. This limits the understanding of outcome durability and highlights the need for longer-term data.

Second, participants were compensated for breath sampling. This was deliberate in this initial attempt at understanding the impact of personal mobile CO breath sampling on adult smokers. The investigators opted for an optimized use scenario, to understand outcomes in the setting of reliable and steady breath sensor use. This may limit the generalizability of the results, particularly those addressing breath sampling behavior. Future research should address how individuals behave when not

incentivized to breath sample, and whether this real-world behavior yields results different from those reported herein. We did take steps to minimize the impact of compensation. First, we instituted a temporal delay between behavior and the associated payment (approximately 3 weeks). Second, payments were structured such that no individual payment was larger than US \$140. Moreover, while compensation was linked to the completion of breath sampling, it was not linked to outcomes such as attitudes, smoking behavior, or the content of participant feedback. The decrease in breath sampling from an average of 3.8 samples per participant per day at 1 week to 1.3 at 12 weeks suggests the study compensation did not unduly influence breath sampling behavior.

There are additional study design considerations to address as well. First, we did not require proof that breath samples were provided only by study participants. We believe this possibility is unlikely, as breath sampling over the course of the study largely followed the expected pattern of single-person use, with decreasing number of samples over time. When designing the study, we considered the drawbacks of implementing monitoring, including further differentiating sensor use in the study from real-world sensor use experience, decreasing autonomy and convenience for participants, and instilling a sense of policing that might have affected participant perception and experience of the sensor. Nonetheless, because we cannot exclude the possibility that someone other than the participant provided breath samples, we acknowledge this as a limitation.

In addition, it is important to consider that the study design, as a prospective cohort study, limits understanding of the influence

of baseline motivation to quit on participant outcomes. We do believe it is beneficial that the majority of study participants (76.9%, 180/234) were thinking of quitting in the next 6 months for 2 reasons: (1) This population had room for both observable improvement (thinking of quitting in the next 30 days) and worsening (not thinking of quitting) of motivation. (2) At this baseline level of motivation (thinking of quitting in the next 6 months), previous work indicates most were unlikely to change their smoking during the duration of the study [47]. Overall, this single-arm study was conducted as an initial assessment of personal mobile breath CO sampling in adult smokers. Now that these initial results have been established, future study via a randomized control trial is an appropriate next step. The aforementioned issues, particularly those of duration of follow-up after the period of breath sensor use and better understanding the role of baseline motivation to quit, should be addressed in any future research.

## Conclusion

In this study, smokers who used the Pivot Breath Sensor over a 12-week period had increased motivation to quit, reduced CPD, and had favorable quit attempt rates. These are meaningful milestones in the process of smoking cessation, conferring increased likelihood of success. Accordingly, the results suggest a role for personal biofeedback via mobile CO breath sampling in smoking cessation, particularly as a means to facilitate motivational advancement and favorable change in smoking behavior.

## Acknowledgments

The authors thank Lydia Tesfamariam for assistance with recruiting, collaborative tracking of participants, and for executing study payments. JG's work is supported, in part, by National Institutes of Health (NIH) Grant P30 DK034987.

## Authors' Contributions

JM, CF, KW, DB, and DU designed the study. KW oversaw participant recruiting. KW and JM ran the study. KW performed the administrative functions of study oversight. CF managed the database, conducted data analyses, and created the figures used in the manuscript. JG provided input on the statistical design of the study and conducted statistical data analyses. JM prepared the original draft of the manuscript. DU, CF, DB, and JG reviewed and edited the manuscript prior to submission.

## Conflicts of Interest

JM, CF, KW, and DB are employees of Carrot Inc, the developer of the Pivot Breath Sensor used in this study. They receive salary and stock options from Carrot Inc. DU is the President and CEO of Carrot Inc and an investor in the company.

## References

1. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention (US); 2014.
2. Lortet-Tieulent J, Goding Sauer A, Siegel RL, Miller KD, Islami F, Fedewa SA, et al. State-Level Cancer Mortality Attributable to Cigarette Smoking in the United States. *JAMA Intern Med* 2016 Dec 01;176(12):1792-1798. [doi: [10.1001/jamainternmed.2016.6530](https://doi.org/10.1001/jamainternmed.2016.6530)] [Medline: [27775761](https://pubmed.ncbi.nlm.nih.gov/27775761/)]
3. Office of the Surgeon General. Health Consequences of Smoking, Surgeon General Fact Sheet. URL: <https://www.hhs.gov/surgeongeneral/reports-and-publications/tobacco/consequences-smoking-factsheet/index.html> [accessed 2020-09-16]
4. Creamer MR, Wang TW, Babb S, Cullen KA, Day H, Willis G, et al. Tobacco Product Use and Cessation Indicators Among Adults - United States, 2018. *MMWR Morb Mortal Wkly Rep* 2019 Nov 15;68(45):1013-1019 [FREE Full text] [doi: [10.15585/mmwr.mm6845a2](https://doi.org/10.15585/mmwr.mm6845a2)] [Medline: [31725711](https://pubmed.ncbi.nlm.nih.gov/31725711/)]

5. Office of Disease Prevention and Health Promotion. Tobacco Use. URL: <https://www.healthypeople.gov/2020/topics-objectives/topic/tobacco-use/objectives> [accessed 2020-09-15]
6. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting Smoking Among Adults - United States, 2000-2015. *MMWR Morb Mortal Wkly Rep* 2017 Jan 06;65(52):1457-1464 [FREE Full text] [doi: [10.15585/mmwr.mm6552a1](https://doi.org/10.15585/mmwr.mm6552a1)] [Medline: [28056007](https://pubmed.ncbi.nlm.nih.gov/28056007/)]
7. Jamal A, King BA, Neff LJ, Whitmill J, Babb SD, Graffunder CM. Current Cigarette Smoking Among Adults - United States, 2005-2015. *MMWR Morb Mortal Wkly Rep* 2016 Nov 11;65(44):1205-1211 [FREE Full text] [doi: [10.15585/mmwr.mm6544a2](https://doi.org/10.15585/mmwr.mm6544a2)] [Medline: [27832052](https://pubmed.ncbi.nlm.nih.gov/27832052/)]
8. Cheatham SW, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. *J Sports Med Phys Fitness* 2018 Apr;58(4):534-548. [doi: [10.23736/S0022-4707.17.07437-0](https://doi.org/10.23736/S0022-4707.17.07437-0)] [Medline: [28488834](https://pubmed.ncbi.nlm.nih.gov/28488834/)]
9. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. *JMIR Mhealth Uhealth* 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: [10.2196/11819](https://doi.org/10.2196/11819)] [Medline: [30977740](https://pubmed.ncbi.nlm.nih.gov/30977740/)]
10. Franssen WMA, Franssen GHLM, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. *Int J Behav Nutr Phys Act* 2020 May 11;17(1):57 [FREE Full text] [doi: [10.1186/s12966-020-00955-2](https://doi.org/10.1186/s12966-020-00955-2)] [Medline: [32393357](https://pubmed.ncbi.nlm.nih.gov/32393357/)]
11. George J, MacDonald T. Home Blood Pressure Monitoring. *Eur Cardiol* 2015 Dec;10(2):95-101 [FREE Full text] [doi: [10.15420/ecr.2015.10.2.95](https://doi.org/10.15420/ecr.2015.10.2.95)] [Medline: [30310433](https://pubmed.ncbi.nlm.nih.gov/30310433/)]
12. Bray EP, Holder R, Mant J, McManus RJ. Does self-monitoring reduce blood pressure? Meta-analysis with meta-regression of randomized controlled trials. *Annals of Medicine* 2010 May 27;42(5):371-386 [FREE Full text] [doi: [10.3109/07853890.2010.489567](https://doi.org/10.3109/07853890.2010.489567)]
13. Yeh H, Brown TT, Maruthur N, Ranasinghe P, Berger Z, Suh YD, et al. Comparative effectiveness and safety of methods of insulin delivery and glucose monitoring for diabetes mellitus: a systematic review and meta-analysis. *Ann Intern Med* 2012 Sep 04;157(5):336-347. [doi: [10.7326/0003-4819-157-5-201209040-00508](https://doi.org/10.7326/0003-4819-157-5-201209040-00508)] [Medline: [22777524](https://pubmed.ncbi.nlm.nih.gov/22777524/)]
14. Sandberg A, Sköld CM, Grunewald J, Eklund A, Wheelock M. Assessing Recent Smoking Status by Measuring Exhaled Carbon Monoxide Levels. *PLoS ONE* 2011 Dec 16;6(12):e28864. [doi: [10.1371/journal.pone.0028864](https://doi.org/10.1371/journal.pone.0028864)] [Medline: [22194931](https://pubmed.ncbi.nlm.nih.gov/22194931/)]
15. Abrams DB, Niaura R, Brown RA, Emmons KM, Goldstein MG, Monti PM. *The Tobacco Dependence Treatment Handbook: A Guide to Best Practices*. New York: Guilford Press; 2008.
16. Beard E, West R. Pilot Study of the Use of Personal Carbon Monoxide Monitoring to Achieve Radical Smoking Reduction. *J. Smok Cessat* 2012 Jul 13;7(1):12-17. [doi: [10.1017/jsc.2012.1](https://doi.org/10.1017/jsc.2012.1)]
17. Patrick H, Fujii CA, Glaser DB, Utley DS, Marler JD. A Comprehensive Digital Program for Smoking Cessation: Assessing Feasibility in a Single-Group Cohort Study. *JMIR Mhealth Uhealth* 2018 Dec 18;6(12):e11708 [FREE Full text] [doi: [10.2196/11708](https://doi.org/10.2196/11708)] [Medline: [30563807](https://pubmed.ncbi.nlm.nih.gov/30563807/)]
18. Marler JD, Fujii CA, Utley DS, Tesfamariam LJ, Galanko JA, Patrick H. Initial Assessment of a Comprehensive Digital Smoking Cessation Program That Incorporates a Mobile App, Breath Sensor, and Coaching: Cohort Study. *JMIR Mhealth Uhealth* 2019 Feb 04;7(2):e12609 [FREE Full text] [doi: [10.2196/12609](https://doi.org/10.2196/12609)] [Medline: [30670372](https://pubmed.ncbi.nlm.nih.gov/30670372/)]
19. Foulds J, Veldheer S, Hrabovsky S, Yingst J, Sciamanna C, Chen G, et al. The effect of motivational lung age feedback on short-term quit rates in smokers seeking intensive group treatment: A randomized controlled pilot study. *Drug Alcohol Depend* 2015 Aug 01;153:271-277 [FREE Full text] [doi: [10.1016/j.drugalcdep.2015.05.007](https://doi.org/10.1016/j.drugalcdep.2015.05.007)] [Medline: [26051163](https://pubmed.ncbi.nlm.nih.gov/26051163/)]
20. Hajek P, West R, Lee A, Foulds J, Owen L, Eiser JR, et al. Randomized controlled trial of a midwife-delivered brief smoking cessation intervention in pregnancy. *Addiction* 2001 Mar;96(3):485-494. [doi: [10.1046/j.1360-0443.2001.96348511.x](https://doi.org/10.1046/j.1360-0443.2001.96348511.x)] [Medline: [11255587](https://pubmed.ncbi.nlm.nih.gov/11255587/)]
21. Shahab L, West R, McNeill A. A randomized, controlled trial of adding expired carbon monoxide feedback to brief stop smoking advice: evaluation of cognitive and behavioral effects. *Health Psychol* 2011 Jan;30(1):49-57. [doi: [10.1037/a0021821](https://doi.org/10.1037/a0021821)] [Medline: [21299294](https://pubmed.ncbi.nlm.nih.gov/21299294/)]
22. McClure JB, Ludman E, Grothaus L, Pabiniak C, Richards J, Mohelnitzky A. Immediate and short-term impact of a brief motivational smoking intervention using a biomedical risk assessment: the Get PHIT trial. *Nicotine Tob Res* 2009 Apr;11(4):394-403 [FREE Full text] [doi: [10.1093/ntr/ntp004](https://doi.org/10.1093/ntr/ntp004)] [Medline: [19299409](https://pubmed.ncbi.nlm.nih.gov/19299409/)]
23. McClure JB, Ludman EJ, Grothaus L, Pabiniak C, Richards J. Impact of a brief motivational smoking cessation intervention the Get PHIT randomized controlled trial. *Am J Prev Med* 2009 Aug;37(2):116-123 [FREE Full text] [doi: [10.1016/j.amepre.2009.03.018](https://doi.org/10.1016/j.amepre.2009.03.018)] [Medline: [19524389](https://pubmed.ncbi.nlm.nih.gov/19524389/)]
24. Choi W, Kim C, Lee O. Effects of brief smoking cessation education with expiratory carbon monoxide measurement on level of motivation to quit smoking. *Korean J Fam Med* 2013 May;34(3):190-198 [FREE Full text] [doi: [10.4082/kjfm.2013.34.3.190](https://doi.org/10.4082/kjfm.2013.34.3.190)] [Medline: [23730486](https://pubmed.ncbi.nlm.nih.gov/23730486/)]
25. Borrelli B, Novak S, Hecht J, Emmons K, Papandonatos G, Abrams D. Home health care nurses as a new channel for smoking cessation treatment: outcomes from project CARES (Community-nurse Assisted Research and Education on Smoking). *Prev Med* 2005;41(5-6):815-821. [doi: [10.1016/j.ypmed.2005.08.004](https://doi.org/10.1016/j.ypmed.2005.08.004)] [Medline: [16182355](https://pubmed.ncbi.nlm.nih.gov/16182355/)]



26. Bureau of Labor Statistics, US Department of Labor. The Employment Situation – July 2019. USDL-19-1388. 2019 Aug 2. URL: [accessed 2020-09-19]
27. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. *Nicotine Tob Res* 2002 May;4(2):149-159. [doi: [10.1080/14622200210123581](https://doi.org/10.1080/14622200210123581)] [Medline: [12028847](#)]
28. Prochaska J, Norcross J, Diclemente C. Applying the stages of change. *Psychother Aust* 2013;19:15. [doi: [10.1093/med:psych/9780199845491.003.0034](https://doi.org/10.1093/med:psych/9780199845491.003.0034)]
29. Walton K, Wang TW, Schauer GL, Hu S, McGruder HF, Jamal A, et al. State-Specific Prevalence of Quit Attempts Among Adult Cigarette Smokers - United States, 2011-2017. *MMWR Morb Mortal Wkly Rep* 2019 Jul 19;68(28):621-626 [FREE Full text] [doi: [10.15585/mmwr.mm6828a1](https://doi.org/10.15585/mmwr.mm6828a1)] [Medline: [31318853](#)]
30. Farkas AJ. When does cigarette fading increase the likelihood of future cessation? *Ann Behav Med* 1999;21(1):71-76. [doi: [10.1007/BF02895036](https://doi.org/10.1007/BF02895036)] [Medline: [18425657](#)]
31. Hughes JR, Cummings KM, Hyland A. Ability of smokers to reduce their smoking and its association with future smoking cessation. *Addiction* 1999 Jan;94(1):109-114. [doi: [10.1046/j.1360-0443.1999.9411097.x](https://doi.org/10.1046/j.1360-0443.1999.9411097.x)] [Medline: [10665102](#)]
32. Falba T, Jofre-Bonet M, Busch S, Duchovny N, Sindelar J. Reduction of quantity smoked predicts future cessation among older smokers. *Addiction* 2004 Jan;99(1):93-102. [doi: [10.1111/j.1360-0443.2004.00574.x](https://doi.org/10.1111/j.1360-0443.2004.00574.x)] [Medline: [14678067](#)]
33. Hyland A, Levy DT, Rezaishiraz H, Hughes JR, Bauer JE, Giovino GA, et al. Reduction in amount smoked predicts future cessation. *Psychol Addict Behav* 2005 Jun;19(2):221-225. [doi: [10.1037/0893-164X.19.2.221](https://doi.org/10.1037/0893-164X.19.2.221)] [Medline: [16011395](#)]
34. Broms U, Korhonen T, Kaprio J. Smoking reduction predicts cessation: longitudinal evidence from the Finnish adult twin cohort. *Nicotine Tob Res* 2008 Mar;10(3):423-427. [doi: [10.1080/14622200801888988](https://doi.org/10.1080/14622200801888988)] [Medline: [18324560](#)]
35. Joseph AM, Hecht SS, Murphy SE, Lando H, Carmella SG, Gross M, et al. Smoking reduction fails to improve clinical and biological markers of cardiac disease: a randomized controlled trial. *Nicotine Tob Res* 2008 Mar;10(3):471-481 [FREE Full text] [doi: [10.1080/14622200801901948](https://doi.org/10.1080/14622200801901948)] [Medline: [18324566](#)]
36. Glasgow RE, Gaglio B, Estabrooks PA, Marcus AC, Ritzwoller DP, Smith TL, et al. Long-term results of a smoking reduction program. *Med Care* 2009 Jan;47(1):115-120. [doi: [10.1097/MLR.0b013e31817e18d1](https://doi.org/10.1097/MLR.0b013e31817e18d1)] [Medline: [19106739](#)]
37. Prochaska JO, Velicer WF, Prochaska JM, Johnson JL. Size, consistency, and stability of stage effects for smoking cessation. *Addict Behav* 2004 Jan;29(1):207-213. [doi: [10.1016/s0306-4603\(03\)00086-8](https://doi.org/10.1016/s0306-4603(03)00086-8)] [Medline: [14667431](#)]
38. Ferguson JA, Patten CA, Schroeder DR, Offord KP, Eberman KM, Hurt RD. Predictors of 6-month tobacco abstinence among 1224 cigarette smokers treated for nicotine dependence. *Addict Behav* 2003 Sep;28(7):1203-1218. [doi: [10.1016/s0306-4603\(02\)00260-5](https://doi.org/10.1016/s0306-4603(02)00260-5)] [Medline: [12915164](#)]
39. Chouinard M, Robichaud-Ekstrand S. Predictive value of the transtheoretical model to smoking cessation in hospitalized patients with cardiovascular disease. *Eur J Cardiovasc Prev Rehabil* 2007 Feb;14(1):51-58. [doi: [10.1097/HJR.0b013e328014027b](https://doi.org/10.1097/HJR.0b013e328014027b)] [Medline: [17301627](#)]
40. Ockene J, Kristeller JL, Goldberg R, Ockene I, Merriam P, Barrett S, et al. Smoking cessation and severity of disease: the Coronary Artery Smoking Intervention Study. *Health Psychol* 1992;11(2):119-126. [doi: [10.1037//0278-6133.11.2.119](https://doi.org/10.1037//0278-6133.11.2.119)] [Medline: [1582380](#)]
41. Zhu S, Lee M, Zhuang Y, Gamst A, Wolfson T. Interventions to increase smoking cessation at the population level: how much progress has been made in the last two decades? *Tob Control* 2012 Mar;21(2):110-118 [FREE Full text] [doi: [10.1136/tobaccocontrol-2011-050371](https://doi.org/10.1136/tobaccocontrol-2011-050371)] [Medline: [22345233](#)]
42. Centers for Disease Control and Prevention. Best Practices for Comprehensive Tobacco Control Programs—2014. In: Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Atlanta: Centers for Disease Control and Prevention; 2014:1-141.
43. Hughes JR. Reduced smoking: an introduction and review of the evidence. *Addiction* 2000 Jan;95 Suppl 1:S3-S7. [Medline: [10723815](#)]
44. Gualtieri L, Rosenbluth S, Phillips J. Can a Free Wearable Activity Tracker Change Behavior? The Impact of Trackers on Adults in a Physician-Led Wellness Group. *JMIR Res Protoc* 2016 Nov 30;5(4):e237 [FREE Full text] [doi: [10.2196/resprot.6534](https://doi.org/10.2196/resprot.6534)] [Medline: [27903490](#)]
45. Yingling L, Allen NA, Litchman ML, Colicchio V, Gibson BS. An Evaluation of Digital Health Tools for Diabetes Self-Management in Hispanic Adults: Exploratory Study. *JMIR Diabetes* 2019 Jul 16;4(3):e12936 [FREE Full text] [doi: [10.2196/12936](https://doi.org/10.2196/12936)] [Medline: [31313657](#)]
46. Alharbi M, Straiton N, Gallagher R. Harnessing the Potential of Wearable Activity Trackers for Heart Failure Self-Care. *Curr Heart Fail Rep* 2017 Dec;14(1):23-29. [doi: [10.1007/s11897-017-0318-z](https://doi.org/10.1007/s11897-017-0318-z)] [Medline: [28181075](#)]
47. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. *Am Psychol* 1992 Sep;47(9):1102-1114. [Medline: [1329589](#)]

## Abbreviations

**CO:** carbon monoxide

**CPD:** cigarettes per day

**ITT:** intention to treat

**PPA:** point prevalence abstinence

**ppm:** parts per million

*Edited by G Eysenbach; submitted 23.07.20; peer-reviewed by D Goyal; comments to author 17.08.20; revised version received 26.08.20; accepted 07.09.20; published 02.10.20.*

*Please cite as:*

*Marler JD, Fujii CA, Wong KS, Galanko JA, Balbierz DJ, Utley DS*

*Assessment of a Personal Interactive Carbon Monoxide Breath Sensor in People Who Smoke Cigarettes: Single-Arm Cohort Study*  
*J Med Internet Res* 2020;22(10):e22811

URL: <https://www.jmir.org/2020/10/e22811>

doi: [10.2196/22811](https://doi.org/10.2196/22811)

PMID: [32894829](https://pubmed.ncbi.nlm.nih.gov/32894829/)

©Jennifer D Marler, Craig A Fujii, Kristine S Wong, Joseph A Galanko, Daniel J Balbierz, David S Utley. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 02.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Predictors and Effects of Usage of an Online Mindfulness Intervention for Distressed Cancer Patients: Usability Study

Linda Cillessen<sup>1,2</sup>, MSc; Monique OM van de Ven<sup>3</sup>, PhD; Félix R Compen<sup>1,2</sup>, PhD; Else M Bisseling<sup>1,2</sup>, MD, PhD; Marije L van der Lee<sup>4</sup>, PhD; Anne EM Speckens<sup>1,2</sup>, MD, PhD

<sup>1</sup>Department of Psychiatry, Center for Mindfulness, Radboud University Medical Center, Nijmegen, Netherlands

<sup>2</sup>Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, Netherlands

<sup>3</sup>Department of Medical Psychology, Laurentius Hospital, Roermond, Netherlands

<sup>4</sup>Scientific Research Department, Centre for Psycho-Oncology, Helen Dowling Institute, Bilthoven, Netherlands

**Corresponding Author:**

Linda Cillessen, MSc

Department of Psychiatry

Center for Mindfulness

Radboud University Medical Center

Postbus 9101

Nijmegen, 6500 HB

Netherlands

Phone: 31 (0)24 361 5445

Email: [linda.cillessen@radboudumc.nl](mailto:linda.cillessen@radboudumc.nl)

## Abstract

**Background:** One in three cancer patients experience high psychological distress. Mindfulness-based interventions are effective in reducing psychological distress in this patient group. However, these interventions lack availability and flexibility, which may compromise participation in the intervention for cancer patients experiencing late symptoms like fatigue or pain. Therefore, mindfulness-based interventions are increasingly offered via the internet. However, little is known about the usage of these online mindfulness-based interventions.

**Objective:** The aim of this study was to (1) predict uptake of and adherence to online mindfulness-based cognitive therapy (eMBCT) using baseline patient characteristics (demographic, cancer-related, personality, and psychological variables) and (2) examine the relations between adherence and treatment outcomes in eMBCT for cancer patients.

**Methods:** A total of 125 cancer patients were assigned to eMBCT in a parent randomized controlled trial comparing MBCT and eMBCT with treatment as usual in distressed cancer patients. Various usage measures of eMBCT were automatically tracked within the online program. Based on activity of use, participants were classified as nonusers, minimal users, low users, and intended users. Questionnaires were used to assess baseline characteristics (preintervention) and outcomes (pre- and postintervention). To answer the research questions, data were analyzed with t tests,  $\chi^2$  tests, and linear regression models.

**Results:** Based on weekly activity, participants were classified as nonusers (n=17, 13.6%), who completed no exercises in MBCT; minimal users (n=31, 24.8%), who completed at least one exercise of one to three sessions; low users (n=12, 9.6%), who completed at least one exercise of four to seven sessions; and intended users (n=65, 52.0%), who completed at least one exercise of eight to nine sessions. Nonusers had more fear of cancer recurrence at baseline than users (uptake), and intended users were more conscientious than minimal and low users (adherence). Intended users reported a larger reduction in psychological distress and more improvement of positive mental health (ie, emotional, psychological, and social well-being) after the intervention than other participants.

**Conclusions:** This study showed that adherence was related to improved patient outcomes. Patients with strong fear of recurrence or low levels of conscientiousness should receive extra attention, as they are less likely to respectively start or complete eMBCT. Future research may focus on the development of flexible and adaptive eMBCT programs to fit individual needs.

(*J Med Internet Res* 2020;22(10):e17526) doi:[10.2196/17526](https://doi.org/10.2196/17526)

**KEYWORDS**

internet intervention; eHealth; mindfulness; mindfulness-based cognitive therapy; usage; log data; uptake; adherence; cancer; oncology

**Introduction**

About one-third of cancer patients and survivors experience high psychological distress due to symptoms of anxiety and depression [1,2]. Mindfulness-based interventions (MBIs) can help to reduce psychological distress in cancer patients, as shown in several large meta-analyses [3,4]. Mindfulness is defined as “paying attention in a particular way: on purpose, in the present moment, and nonjudgmentally” [5]. The two most widely used MBIs (mindfulness-based stress reduction [MBSR] and mindfulness-based cognitive therapy [MBCT]) are usually delivered as group interventions with eight weekly group meetings and a silent day [6,7]. All meetings include meditation exercises (bodyscan, sitting meditations, and gentle movements), psychoeducation, and group discussions. Participants are instructed to practice meditation on a daily basis to cultivate their mindfulness skills.

Despite its beneficial effects, not all distressed cancer patients and survivors are able to participate in a regular face-to-face MBI. As a high proportion of patients have physical problems like severe fatigue [8] or pain [9,10], traveling on a weekly basis to a lengthy meeting at a fixed date and time might be impossible. Therefore, researchers started to investigate whether online MBIs, with their higher accessibility and flexibility, have similar effects as regular face-to-face group MBIs. These online MBIs reduce travel costs and efforts, have 24/7 availability, and may avoid waitlists [11]. Furthermore, participants are able to practice at a location and time they prefer, and at their own pace [12,13].

A recent meta-analysis of online MBIs in various populations showed moderate effects on stress and small effects on depression and anxiety [14]. Focusing on cancer patients, studies on online MBIs revealed promising results. For instance, Zernicke et al [15] found that online mindfulness-based cancer recovery was feasible and reduced symptoms of mood disturbance and stress in a randomized pilot study. In an actively controlled study, online MBCT (eMBCT) reduced fatigue severity in cancer patients [16]. Furthermore, our own trial (the BeMind project) showed that cancer patients reported a moderate reduction in psychological distress after participating in individual eMBCT in comparison with those receiving treatment as usual [17]. Over the course of a 9-month follow-up, eMBCT resulted in an even greater reduction in psychological distress than the classical group-based MBCT [18]. In terms of cost-effectiveness, both treatments were equally cost-effective compared with treatment as usual [19].

Despite growing evidence on the cost-effectiveness of online MBIs, less is known about usage of these online interventions. Usage is relevant to address, as it directly relates to intervention outcomes [20] and can help to optimize interventions [21]. Two relevant aspects of usage are uptake (ie, starting with the intervention) and adherence (ie, receiving an intended dose of the intervention) [22,23]. Uptake and adherence can be tracked

with log data from the intervention website. Commonly reported log measures are frequency of use (eg, number of logins), duration of use (eg, duration of each login), and activity (eg, number of completed exercises) [24]. These log measures can be tracked with great objectivity [24], although they only reflect online and not offline engagement with the intervention.

A meta-analysis on predictors of adherence in online interventions among adults found that women demonstrated greater adherence than men. Mixed findings were observed for the relation with age and severity of symptoms [22]. Personality also seems to affect adherence to online interventions. For instance, a study on a web-based occupational health intervention showed that lower levels of negative affectivity and impulsivity and higher levels of alexithymia correlated with less usage of the online intervention [25]. In adults with cancer, uptake was the highest among female patients, while older patients demonstrated a greater adherence to online cognitive behavioral therapy than younger patients [23]. Another study in adults with cancer participating in a web-based cognitive behavioral intervention found that different user groups did not differ in age, education level, or psychological distress [26]. As far as we know, predictors of adherence and its relationship to outcome of online MBIs for cancer patients have not yet been studied.

The aim of this study was to examine the usage of individual eMBCT for distressed cancer patients in relation to outcome. First, we explored whether various baseline patient characteristics (demographic, cancer-related, psychological, and personality variables) could predict uptake and adherence. Second, we tested whether adherence and separate usage measures (number of logins, total time logged in, mean time logged in, number of emails sent to the therapist, and number of assignments completed) were related to treatment outcome in terms of both psychological distress and positive mental health (ie, emotional, psychological, and social well-being). We expected that intended users would gain more from the intervention than low and minimal users.

**Methods****Study Design**

This study concerns secondary analyses of the data from the parent BeMind study, a multicenter randomized controlled trial (RCT) studying the effects of group face-to-face MBCT and eMBCT versus treatment as usual in distressed cancer patients [17,18,27]. This study only includes data of participants immediately randomized to eMBCT or those randomized to eMBCT after treatment as usual. All participants provided written informed consent prior to participation.

**Participants**

Distressed cancer patients were recruited through various online and offline media, and were directed to a study website, where they could self-enroll for the study. The inclusion criteria were

having any cancer diagnosis; experiencing at least mild psychological distress (a score of  $\geq 11$  on the Hospital Anxiety and Depression Scale [HADS] [28,29]); internet access and computer literacy; good command of the Dutch language; and willingness to participate in a mindfulness intervention. The exclusion criteria were severe psychiatric morbidity; change in psychotropic medication within 3 months prior to baseline; and current or previous participation in MBCT or MBSR. More details about the recruitment procedure can be found elsewhere [27].

## Intervention

The online MBCT intervention followed the protocol of Segal et al [30], with some adaptations to fit the needs of the target group, for instance, psychoeducation about grief and cancer-related fatigue. The intervention was designed to be completed in 9 weeks; however, the average time to complete the program was 10.4 weeks (SD 4.0 weeks). Each session was spent on a specific theme, for instance, automatic pilot, communication, or self-care. Participants were provided with information, audio files of guided meditation, and assignments around the theme of the session through a personal secure webpage. The assignments included, for instance, the recording of pleasant or unpleasant events, or how they experienced the meditation exercises. Participants were encouraged to read the information and perform the assigned meditation exercises and assignments within 1 week. The therapist provided feedback on a predetermined day of the week. All therapists had experience in psycho-oncology and were qualified mindfulness trainers according to the criteria of the UK Mindfulness-Based Teacher Network [31]. More details of eMBCT, including screenshots of the intervention, can be found elsewhere [32].

## Measures

### Log Data

We measured different aspects of usage with log data, which is recommended because different aspects reflect different types of usage [24]. Log data of eMBCT were retrieved from the study website. These included for each login, the time logging in and logging out and the number of assignments saved and submitted. It should be noted that after 30 minutes of inactivity participants were automatically logged out. Furthermore, all emails from participants to therapists were available. From the available data, the following measures were calculated: total time logged in, mean time logged in per login, number of logins (of at least 1 minute), number of completed assignments (which included both those saved by the user and those submitted to the therapist), and number of emails sent to the therapist.

Based on usage, participants were divided into usage groups, which is a common practice in this field [26]. We chose to create categories based on one aspect of activity, namely the amount of sessions in which at least one exercise had been completed, as frequency and duration may be more biased. Regarding frequency, some participants may write their experiences during meditation on paper and add them to the online program at a later moment. Regarding duration, we expected large variations in duration of use, as the ease and speed of writing may differ. Furthermore, activity may be more likely to reflect treatment

engagement compared with other usage measures [33]. As eMBCT is a complete program, in which each week provides new knowledge and skills training that builds on the previous week, intended usage was defined as completing at least one exercise of eight to nine sessions. Low usage was defined as completing at least half of the program (ie, four to seven sessions) [34,35]. Minimal users completed at least one exercise of one to three sessions, while nonusers did not complete any of the exercises. When focusing on uptake, nonusers were compared with users (minimal/low/intended), and when focusing on adherence, intended users were compared with low/minimal users.

### Baseline Characteristics

The following self-reported baseline characteristics were explored as possible predictors: gender, age, education level, cancer type (breast vs other), anticancer treatment intent (curative vs palliative), personality (openness, conscientiousness, extraversion, agreeableness, and neuroticism), baseline psychological distress, positive mental health, rumination, fear of cancer recurrence, and mindfulness skills. Sociodemographic characteristics and cancer-related variables were assessed in the baseline interview and via self-report questionnaires. Psychological predictors included baseline psychological distress (described below), baseline positive mental health (described below), rumination, fear of cancer recurrence, and mindfulness skills. Rumination was measured with the 12-item rumination subscale of the Rumination and Reflection Questionnaire (RRQ) [36]. Fear of cancer recurrence was measured with the nine-item Severity subscale of the Fear of Cancer Recurrence Inventory (FCRI) [37,38]. Mindfulness skills were measured with the Five Facet Mindfulness Questionnaire Short Form (FFMQ-SF), a 24-item self-report questionnaire [39]. Personality was assessed with the NEO Five Factor Inventory (NEO-FFI) [40]. This 60-item self-report questionnaire measures five personality characteristics (openness to experiences, conscientiousness, extraversion, agreeableness, and neuroticism).

### Outcome Measures

The outcome measures were self-reported psychological distress and positive mental health. Psychological distress was measured with the 14-item HADS (theoretical range 0-42), developed to measure depression and anxiety [28,29]. The HADS has adequate psychometric properties to detect distress in cancer patients [41,42]. Internal consistency in the present study was good (Cronbach  $\alpha$  at  $T_0=0.87$ ). Positive mental health was measured with the Mental Health Continuum Short Form (MHC-SF), a 14-item questionnaire measuring emotional, psychological, and social well-being (theoretical range 0-70) [43]. The MHC-SF has adequate psychometric properties [44]. Internal consistency was excellent ( $\alpha=.93$ ). Both measures were assessed before and after the intervention.

### Data Analysis

Data were analyzed with SPSS version 22 (IBM Corp). Statistical significance was determined at  $P<.05$  (two-sided). Descriptive statistics for participants and usage measures were calculated. With regard to the usage data, 16 participants (12.8%) did not login at all. For these patients, the number of

logins was recorded as zero. For the other usage measures, missing values were maintained, as recording them to zero would lead to bias (eg, the mean time logged-in over all participants would be artificially lowered).

Visual inspection of histograms revealed all measures of usage were normally distributed, except for the number of completed assignments, which had a negative skewness, indicating that a large proportion of participants saved all assignments (ceiling effect). For further analyses, the number of completed assignments was dichotomized with median split. A score of one (above the median) indicated that participants saved or submitted either all 58 assignments or all but one assignment (57 assignments), while a score of zero indicated less assignments were completed (<57 assignments). We refer to this variable as completed all assignments (yes/no). Number of logins, total time logged in, mean time logged in, number of emails sent, and number of exercises completed were calculated for each group. In the analyses, the low and minimal user groups were combined to create more equal group sizes.

The first research question focused on prediction of uptake (nonusers vs users) and adherence (intended vs minimal/low users). For each of the user groups, baseline patient characteristics (demographic, cancer-related, personality, and psychological variables) were described. Independent sample *t* tests and  $\chi^2$  tests were used to test for significant differences between the user groups regarding uptake or adherence.

To study the relationship between usage and outcome (research question 2), linear regression models were used. These analyses

only included participants who actually used the intervention (ie, minimal, low, and intended users). Separate models were run for usage group (minimal/low vs intended), and per usage measure (total time logged in, mean time logged in, number of log-ins, number of assignments completed, and number of emails sent to the therapist) and outcome measure (psychological distress and positive mental health). All models were controlled for baseline levels of outcome measures.

## Results

### Participants

In total, 125 patients with cancer participated in eMBCT. The mean age of the participants was 52 years (SD 10.2). Most participants were female (*n*=109, 87.2%), had breast cancer (*n*=76, 60.8%), and were treated with curative intent (*n*=102, 81.6%). The mean level of psychological distress on the HADS was 17 (SD 6.9).

Regarding the different measures of usage in the total group, participants logged in on average 30.5 times (SD 28.1), with a mean time logged in of 28.1 minutes (SD 19.1) and a total time logged in of 1066 minutes (SD 1217). Participants sent on average nine emails (SD 5.8) and completed most assignments (median 57.5, range 1-58). Seventeen participants (13.6%) were classified as nonusers, 31 (24.8%) as minimal users, 12 (9.6%) as low users, and 65 (52.0%) as intended users. Usage in each of the user groups is displayed in [Table 1](#).

**Table 1.** Descriptive statistics of usage measures in the user groups (N=125).

Variable	Nonusers (n=17)	Users (n=108)		
		Minimal users (n=31)	Low users (n=12)	Intended users (n=65)
Duration: average login time, mean (SD) <sup>a</sup>	3.5 (N/A) <sup>b,c</sup>	21.4 (17.9)	24.4 (13.1)	32.3 (19.6)
Duration: total login time, mean (SD) <sup>a</sup>	7.0 (N/A) <sup>c</sup>	153.7 (232.5)	588.8 (404.3)	1606.3 (1299.9)
Frequency: number of logins, mean (SD)	0.12 (0.49)	6.6 (6.5)	24.1 (9.4)	51.1 (23.1)
Activity: emails sent	0 (N/A)	3.0 (1.6)	7.7 (4.6)	11.6 (5.3)
Activity: exercises completed <sup>d</sup>	0 (N/A)	8.6 (6.5)	36.8 (7.2)	57.7 (9.2)

<sup>a</sup>Time was measured in minutes.

<sup>b</sup>N/A: not applicable.

<sup>c</sup>One participant in the nonuser group logged in twice, but did not complete any of the exercises, therefore belonging to the nonuser group. The duration scores reflect scores of this participant. Standard deviation cannot be calculated for one score.

<sup>d</sup>As this variable had a strongly skewed distribution, it was dichotomized (median split) for analysis.

### Prediction of Uptake and Adherence

#### Prediction of Uptake

[Table 2](#) presents the baseline characteristics for users and nonusers, and the results of the statistical tests comparing

differences between these groups. Nonusers had higher levels of baseline fear of cancer recurrence compared with users ( $t_{118}=2.27$ ,  $P=.03$ ). This effect was of a medium to large size ( $D=0.69$ ). There were no other differences between users and nonusers at baseline.

**Table 2.** Descriptive statistics of baseline characteristics of uptake (users vs nonusers) and prediction of uptake assessed with independent sample t tests or  $\chi^2$  tests (N=125).

Characteristic	Nonusers (n=17)	Users (n=108)	Test value (t value or $\chi^2$ value) (df)	P value	Cohen d
Age (years), mean (SD)	49.4 (11.4)	52.2 (10.1)	-1.04 (123)	.30	0.26
Gender (female), n (%)	17 (100%)	92 (85%)	2.89 (1)	.09	N/A <sup>a</sup>
Higher education (yes vs no), n (%)	8 (47%)	73 (68%)	2.72 (1)	.10	N/A
Cancer (breast vs other), n (%)	12 (71%)	64 (59%)	0.79 (1)	.37	N/A
Treatment intent (curative vs palliative), n (%)	14 (82%)	88 (82%)	0.01 (1)	.93	N/A
Neuroticism, mean (SD)	38.3 (8.1)	35.8 (7.8)	1.20 (122)	.23	0.31
Extraversion, mean (SD)	37.9 (5.8)	37.9 (6.5)	-0.04 (122)	.97	0.00
Openness, mean (SD)	40.0 (6.0)	40.4 (5.2)	-0.30 (122)	.76	0.07
Altruism, mean (SD)	45.8 (5.6)	46.6 (4.2)	-0.73 (122)	.47	0.16
Conscientiousness, mean (SD)	42.4 (6.8)	41.8 (5.9)	0.40 (122)	.69	0.09
Psychological distress, mean (SD)	16.4 (7.1)	16.8 (6.9)	-0.20 (119)	.84	0.06
Positive mental health, mean (SD)	39.7 (16.0)	37.4 (13.4)	0.59 (119)	.56	0.16
Rumination, mean (SD)	43.2 (9.5)	42.3 (8.3)	0.39 (118)	.70	0.10
Fear of cancer recurrence, mean (SD)	91.5 (18.7)	78.1 (20.3)	2.27 (118)	.03	0.69
Mindfulness, mean (SD)	78.4 (14.4)	77.0 (10.8)	0.44 (119)	.66	0.11

<sup>a</sup>N/A: not applicable.

### Prediction of Adherence

Table 3 presents the baseline characteristics for low/minimal and intended users, and the results of the statistical tests comparing differences between these groups. Intended users were more conscientious when compared with the combined group of minimal and low users ( $t_{106}=-2.04$ ,  $P=.04$ ). This effect was of a small to medium size ( $D=0.39$ ). There were no other differences between the two groups.

### Relations Between Adherence and Outcomes

Relations between adherence and outcomes are displayed in Table 4. Participants who used the intervention as intended

reported less psychological distress ( $t_{86}=-2.47$ ,  $P=.02$ ) and more positive mental health after the intervention than minimal and low users ( $t_{86}=5.18$ ,  $P=.02$ ). When focusing on specific usage measures, we found that participants who completed all exercises reported less psychological distress ( $t_{86}=-2.80$ ,  $P=.01$ ) and more positive mental health ( $t_{86}=5.24$ ,  $P=.01$ ) after the intervention than those who did not. Mean time logged in, total time logged in, number of logins, and number of emails sent did not appear to be related to psychological distress or positive mental health after the intervention.



**Table 3.** Descriptive statistics of baseline characteristics of adherence (minimal/low vs intended users) and prediction of adherence assessed with independent sample *t* tests or  $\chi^2$  tests (N=108).

Characteristic	Minimal/low users (n=43)	Intended users (n=65)	Test value ( <i>t</i> value or $\chi^2$ value) ( <i>df</i> )	<i>P</i> value	Cohen <i>d</i>
Age (years), mean (SD)	52.7 (10.1)	51.8 (10.1)	0.41 (106)	.68	0.09
Gender (female), n (%)	38 (88%)	54 (83%)	0.58 (1)	.45	N/A <sup>a</sup>
Higher education (yes vs no), n (%)	25 (58%)	48 (74%)	2.92 (1)	.09	N/A
Cancer (breast vs other), n (%)	25 (58%)	39 (60%)	0.04 (1)	.85	N/A
Treatment intent (curative vs palliative) n (%)	32 (74%)	56 (86%)	2.36 (1)	.12	N/A
Neuroticism, mean (SD)	36.8 (7.5)	35.1 (8.1)	1.13 (106)	.26	0.22
Extraversion, mean (SD)	36.8 (6.7)	38.7 (6.2)	-1.54 (106)	.13	0.29
Openness, mean (SD)	39.4 (5.8)	41.1 (4.8)	-1.62 (106)	.11	0.32
Altruism, mean (SD)	46.1 (3.7)	47.0 (4.5)	-1.13 (106)	.26	0.22
Conscientiousness, mean (SD)	40.4 (5.8)	42.7 (5.9)	-2.04 (106)	.04	0.39
Psychological distress, mean (SD)	16.6 (6.4)	17.0 (7.4)	-0.31 (105)	.76	0.06
Positive mental health, mean (SD)	37.9 (13.3)	37.1 (13.5)	0.31 (105)	.76	0.06
Rumination, mean (SD)	43.1 (6.7)	41.7 (9.2)	0.88 (105)	.38	0.17
Fear of cancer recurrence, mean (SD)	77.4 (19.7)	78.6 (20.8)	-0.31 (105)	.76	0.06
Mindfulness, mean (SD)	76.1 (9.7)	77.6 (11.6)	-0.69 (105)	.49	0.14

<sup>a</sup>N/A: not applicable.**Table 4.** Results of linear regression analyses predicting outcomes from usage group (intended vs minimal/low users) and separate usage measures of intervention users (N=89).

Variable <sup>a</sup>	Full model <sup>b</sup>			Predictor			
	Adjusted R <sup>2</sup>	<i>F</i> ( <i>df</i> )	<i>P</i>	B	<i>t</i> ( <i>df</i> )	<i>P</i>	95% CI
<b>Psychological distress</b>							
Adherence (minimal/low vs intended users)	0.34	23.76 (2,86)	<.001	-2.87	-2.47 (86)	.02	-5.19 to -0.56
Number of logins	0.32	20.46 (2,86)	<.001	-0.03	-1.23 (86)	.22	-0.07 to 0.02
Mean time logged in	0.31	20.69 (2,86)	<.001	-0.04	-1.35 (86)	.18	-0.09 to 0.02
Total time logged in	0.32	20.40 (2,86)	<.001	-0.001	-1.20 (86)	.23	-0.01 to 0.00
Emails sent	0.32	19.62 (2,83)	<.001	-0.06	-0.60 (83)	.55	-0.27 to 0.14
All exercises completed (yes/no)	0.34	24.48 (2,86)	<.001	-2.80	-2.51 (86)	.01	-5.02 to -0.59
<b>Positive mental health</b>							
Adherence (minimal/low vs intended users)	0.49	41.86 (2,86)	<.001	5.18	2.40 (86)	.02	0.88 to 9.48
Number of logins	0.45	37.28 (2,86)	<.001	0.03	0.90 (86)	.37	-0.04 to 0.11
Mean time logged in	0.46	36.77 (2,86)	<.001	0.03	0.50 (86)	.62	-0.08 to 0.13
Total time logged in	0.45	36.64 (2,86)	<.001	0.00	0.33 (86)	.74	-0.01 to 0.01
Emails sent	0.46	34.77 (2,83)	<.001	0.05	0.28 (83)	.78	-0.32 to 0.43
All exercises completed (yes/no)	0.51	45.51 (2,86)	<.001	5.24	2.53 (86)	.01	1.12 to 9.36

<sup>a</sup>Of the 108 users, 19 missed either the premeasure or the postmeasure of the outcomes, resulting in a sample size of 89 for these analyses. Three participants missed a score on emails sent, resulting in a sample size of 86 for these analyses.<sup>b</sup>The full models were controlled for baseline levels of psychological distress or positive mental health.

## Discussion

The aim of this study was to predict uptake and adherence of eMBCT and to examine the association between adherence and treatment outcome of eMBCT in distressed cancer patients. We divided participants into different user groups as follows: nonusers ( $n=17$ , 13.6%), minimal users ( $n=31$ , 24.8%), low users ( $n=12$ , 9.6%), and intended users ( $n=65$ , 52.0%). Regarding uptake, nonusers appeared to have more fear of cancer recurrence than users. Regarding adherence, intended users were more conscientious than minimal and low users. Finally, intended users showed a larger reduction in psychological distress and stronger improvement in positive mental health than minimal and low users. We did not find any relations of number of logins, mean and total time logged in, and number of emails sent with treatment outcome. Patients who completed all assignments, however, showed less psychological distress and more positive mental health after the intervention than those who did not.

About half of the participants used eMBCT as intended. A previous study on online cognitive behavioral therapy for cancer patients found a similar percentage of high users (44%) [26].

Regarding prediction of uptake, we found that nonusers had more fear of cancer recurrence than users. It is possible that patients with a high fear of recurrence are more likely to avoid participation, as it means being confronted with their fear. This is in line with a recent study on online self-help for fear of cancer recurrence that found that most patients do not login or express a need for support [45]. As fear of recurrence is an unmet need among cancer patients [46], this group might need more information prior to enrollment in an intervention like eMBCT [47]. A study on an internet-based intervention for depressive symptoms in primary care showed that a brief preparatory informational video increased acceptance of the intervention [48]. A video about eMBCT for distressed cancer patients could discuss evidence for the effectiveness of the intervention, and possible facilitators and barriers (such as confrontation with cancer and fear of recurrence). Preparatory face-to-face conversations with a health care provider prior to eMBCT, in which these topics could be discussed, may also be useful. Involving a partner or a close friend of the patient during eMBCT may help as well [49].

Regarding prediction of adherence, we showed that participants with low conscientiousness have trouble completing eMBCT. This is not surprising, as one of the aspects of conscientiousness is self-discipline [40]. Furthermore, increased conscientiousness is associated with increased mindfulness [50], which might catalyze participation in an online mindfulness program like this. As we previously found that participants with low conscientiousness had better results in eMBCT compared with regular group MBCT [18], more individual feedback by a mindfulness teacher might support them. Alternatively, patients could be offered a preintervention. For example, a study on health behavior showed that participants who were poor in planning benefitted from a preintervention aimed at planning and implementation intentions [51].

Besides fear of cancer recurrence and conscientiousness, we did not find any other predictors of usage of eMBCT. This is in accordance with a recent meta-analysis on predictors of online intervention adherence, which did not find many predictors for adherence either [22]. A proposed behavioral change model for internet interventions suggests that adherence is not only determined by personal characteristics, but also by environmental factors, support, and website characteristics [52], which we also showed in a qualitative study on eMBCT [47]. Future researchers might want to consider improving prediction by including all of these factors. In addition, qualitative methods can deepen our insight into the possible contributors to the usage of eMBCT [47].

Finally, we found that intended users reported less psychological distress and more positive mental health after the intervention than minimal and low users. A similar pattern was found for participants who completed all assignments as compared with those who did not. Thus, it seems that greater activity in eMBCT, rather than the sheer frequency or duration of use, is related to improved treatment outcome. A previous study on an online intervention for depression also found that clinically relevant improvement was related to activity rather than frequency and duration [53]. Our results also match with findings of a meta-analysis on the relation between adherence to homework and treatment outcome in MBIs [21]. The authors concluded that this relation is not linear, suggesting only frequency and duration of practice do not provide a complete picture. Thus, different individual usage patterns may be beneficial (eg, shorter and frequent sessions or longer and less frequent sessions), as long as activity is high.

Regarding clinical implications, our study showed that half of the participants used the program as intended and that higher activity was related to better outcome. As half of the participants did not use the program as intended, participation in these kinds of interventions should be closely monitored. In our case, monitoring took place in the form of weekly written asynchronous contact with a qualified trainer. Possibly, more intensive synchronous digital contact or even blended forms including face-to-face contact might be necessary for particular subgroups of patients [47]. To increase uptake of and adherence to this and similar programs, these programs could be designed as adaptive interventions in which the type (eg, regular group MBCT or eMBCT or hybrid/blended forms), dosage, and even content are individualized and flexible based on patients' characteristics, preferences, and clinical presentations, as well as their reactions to the intervention [54,55]. However, we should note that although only half of the participants did not use eMBCT as intended, previous results of the BeMind study showed eMBCT to be superior to treatment as usual [17] and, at long-term follow-up, even to regular face-to-face group MBCT [18].

The strengths of this study include comparison of different usage groups in terms of a large number of baseline characteristics and examination of the relationship between usage and treatment outcome. Another strength is that this study was part of an RCT that also included regular group-based MBCT as an intervention arm. Therefore, our sample might be less biased toward participants interested in participating in an online intervention.

Some limitations should be mentioned as well. First, the log measures tracked usage of eMBCT. However, participants may have been engaged with the mindfulness training in ways that were not reflected in the eMBCT log measures, as meditation audio files could be downloaded or found on YouTube. Although we assume that usage in eMBCT and engagement with training in other ways highly overlap, we cannot be sure we measured *actual* meditation practice in this study. As meditation is an important aspect of any mindfulness intervention, this is an important point for future research. Second, the amount of emails sent might be a less valid measurement of activity. Participants wrote their personal logs

and therapists could reply to those personal logs, creating an additional communication system next to email. Finally, owing to the low number of nonusers, the power to detect possible differences with *t* tests comparing nonusers and users was low, so it is likely that only larger effects were detected.

In conclusion, our study showed that adherence was related to improved patient outcomes. Patients with a strong fear of recurrence or low levels of conscientiousness should receive extra attention, as they are less likely to start or complete eMBCT. Future research may focus on the development of flexible and adaptive eMBCT programs.

## Acknowledgments

The BeMind project was financed by a fund of Pink Ribbon/KWF (no: 2012.WO14.C153).

## Conflicts of Interest

None declared.

## References

1. Carlson LE, Angen M, Cullum J, Goodey E, Koopmans J, Lamont L, et al. High levels of untreated distress and fatigue in cancer patients. *Br J Cancer* 2004 Jun 14;90(12):2297-2304 [FREE Full text] [doi: [10.1038/sj.bjc.6601887](https://doi.org/10.1038/sj.bjc.6601887)] [Medline: [15162149](https://pubmed.ncbi.nlm.nih.gov/15162149/)]
2. Zabora J, BrintzenhofeSzoc K, Curbow B, Hooker C, Piantadosi S. The prevalence of psychological distress by cancer site. *Psychooncology* 2001;10(1):19-28. [doi: [10.1002/1099-1611\(200101/02\)10:1<19::aid-pon501>3.0.co;2-6](https://doi.org/10.1002/1099-1611(200101/02)10:1<19::aid-pon501>3.0.co;2-6)] [Medline: [11180574](https://pubmed.ncbi.nlm.nih.gov/11180574/)]
3. Cillessen L, Johannsen M, Speckens AE, Zachariae R. Mindfulness-based interventions for psychological and physical health outcomes in cancer patients and survivors: A systematic review and meta-analysis of randomized controlled trials. *Psychooncology* 2019 Dec 11;28(12):2257-2269 [FREE Full text] [doi: [10.1002/pon.5214](https://doi.org/10.1002/pon.5214)] [Medline: [31464026](https://pubmed.ncbi.nlm.nih.gov/31464026/)]
4. Haller H, Winkler MM, Klose P, Dobos G, Kümmel S, Cramer H. Mindfulness-based interventions for women with breast cancer: an updated systematic review and meta-analysis. *Acta Oncol* 2017 Dec;56(12):1665-1676. [doi: [10.1080/0284186X.2017.1342862](https://doi.org/10.1080/0284186X.2017.1342862)] [Medline: [28686520](https://pubmed.ncbi.nlm.nih.gov/28686520/)]
5. Kabat-Zinn J. Full Catastrophe Living, Revised Edition: How to cope with stress, pain and illness using mindfulness meditation. UK: Hachette; 2013.
6. Kabat-Zinn J. Wherever You Go, There You Are: Mindfulness Meditation in Everyday Life. UK: Hachette; 2009.
7. Segal Z, Williams J, Teasdale J. Mindfulness-based cognitive therapy for depression. New York: Guilford Press; 2013.
8. Lawrence DP, Kupelnick B, Miller K, Devine D, Lau J. Evidence report on the occurrence, assessment, and treatment of fatigue in cancer patients. *J Natl Cancer Inst Monogr* 2004(32):40-50. [doi: [10.1093/jncimonographs/lgh027](https://doi.org/10.1093/jncimonographs/lgh027)] [Medline: [15263040](https://pubmed.ncbi.nlm.nih.gov/15263040/)]
9. Meretoja TJ, Leidenius MH, Tasmuth T, Sipilä R, Kalso E. Pain at 12 months after surgery for breast cancer. *JAMA* 2014 Jan 01;311(1):90-92. [doi: [10.1001/jama.2013.278795](https://doi.org/10.1001/jama.2013.278795)] [Medline: [24381969](https://pubmed.ncbi.nlm.nih.gov/24381969/)]
10. Smith HS, Wu SX. Persistent pain after breast cancer treatment. *Annals of palliative medicine* 2013;1(3):182-194. [doi: [10.1093/med/9780199796410.003.0013](https://doi.org/10.1093/med/9780199796410.003.0013)] [Medline: [32966010](https://pubmed.ncbi.nlm.nih.gov/32966010/)]
11. Proudfoot J, Klein B, Barak A, Carlbring P, Cuijpers P, Lange A, et al. Establishing guidelines for executing and reporting Internet intervention research. *Cogn Behav Ther* 2011;40(2):82-97. [doi: [10.1080/16506073.2011.573807](https://doi.org/10.1080/16506073.2011.573807)] [Medline: [25155812](https://pubmed.ncbi.nlm.nih.gov/25155812/)]
12. Carlbring P, Andersson G. Internet and psychological treatment. How well can they be combined? *Computers in Human Behavior* 2006 May;22(3):545-553 [FREE Full text] [doi: [10.1016/j.chb.2004.10.009](https://doi.org/10.1016/j.chb.2004.10.009)]
13. Christensen H. Computerised therapy for psychiatric disorders. *The Lancet* 2007 Jul 14;370(9582):112-113. [doi: [10.1016/S0140-6736\(07\)61065-4](https://doi.org/10.1016/S0140-6736(07)61065-4)] [Medline: [17630022](https://pubmed.ncbi.nlm.nih.gov/17630022/)]
14. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. *Clin Psychol Rev* 2016 Apr;45:102-114 [FREE Full text] [doi: [10.1016/j.cpr.2016.03.009](https://doi.org/10.1016/j.cpr.2016.03.009)] [Medline: [27111302](https://pubmed.ncbi.nlm.nih.gov/27111302/)]
15. Zernicke KA, Campbell TS, Specia M, McCabe-Ruff K, Flowers S, Carlson LE. A randomized wait-list controlled trial of feasibility and efficacy of an online mindfulness-based cancer recovery program: the eTherapy for cancer applying mindfulness trial. *Psychosom Med* 2014 May;76(4):257-267. [doi: [10.1097/PSY.000000000000053](https://doi.org/10.1097/PSY.000000000000053)] [Medline: [24804884](https://pubmed.ncbi.nlm.nih.gov/24804884/)]

16. Bruggeman-Everts FZ, Wolvers MD, van de Schoot R, Vollenbroek-Hutten MM, Van der Lee ML. Effectiveness of Two Web-Based Interventions for Chronic Cancer-Related Fatigue Compared to an Active Control Condition: Results of the "Fitter na kanker" Randomized Controlled Trial. *J Med Internet Res* 2017 Oct 19;19(10):e336 [FREE Full text] [doi: [10.2196/jmir.7180](https://doi.org/10.2196/jmir.7180)] [Medline: [29051138](https://pubmed.ncbi.nlm.nih.gov/29051138/)]
17. Compen F, Bisseling E, Schellekens M, Donders R, Carlson L, van der Lee M, et al. Face-to-Face and Internet-Based Mindfulness-Based Cognitive Therapy Compared With Treatment as Usual in Reducing Psychological Distress in Patients With Cancer: A Multicenter Randomized Controlled Trial. *J Clin Oncol* 2018 Aug 10;36(23):2413-2421. [doi: [10.1200/JCO.2017.76.5669](https://doi.org/10.1200/JCO.2017.76.5669)] [Medline: [29953304](https://pubmed.ncbi.nlm.nih.gov/29953304/)]
18. Cillessen L, Schellekens MP, Van de Ven MO, Donders AR, Compen FR, Bisseling EM, et al. Consolidation and prediction of long-term treatment effect of group and online mindfulness-based cognitive therapy for distressed cancer patients. *Acta Oncol* 2018 Oct;57(10):1293-1302. [doi: [10.1080/0284186X.2018.1479071](https://doi.org/10.1080/0284186X.2018.1479071)] [Medline: [29932784](https://pubmed.ncbi.nlm.nih.gov/29932784/)]
19. Compen F, Adang E, Bisseling E, van der Lee M, Speckens A. Cost-utility of individual internet-based and face-to-face Mindfulness-Based Cognitive Therapy compared with treatment as usual in reducing psychological distress in cancer patients. *Psychooncology* 2020 Feb 26;29(2):294-303 [FREE Full text] [doi: [10.1002/pon.5246](https://doi.org/10.1002/pon.5246)] [Medline: [31650662](https://pubmed.ncbi.nlm.nih.gov/31650662/)]
20. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the Relationship Between Usage and Outcomes of an Internet-Based Intervention for Individuals With Depressive Symptoms: Secondary Analysis of Data From a Randomized Controlled Trial. *J Med Internet Res* 2019 Aug 01;21(8):e12775 [FREE Full text] [doi: [10.2196/12775](https://doi.org/10.2196/12775)] [Medline: [31373272](https://pubmed.ncbi.nlm.nih.gov/31373272/)]
21. Parsons CE, Crane C, Parsons LJ, Fjorback LO, Kuyken W. Home practice in Mindfulness-Based Cognitive Therapy and Mindfulness-Based Stress Reduction: A systematic review and meta-analysis of participants' mindfulness practice and its association with outcomes. *Behav Res Ther* 2017 Aug;95:29-41 [FREE Full text] [doi: [10.1016/j.brat.2017.05.004](https://doi.org/10.1016/j.brat.2017.05.004)] [Medline: [28527330](https://pubmed.ncbi.nlm.nih.gov/28527330/)]
22. Beatty L, Binnion C. A Systematic Review of Predictors of, and Reasons for, Adherence to Online Psychological Interventions. *Int J Behav Med* 2016 Dec;23(6):776-794. [doi: [10.1007/s12529-016-9556-9](https://doi.org/10.1007/s12529-016-9556-9)] [Medline: [26957109](https://pubmed.ncbi.nlm.nih.gov/26957109/)]
23. Beatty L, Kemp E, Binnion C, Turner J, Milne D, Butow P, et al. Uptake and adherence to an online intervention for cancer-related distress: older age is not a barrier to adherence but may be a barrier to uptake. *Support Care Cancer* 2017 Jun;25(6):1905-1914. [doi: [10.1007/s00520-017-3591-1](https://doi.org/10.1007/s00520-017-3591-1)] [Medline: [28155018](https://pubmed.ncbi.nlm.nih.gov/28155018/)]
24. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Strategies to facilitate exposure to internet-delivered health behavior change interventions aimed at adolescents or young adults: a systematic review. *Health Educ Behav* 2011 Feb;38(1):49-62. [doi: [10.1177/1090198110372878](https://doi.org/10.1177/1090198110372878)] [Medline: [21189422](https://pubmed.ncbi.nlm.nih.gov/21189422/)]
25. Villaume K, Tafvelin S, Hasson D. Health-relevant personality traits in relation to adherence to a web-based occupational health promotion and stress management intervention. *Intl J of Workplace Health Mgt* 2018 Jun 04;11(3):143-158. [doi: [10.1108/ijwhm-11-2017-0092](https://doi.org/10.1108/ijwhm-11-2017-0092)]
26. van den Berg SW, Peters EJ, Kraaijeveld JF, Gielissen MF, Prins JB. Usage of a generic web-based self-management intervention for breast cancer survivors: substudy analysis of the BREATH trial. *J Med Internet Res* 2013 Aug 19;15(8):e170 [FREE Full text] [doi: [10.2196/jmir.2566](https://doi.org/10.2196/jmir.2566)] [Medline: [23958584](https://pubmed.ncbi.nlm.nih.gov/23958584/)]
27. Compen FR, Bisseling EM, Van der Lee ML, Adang EM, Donders AR, Speckens AE. Study protocol of a multicenter randomized controlled trial comparing the effectiveness of group and individual internet-based Mindfulness-Based Cognitive Therapy with treatment as usual in reducing psychological distress in cancer patients: the BeMind study. *BMC Psychol* 2015;3:27 [FREE Full text] [doi: [10.1186/s40359-015-0084-1](https://doi.org/10.1186/s40359-015-0084-1)] [Medline: [26273472](https://pubmed.ncbi.nlm.nih.gov/26273472/)]
28. Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychol Med* 1997 Mar;27(2):363-370. [doi: [10.1017/s0033291796004382](https://doi.org/10.1017/s0033291796004382)] [Medline: [9089829](https://pubmed.ncbi.nlm.nih.gov/9089829/)]
29. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983 Jun;67(6):361-370. [doi: [10.1111/j.1600-0447.1983.tb09716.x](https://doi.org/10.1111/j.1600-0447.1983.tb09716.x)] [Medline: [6880820](https://pubmed.ncbi.nlm.nih.gov/6880820/)]
30. Segal Z, Williams J, Teasdale J. Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse. New York: Guilford Press; 2002.
31. Good practice guidelines for teaching mindfulness-based courses. UK Network for Mindfulness-Based Teachers. 2015. URL: <https://bamba.org.uk/wp-content/uploads/2019/06/UK-MB-teacher-GPG-2015-final-2.pdf>
32. Bruggeman Everts F, van der Lee M, de Jager Meezenbroek E. Web-based individual Mindfulness-Based Cognitive Therapy for cancer-related fatigue — A pilot study. *Internet Interventions* 2015 May;2(2):200-213 [FREE Full text] [doi: [10.1016/j.invent.2015.03.004](https://doi.org/10.1016/j.invent.2015.03.004)]
33. Dobkin PL, Bernardi NF, Bagnis CI. Enhancing Clinicians' Well-Being and Patient-Centered Care Through Mindfulness. *Journal of Continuing Education in the Health Professions* 2016;36(1):11-16. [doi: [10.1097/ceh.0000000000000021](https://doi.org/10.1097/ceh.0000000000000021)]
34. Teasdale JD, Segal ZV, Williams JM, Ridgeway VA, Soulsby JM, Lau MA. Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *J Consult Clin Psychol* 2000 Aug;68(4):615-623. [doi: [10.1037/0022-006X.68.4.615](https://doi.org/10.1037/0022-006X.68.4.615)] [Medline: [10965637](https://pubmed.ncbi.nlm.nih.gov/10965637/)]
35. Kuyken W, Byford S, Taylor RS, Watkins E, Holden E, White K, et al. Mindfulness-based cognitive therapy to prevent relapse in recurrent depression. *J Consult Clin Psychol* 2008 Dec;76(6):966-978. [doi: [10.1037/a0013786](https://doi.org/10.1037/a0013786)] [Medline: [19045965](https://pubmed.ncbi.nlm.nih.gov/19045965/)]



36. Trapnell PD, Campbell JD. Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology* 1999 Feb;76(2):284-304. [doi: [10.1037/0022-3514.76.2.284](https://doi.org/10.1037/0022-3514.76.2.284)] [Medline: [10074710](https://pubmed.ncbi.nlm.nih.gov/10074710/)]
37. Simard S, Savard J. Fear of Cancer Recurrence Inventory: development and initial validation of a multidimensional measure of fear of cancer recurrence. *Support Care Cancer* 2009 Mar;17(3):241-251. [doi: [10.1007/s00520-008-0444-y](https://doi.org/10.1007/s00520-008-0444-y)] [Medline: [18414902](https://pubmed.ncbi.nlm.nih.gov/18414902/)]
38. van Helmond SJ, van der Lee ML, de Vries J. Translation and validation of the Dutch version of the Fear of Cancer Recurrence Inventory (FCRI-NL). *J Psychosom Res* 2017 Nov;102:21-28. [doi: [10.1016/j.jpsychores.2017.09.001](https://doi.org/10.1016/j.jpsychores.2017.09.001)] [Medline: [28992893](https://pubmed.ncbi.nlm.nih.gov/28992893/)]
39. Bohlmeijer E, ten Klooster PM, Fledderus M, Veehof M, Baer R. Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. *Assessment* 2011 Sep;18(3):308-320. [doi: [10.1177/1073191111408231](https://doi.org/10.1177/1073191111408231)] [Medline: [21586480](https://pubmed.ncbi.nlm.nih.gov/21586480/)]
40. Costa PT, MacCrae RR. Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI). Florida, USA: Psychological Assessment Resources; 1992.
41. Vodermaier A, Linden W, Siu C. Screening for emotional distress in cancer patients: a systematic review of assessment instruments. *J Natl Cancer Inst* 2009 Nov 04;101(21):1464-1488 [FREE Full text] [doi: [10.1093/jnci/djp336](https://doi.org/10.1093/jnci/djp336)] [Medline: [19826136](https://pubmed.ncbi.nlm.nih.gov/19826136/)]
42. Mitchell AJ, Meader N, Symonds P. Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: a meta-analysis. *J Affect Disord* 2010 Nov;126(3):335-348. [doi: [10.1016/j.jad.2010.01.067](https://doi.org/10.1016/j.jad.2010.01.067)] [Medline: [20207007](https://pubmed.ncbi.nlm.nih.gov/20207007/)]
43. Keyes CL. The Mental Health Continuum: From Languishing to Flourishing in Life. *Journal of Health and Social Behavior* 2002 Jun;43(2):207. [doi: [10.2307/3090197](https://doi.org/10.2307/3090197)]
44. Lamers SM, Westerhof GJ, Bohlmeijer ET, ten Klooster PM, Keyes CL. Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF). *J Clin Psychol* 2011 Jan 22;67(1):99-110. [doi: [10.1002/jclp.20741](https://doi.org/10.1002/jclp.20741)] [Medline: [20973032](https://pubmed.ncbi.nlm.nih.gov/20973032/)]
45. van Helmond SJ, van der Lee ML, van Woezik RA, Lodder P, de Vries J. No effect of CBT-based online self-help training to reduce fear of cancer recurrence: First results of the CAREST multicenter randomized controlled trial. *Psychooncology* 2020 Jan 13;29(1):86-97. [doi: [10.1002/pon.5233](https://doi.org/10.1002/pon.5233)] [Medline: [31595627](https://pubmed.ncbi.nlm.nih.gov/31595627/)]
46. Simard S, Thewes B, Humphris G, Dixon M, Hayden C, Mireskandari S, et al. Fear of cancer recurrence in adult cancer survivors: a systematic review of quantitative studies. *J Cancer Surviv* 2013 Sep;7(3):300-322. [doi: [10.1007/s11764-013-0272-z](https://doi.org/10.1007/s11764-013-0272-z)] [Medline: [23475398](https://pubmed.ncbi.nlm.nih.gov/23475398/)]
47. Compen FR, Bisseling EM, Schellekens MP, Jansen ET, van der Lee ML, Speckens AE. Mindfulness-Based Cognitive Therapy for Cancer Patients Delivered via Internet: Qualitative Study of Patient and Therapist Barriers and Facilitators. *J Med Internet Res* 2017 Dec 18;19(12):e407 [FREE Full text] [doi: [10.2196/jmir.7783](https://doi.org/10.2196/jmir.7783)] [Medline: [29254912](https://pubmed.ncbi.nlm.nih.gov/29254912/)]
48. Ebert D, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. *J Affect Disord* 2015 May 01;176:9-17. [doi: [10.1016/j.jad.2015.01.056](https://doi.org/10.1016/j.jad.2015.01.056)] [Medline: [25682378](https://pubmed.ncbi.nlm.nih.gov/25682378/)]
49. van den Hurk DG, Schellekens MP, Molema J, Speckens AE, van der Drift MA. Mindfulness-Based Stress Reduction for lung cancer patients and their partners: Results of a mixed methods pilot study. *Palliat Med* 2015 Jul 20;29(7):652-660. [doi: [10.1177/0269216315572720](https://doi.org/10.1177/0269216315572720)] [Medline: [25701663](https://pubmed.ncbi.nlm.nih.gov/25701663/)]
50. Giluk T. Mindfulness, Big Five personality, and affect: A meta-analysis. *Personality and Individual Differences* 2009 Dec;47(8):805-811 [FREE Full text] [doi: [10.1016/j.paid.2009.06.026](https://doi.org/10.1016/j.paid.2009.06.026)]
51. Allan JL, Sniehotta FF, Johnston M. The best laid plans: planning skill determines the effectiveness of action plans and implementation intentions. *Ann Behav Med* 2013 Aug 28;46(1):114-120. [doi: [10.1007/s12160-013-9483-9](https://doi.org/10.1007/s12160-013-9483-9)] [Medline: [23456214](https://pubmed.ncbi.nlm.nih.gov/23456214/)]
52. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. *Ann Behav Med* 2009 Aug;38(1):18-27 [FREE Full text] [doi: [10.1007/s12160-009-9133-4](https://doi.org/10.1007/s12160-009-9133-4)] [Medline: [19802647](https://pubmed.ncbi.nlm.nih.gov/19802647/)]
53. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. *J Med Internet Res* 2013 Oct 17;15(10):e231 [FREE Full text] [doi: [10.2196/jmir.2771](https://doi.org/10.2196/jmir.2771)] [Medline: [24135213](https://pubmed.ncbi.nlm.nih.gov/24135213/)]
54. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Stat Med* 2005 May 30;24(10):1455-1481. [doi: [10.1002/sim.2022](https://doi.org/10.1002/sim.2022)] [Medline: [15586395](https://pubmed.ncbi.nlm.nih.gov/15586395/)]
55. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med* 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: [10.1016/j.amepre.2007.01.022](https://doi.org/10.1016/j.amepre.2007.01.022)] [Medline: [17466815](https://pubmed.ncbi.nlm.nih.gov/17466815/)]



## Abbreviations

**eMBCT:** online mindfulness-based cognitive therapy  
**HADS:** Hospital Anxiety and Depression Scale  
**MBCT:** mindfulness-based cognitive therapy  
**MBI:** mindfulness-based intervention  
**MBSR:** mindfulness-based stress reduction  
**MHC-SF:** Mental Health Continuum Short Form  
**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 18.12.19; peer-reviewed by KS Pheh, S Martin, E Cox-martin; comments to author 23.03.20; revised version received 15.07.20; accepted 19.07.20; published 02.10.20.*

*Please cite as:*

Cillessen L, van de Ven MOM, Compen FR, Bisseling EM, van der Lee ML, Speckens AEM

Predictors and Effects of Usage of an Online Mindfulness Intervention for Distressed Cancer Patients: Usability Study

J Med Internet Res 2020;22(10):e17526

URL: <https://www.jmir.org/2020/10/e17526>

doi: [10.2196/17526](https://doi.org/10.2196/17526)

PMID: [33006567](https://pubmed.ncbi.nlm.nih.gov/33006567/)

©Linda Cillessen, Monique OM van de Ven, Félix R Compen, Else M Bisseling, Marije L van der Lee, Anne EM Speckens. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 02.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Twelve-Month Follow-Up to a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention for Rural Adults With Depression Symptoms: Single-Arm Longitudinal Study

Mark Schure<sup>1</sup>, BSc, MSc, PhD; Bernadette McCrory<sup>2</sup>, MPH, PhD; Kathryn Tuchscherer Franklin<sup>1</sup>, MEd, PhD; John Greist<sup>3</sup>, MD; Ruth Striegel Weissman<sup>3</sup>, PhD

<sup>1</sup>Department of Health & Human Development, Montana State University, Bozeman, MT, United States

<sup>2</sup>Department of Mechanical & Industrial Engineering, Montana State University, Bozeman, MT, United States

<sup>3</sup>Center for Mental Health Research and Recovery, Montana State University, Bozeman, MT, United States

**Corresponding Author:**

Mark Schure, BSc, MSc, PhD

Department of Health & Human Development

Montana State University

305 Herrick Hall

Bozeman, MT, 59717

United States

Phone: 1 406 994 3248

Email: [mark.schure@montana.edu](mailto:mark.schure@montana.edu)

**Related Article:**

This is a corrected version. See correction statement: <https://www.jmir.org/2020/10/e25146/>

## Abstract

**Background:** Internet-based cognitive behavior therapy (iCBT) interventions have the potential to help individuals with depression, regardless of time and location. Yet, limited information exists on the longer-term (>6 months) effects of iCBT and adherence to these interventions.

**Objective:** The primary aim of this study was to evaluate the longitudinal (12 months) effectiveness of a fully automated, self-guided iCBT intervention called Thrive, designed to enhance engagement with a rural population of adults with depression symptoms. The secondary aim was to determine whether the program adherence enhanced the effectiveness of the Thrive intervention.

**Methods:** We analyzed data from 181 adults who used the Thrive intervention. Using self-reports, participants were evaluated at baseline, 8 weeks, 6 months, and 12 months for the primary outcome of depression symptom severity using the Patient Health Questionnaire-9 (PHQ-9) scale and secondary outcome measures, namely, the Generalized Anxiety Disorder Scale-7 (GAD-7) scores, Work and Social Adjustment Scale (WSAS) scores, Conner-Davidson Resilience Scale-10 (CD-RISC-10) scores, and suicidal ideation (ninth item of the PHQ-9 scale) scores. The Thrive program adherence was measured using the numbers of program logins, page views, and lessons completed.

**Results:** The assessment response rates for 8-week, 6-month, and 12-month outcomes were 58.6% (106/181), 50.3% (91/181), and 51.4% (93/181), respectively. By 8 weeks, significant improvements were observed for all outcome measures. These improvements were maintained at 12 months with mean reductions in severities of depression (mean  $-6.5$ ;  $P<.001$ ) and anxiety symptoms (mean  $-4.3$ ;  $P<.001$ ). Improvements were also observed in work and social functioning (mean  $-6.9$ ;  $P<.001$ ) and resilience (mean  $4.3$ ;  $P<.001$ ). Marked decreases were observed in suicidal ideation (PHQ-9 ninth item score  $>1$ ) at 6 months (16.5%) and 12 months (17.2%) compared to baseline (39.8%) ( $P<.001$ ). In regard to the program adherence, cumulative counts of page views and lessons completed were significantly related to lower PHQ-9, GAD-7, and WSAS scores and higher CD-RISC-10 scores (all  $P$  values  $<.001$  with an exception of page views with WSAS for which  $P$  value was .02).

**Conclusions:** The Thrive intervention was effective at reducing depression and anxiety symptom severity and improving functioning and resilience among a population of adults from mostly rural communities in the United States. These gains were

maintained at 1 year. Program adherence, measured by the number of logins and lessons completed, indicates that users who engage more with the program benefit more from the intervention.

**Trial Registration:** ClinicalTrials.gov NCT03244878; <https://clinicaltrials.gov/ct2/show/NCT03244878>

(*J Med Internet Res* 2020;22(10):e21336) doi:[10.2196/21336](https://doi.org/10.2196/21336)

## KEYWORDS

internet-based cognitive behavior therapy; depression; anxiety; long-term outcomes; iCBT; CBT; therapy; mental health; outcome

## Introduction

Clinician-delivered cognitive behavior therapy (CBT) is a long-standing evidence-based psychotherapy for depression and anxiety symptoms and disorders [1,2]. Computerized forms of CBT were introduced three decades ago, paving the way for present efforts to implement such programs via the internet [3]. Compared to clinician-delivered CBT, internet-based cognitive behavior therapy (iCBT) programs have potential for greater reach and scalability, greater standardization of content delivery, and reduced risk of stigmatization [4-6]. Even more, they have demonstrated equivalent effectiveness for reducing depression and anxiety symptoms [7].

Studies support the feasibility, acceptability, and effectiveness of self-guided (no supportive contacts by email, text, telephone, or face-to-face) iCBT interventions on depression and anxiety symptoms [4,6,8-10]. These findings are particularly promising for people living in rural and frontier communities, which, nationally and internationally, have greater behavioral health care access challenges [11-13]. Compared with urban residents, rural and frontier residents have fewer qualified mental and behavioral health care providers and longer travel times to clinical services; moreover, they report greater concerns about privacy and higher levels of stigma [8,14-16]. Outside of the United States, other identified barriers to care in rural settings include long wait times for appointments, cost of care, transportation, lack of education, and stigma toward seeking mental health care [17,18].

A meta-analysis of 13 randomized controlled trials (RCTs), all conducted outside the United States [4], evaluated the efficacy of self-guided iCBT interventions for adults with depression symptoms. Compared to controls, iCBT was significantly more effective for the intervention groups on depressive symptoms severity. Furthermore, program adherence was significantly associated with reduction in depression symptoms. Another meta-analysis of 64 RCTs evaluated the effectiveness of iCBT, usually delivered with support from a health care provider or other individual, compared to the usual care, wait list, or placebo intervention in the treatment of depression or anxiety [6]. The meta-analysis reported superior outcomes among the intervention group compared to the control group. Most self-guided iCBT trials with outcomes of depression and anxiety have shorter follow-up assessment periods ( $\leq 6$  months). We identified only 2 studies reporting longer-term effects of completely self-guided iCBT on depression symptoms. In Mira and colleagues' trial [19], 12-month effect sizes for depression symptoms demonstrated a moderate within-group effect size ( $d = 0.67$ ). Clarke and colleagues [20] reported on an 8-month

follow-up on depression symptoms, indicating a comparable effect size of 0.72 (calculated by authors).

In the context of online programming for mental health interventions, adherence has been defined as "the extent to which individuals experience or engage with content" [21]. However, operationalizing adherence to iCBT interventions poses challenges. For example, while in-person CBT typically involves treatment sessions (with a mutually agreed upon start and end time), electronic delivery of treatment has varying degrees of engagement and length of use. The content of iCBT may be designed to comprise numerous short videos that can be viewed in a variety of temporal orders and explored in a variety of timeframes different from what would typically be encountered in an in-person session. Hence, the concept of a session may not easily translate from in-person CBT to iCBT or from one iCBT program to another.

Most published RCTs evaluating the efficacy of iCBT interventions have been implemented in non-US urban settings. Our research team evaluated an iCBT intervention, called Thrive, designed to help improve depression and anxiety symptoms for adults residing in the western rural communities in the United States [8,9]. Thrive is an interactive, fully automated, self-guided intervention that uses a video-based platform to deliver CBT curriculum and provide supportive feedback to users [22]. We implemented a pragmatic usual care waitlist control (WLC) RCT, enrolling 343 adults with at least mild depression symptoms (a Patient Health Questionnaire-9 [PHQ-9] score  $> 5$ ), who were randomized either to immediate access to the Thrive intervention or to 8-week delayed access [9]. In models adjusted for potential confounders, depression severity following 8 weeks of intervention was significantly lower for the immediate access group than for the WLC group ( $d = 0.63$ ). Superior 8-week outcomes in the immediate access group versus WLC group were also observed for secondary outcomes, including anxiety symptoms ( $d = 0.47$ ), work and social functioning ( $d = 0.39$ ), and resilience ( $d = 0.55$ ). Although not statistically significant, the immediate access group was 45% less likely (odds ratio [OR] 0.55) to report suicidal ideation compared to the WLC group.

Since there are few studies examining the long-term impact of self-guided iCBT interventions, the primary aim of this within-group analysis was to assess 6- and 12-month follow-up outcomes of trial participants who received immediate access to the Thrive intervention. Moreover, to date, there has been no consensus on the operationalization or impact of adherence on outcomes in iCBT interventions [21]. Thus, our secondary aim was to determine whether program adherence enhanced the effectiveness of the Thrive intervention.

## Methods

### Trial Design

A longitudinal study design was used to evaluate the effectiveness of the Thrive intervention with participants receiving immediate access to the iCBT program. Study participants were enrolled in the study between September 2017 and January 2018 and were provided free online access to the intervention for 1 year. The original study design was a waitlist RCT in which participants were randomly assigned to receive either an immediate access to the iCBT intervention or access to the intervention delayed 8 weeks. The WLC group received a link to the National Institute of Mental Health's depression information webpage. All participants received a link to the resource webpage of Montana's chapter of the National Alliance on Mental Illness. All participants were also permitted to continue or begin whatever usual care was available to them. The Montana State University Institutional Review Board (IRB) approved the protocol and all related materials (#MS033017-FC) prior to study initiation. The study is registered at ClinicalTrials.gov (NCT03244878).

### Study Recruitment

We used several strategies for promoting the study to state residents. We first partnered with Montana State University Extension faculty to disseminate study brochures and flyers that guided potential participants to the study website, where they could learn more and sign up. Research team members also conducted 12 community meetings throughout all regions of the state, which were promoted by local extension agents. Other recruitment sources included public service announcements, local newspaper articles, social media (Facebook and Craigslist community pages), email listservs, large employers, and local health care providers.

### Study Eligibility and Participants

Eligibility requirements for the study included age >18 years; Montana state residency; having regular access to broadband internet via a computer, tablet, or smartphone; and reporting at least mild depression symptoms (PHQ-9 score >5) at baseline. Potential participants were directed to a study website where they were informed about study participation; self-screened for eligibility; and, if eligible, guided through the informed consent, randomization, and online assessment process [23]. The Montana State University logo was displayed throughout the study website pages. Of the 573 individuals assessed for eligibility, 463 individuals were deemed study eligible and enrolled in the study; yet 109 were later identified as fraudulent and removed from the study. The sample for the current study included 181 eligible study participants who had immediate access to the Thrive intervention (see [Multimedia Appendix 1](#) for the CONSORT (Consolidated Standards of Reporting Trials) flowchart).

### Intervention Description

Thrive, developed by Waypoint Health Innovations [22], is a self-guided iCBT intervention for depression and anxiety that distills best practices from CBT and delivers them through a rich, structured, and guided curriculum. Thrive uses video,

interactive tools, and sophisticated algorithms that dynamically adjust the individual's course through the intervention. The intervention is comprised of 320 videos, averaging 80 seconds in length, to deliver content. Videos explain CBT concepts, demonstrate skills, provide feedback and recommendations, and portray actual case histories of individuals who used CBT skills to improve depression symptoms. The intervention also provides periodic PHQ-9 self-assessments and tailored feedback based on the scores. For this study, over a third of the demonstration and case history videos were replaced with new videos featuring rural characters, story lines, and settings. Other features of the Thrive program (ie, didactic and feedback videos, interactive tools, and algorithms) were not modified for this study. Thrive incorporates classic cognitive behavior therapy themes in modules (series of the didactic and feedback videos and interactive tools) on Constructive Thinking (cognitive restructuring), Pleasant Activities (behavioral activation), and Assertive Communication (social skills training). Each module has 10 lessons and suggested exercises for users to practice offline as homework pertinent to their own goals.

The cost for an individual to use Thrive for 6 months is roughly equivalent to the cost of 1 session with a therapist.

### Assessments

All participants were assessed at baseline, 8 weeks, 6 months, and 12 months after study enrollment for each outcome measure. Each participant received email reminders when assessments were due, and 2 additional reminders within 7 days were issued for those who had not yet completed their assessment. Data were designated as lost to follow-up when no assessment was completed within 10 days of the due date. Participants with completed interim assessments were rewarded with a US \$25 Amazon gift code, and those with the completed final assessment (12 months) were rewarded with a US \$30 Amazon gift code.

### Measures

All outcome measures and other demographic and treatment measures were administered electronically via the study assessment portal. The primary outcome measure was depression symptom severity measured by the PHQ-9, (score range 0-27; higher scores indicate greater severity) [24]. Secondary outcome measures included anxiety symptom severity, daily functioning, resilience, and suicidal ideation. Anxiety symptom severity was measured with the Generalized Anxiety Disorder Scale-7 (GAD-7) (score range 0-21; higher scores indicate greater severity) [25]. Daily functioning was measured with the Work and Social Adjustment Scale (WSAS) (score range 0-40; higher scores indicate worse daily functioning) [26]. Resilience was measured with the Connor-Davidson Resilience Scale-10 (CD-RISC-10) (score range 0-40; higher scores indicate greater resilience) [27]. Frequency of suicidal ideation was measured with the ninth item of the PHQ-9 (score range 0-3; higher scores indicate greater suicidal ideation).

We assessed rates of remission and relapse. Remission was defined as a treatment response in which an individual with mild, moderate, moderately severe, or severe depression at baseline (PHQ-9 scores  $\geq 5$ ) achieved a subsequent PHQ-9 score



lower than 5 at 6 months and 12 months. Relapse was defined as a PHQ-9 score  $\geq 10$  for those who had achieved remission at 6 months or 12 months. Cumulative program adherence was measured with the following indicators: (1) number of logins, (2) number of page views, and (3) number of lessons completed within the program. Cumulative counts of each of these 3 program usage measures were assessed as explanatory variables for each of the 4 continuous outcome measures (PHQ-9, GAD-7, WSAS, CD-RISC-10) and the PHQ-9 ninth item, suicidal ideation.

Demographic variables included age (years), gender (female vs male), race (White vs other), marital status (single vs married/domestic relationship), employment status (employed full-time, employed part-time, unemployed/retired/student), veteran status (yes vs no), educational attainment ( $\leq$  college without a degree, bachelor's degree,  $\geq$  master's degree), health insurance (private, public, other, none), and rural classification (urban, large rural, small rural, isolated).

### Participant Safety

Participants were encouraged to seek or continue other available care throughout the study. During all assessments, participants who reported any frequency of suicidal ideation (PHQ-9 ninth item score  $>0$ ) on the assessment portal were encouraged to seek help from multiple sources. They were also asked whether they could keep themselves safe from self-harm. Those responding they could not keep themselves safe would be told not to continue in the study and were provided a list of things to do to seek professional help. All participants were provided a resource list for seeking additional support ([Multimedia Appendix 2](#)). However, none of the participants responded they could not be safe during any assessment. Additionally, in the Thrive intervention, any self-assessed PHQ-9 scores  $>20$  and a PHQ-9 score  $>10$  on the third self-assessment recommended seeking a clinician's help. Participants were also provided contact information of the institution's IRB director and the study's principal investigator to report any adverse events. No adverse events were reported in this study.

### Statistical Analysis

The longitudinal change over time in each continuous outcome was assessed using a linear mixed model analysis of repeated measures. Separate models were created for each outcome measure (ie, PHQ-9, GAD-7, WSAS, and CD-RISC-10) to assess the fixed effect of time adjusting for baseline scores and receiving therapy for depression. Similar separate models were used to assess the relationship program logins, page views, and lessons completed on each outcome, adjusting for baseline scores, therapy for depression, and time.

The PHQ-9 ninth item, suicidal ideation, was treated as an ordinal outcome that ranged from 0 ("not at all," no suicidal ideation) to 3 ("nearly every day") where the cumulative probabilities were modeled over the higher-ordered suicidal ideation scale scores (indicating greater suicidal ideation). Using an ordinal logistic regression model within a generalized estimating equation framework, the PHQ-9 ninth item was assessed with the fixed effect of time adjusting for baseline scores and receiving therapy for depression. Similar separate ordinal logistic regression models were used to assess the relationship program logins, page views, and lessons completed, adjusting for baseline scores, therapy for depression, and time.

Statistical analyses were performed using SAS software, version 9.4 (SAS/STAT 14.2, SAS Institute Inc). Maximum likelihood estimators allow efficient parameter estimation using only available data under an assumption of missing at random [28-30]. Sensitivity analyses were also conducted using only participants with complete data (ie, complete cases), and no significant differences were found between those who were lost to follow-up/noncompleters and those who completed the trial. Attrition/loss to follow-up was assessed to ensure the key covariates, and baseline measures did not differ from those that completed and those that did not complete the trial. The baseline primary and secondary outcomes each showed no significant differences between trial completers and noncompleters. There were no significant differences found among completers for each outcome variable of interest based on age, gender, marital status, employment, rural-urban commuting area codes, or baseline therapy usage. Due to extremely low sample sizes, variables of race, veteran status, and other or no insurance could not be assessed. The level of significance was set at  $\alpha=.05$  (two-tailed), and the Bonferroni method was implemented to control false positives over the multiple tests.

## Results

### Participant Characteristics

A total of 181 immediate intervention group participants (iCBT) were included in this longitudinal outcome assessment. As detailed in [Table 1](#), participants were on average 42 years old (SD 12.8); and most were female (88.9%), White (93.9%), and nonveterans (96.1%). A majority was married or in a domestic relationship (56.9%), employed full-time (61.9%), had obtained at least a bachelor's degree (56.9%), and had private health insurance (77.3%). Nearly 15% of participants lived in urban, over 56% in rural, and nearly 29% in isolated communities. Nearly 57% of participants reported receiving clinical care for mental health.



**Table 1.** Baseline characteristics of the analytic sample (N=181).

Characteristics	Values
Age (years), mean (SD)	42.1 (12.8)
Female, n (%)	160 (88.4)
<b>Race, n (%)</b>	
White	170 (93.9)
Other	11 (6.1)
<b>Marital status, n (%)</b>	
Single	78 (43.1)
Married/domestic partnership	103 (56.9)
<b>Employment status, n (%)</b>	
Employed full-time	112 (61.9)
Employed part-time	39 (21.6)
Unemployed/retired/student	30 (16.8)
Veteran, n (%)	7 (3.9)
<b>Education, n (%)</b>	
Some college or less	78 (43.1)
Bachelor's degree	61 (33.7)
Graduate or professional degree	42 (23.2)
<b>Health insurance, n (%)</b>	
Private	140 (77.3)
Public	32 (17.7)
Other	5 (2.8)
None	4 (2.2)
<b>Rural classification<sup>a</sup>, n (%)</b>	
Urban	27 (14.9)
Large rural	42 (23.2)
Small rural	60 (33.2)
Isolated	52 (28.7)
Receiving mental health treatment <sup>b</sup> , n (%)	103 (56.9)

<sup>a</sup>Defined using the rural-urban commuting area codes.

<sup>b</sup>Defined as receiving any clinical care or taking medication(s) for depression symptoms.

## Clinical Outcomes

The assessment response rates for the 8-week, 6-month, and 12-month outcomes were 58.6%, 50.3%, and 51.4%, respectively. Thus, the respective attrition rates were 41.4%, 49.7%, and 48.6%. To assess remission and relapse, participants' PHQ-9 scores were assessed for changes from baseline to 8 weeks, 6 months, and 12 months. Of the 107 participants who completed week 8 assessments, 42 (39.3%) achieved remission. Among the 42 participants who achieved remission, 22 (52.4%) and 24 (57.1%) maintained remission at 6 months and 12 months, respectively. Only 4 participants (9.5%) who had

achieved remission at week 8 subsequently relapsed at 6 months (n=2) and 12 months (n=2).

Further, 72.9% (78/107) participants had moderate or greater depression symptoms at baseline (PHQ-9 score  $\geq 10$ ). From this subgroup of 78, 28 (35.9%) achieved remission at week 8. Of these 28 participants who achieved remission within this subgroup, 11 (39.3%) and 13 (46.4%) maintained remission at 6 months and 12 months, respectively. Only 4 (14.2%) participants who had achieved remission at week 8 subsequently relapsed at 6 months (n=2) and 12 months (n=2).

Longitudinal mean outcome scores and adherence metrics are presented in [Table 2](#).

**Table 2.** Longitudinal mean trend of clinical and adherence measures.

Measures	Baseline, mean (SD)	8 weeks, mean (SD)	6 months, mean (SD)	12 months, mean (SD)	Significance <sup>a</sup>	
					<i>F</i> test ( <i>df</i> )	<i>P</i> value (adj)
Primary outcome measure (Depression symptom severity) <sup>b</sup>	13.7 (5.0)	7.2 (5.1)	7.0 (5.4)	7.2 (5.6)	121.6 (3)	<.001
<b>Secondary outcome measures</b>						
Anxiety symptom severity <sup>c</sup>	10.3 (4.7)	6.4 (4.7)	6.2 (4.8)	6.0 (5.4)	50.4 (3)	<.001
Work and social functioning <sup>d</sup>	20.2 (8.0)	15.2 (9.4)	13.7 (9.7)	13.3 (10.5)	29.1 (3)	<.001
Resilience <sup>e</sup>	22.2 (6.6)	25.6 (6.1)	26.1 (6.3)	26.5 (6.7)	28.9 (3)	<.001
Suicidal ideation <sup>f</sup> (score ≥1), n (%)	72 (39.8)	13 (12.2)	15 (16.5)	16 (17.2)	48.9 (3) <sup>g</sup>	<.001 <sup>h</sup>
<b>Adherence measures</b>						
Logins (cumulative counts)	N/A <sup>i</sup>	8.9 (9.6)	11.1 (13.3)	11.7 (14.3)	N/A	N/A
Page views (cumulative counts)	N/A	66.3 (70.4)	80.9 (89.1)	84.8 (93.5)	N/A	N/A
Lessons completed (cumulative counts)	N/A	7.7 (7.9)	8.9 (9.1)	9.2 (9.3)	N/A	N/A

<sup>a</sup>*P* value was associated with the test (*F* test, Type 3 tests of fixed effects) of the overall time period (week) difference. *P* values adjusted by the Bonferroni method (adj) are shown.

<sup>b</sup>Patient Health Questionnaire-9 (PHQ-9) score range = 0 to 27.

<sup>c</sup>Generalized Anxiety Disorder Scale-7 (GAD-7) score range = 0 to 21.

<sup>d</sup>Work and Social Adjustment Scale (WSAS) score range = 0 to 40.

<sup>e</sup>Connor-Davidson Resilience Scale-10 (CD-RISC-10) score range = 0 to 40.

<sup>f</sup>PHQ-9 Item 9, Suicidal Ideation, score range = 0 to 3.

<sup>g</sup>Chi-square statistic,  $\chi^2$  (degrees of freedom).

<sup>h</sup>*P* value was associated with the test ( $\chi^2$ , Type 3 generalized estimating equations analysis) of the overall time period (week) difference. *P* value adjusted by the Bonferroni method (adj) is shown.

<sup>i</sup>N/A: not applicable.

Mean effect of treatment over time by outcome measure with 95% CI is illustrated in [Figure 1](#).

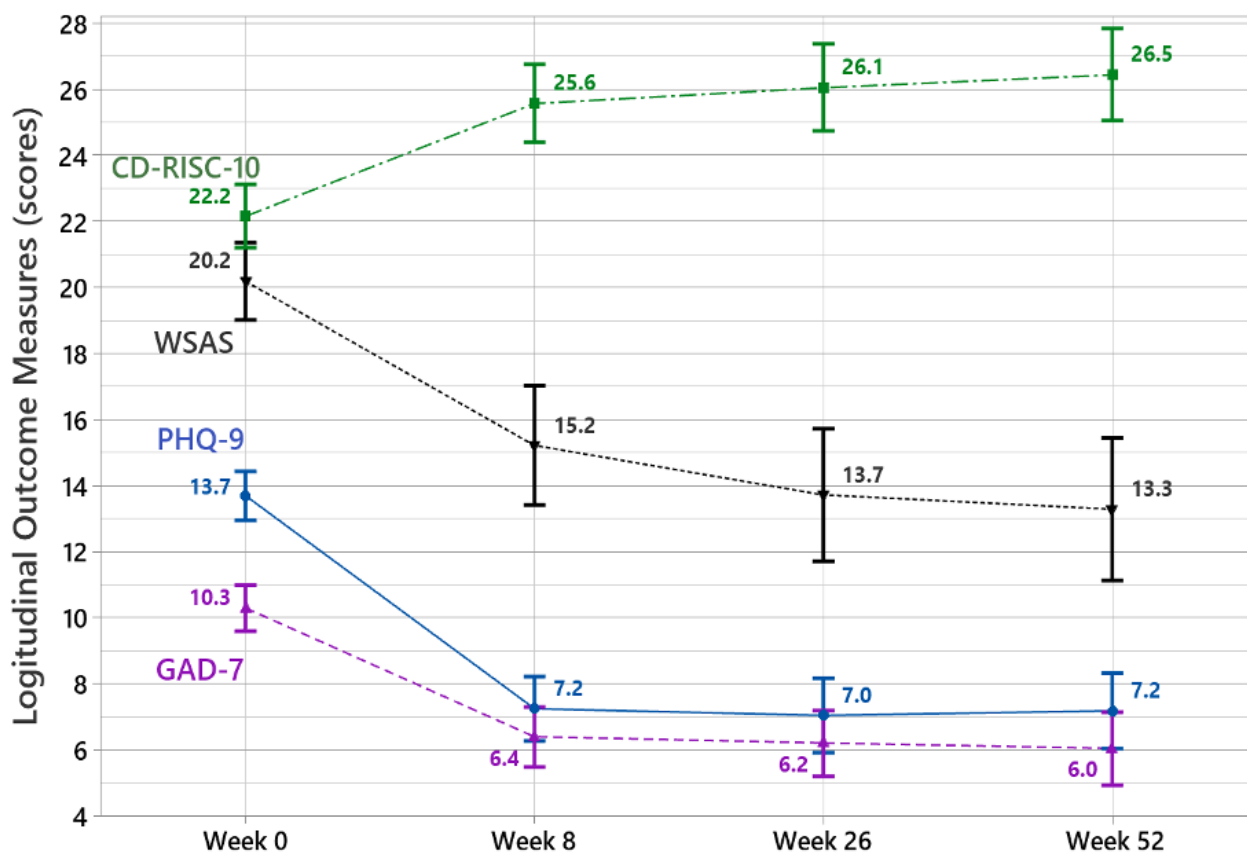
By 8 weeks, significant improvements were observed for all outcome measures. These improvements were maintained at 6 and 12 months. We observed 6-month mean reductions and respective effect sizes in the severity of depression (mean  $-6.3$ ;  $d=1.27$ ) and anxiety symptoms (mean  $-4.1$ ;  $d=0.86$ ). Improvements were also observed in work and social functioning (Mean  $-6.5$ ;  $d=0.73$ ) and resilience (mean  $3.9$ ;  $d=0.60$ ). A total of 23% fewer participants endorsed suicidal ideation (PHQ-9 ninth item score  $>1$ ) at 6 months (16.5%) and 12 months (17.2%) compared to baseline (39.8%).

We observed 12-month mean reductions and respective effect sizes in the severity of depression (mean  $-6.5$ ;  $d=1.23$ ) and anxiety symptoms (mean  $-4.3$ ;  $d=0.93$ ). Improvements were

also observed in work and social functioning (mean  $-6.9$ ;  $d=0.76$ ) and resilience (mean  $4.3$ ;  $d=0.62$ ). Marked decreases were observed on suicidal ideation (PHQ-9 ninth item score  $>1$ ) from baseline (39.8%) to 6 months (16.5%) and 12 months (17.2%). Longitudinal trends from baseline for all outcome measures were statistically significant ( $P<.001$ ).

[Table 3](#) presents the effects of program adherence on each clinical measure. The number of lessons completed was significantly associated to lower PHQ-9 ( $P<.001$ ), GAD-7 ( $P<.001$ ), and WSAS scores ( $P<.001$ ) and higher CD-RISC-10 scores ( $P<.001$ ). The number of page views was significantly associated to lower PHQ-9 ( $P<.001$ ), GAD-7 ( $P<.001$ ), and WSAS scores ( $P=.02$ ) and higher CD-RISC-10 scores ( $P<.001$ ). The number of logins was significantly associated only with the PHQ-9 ( $P=.02$ ). No adherence metrics were significantly associated with suicidal ideation.

**Figure 1.** Mean effect of treatment over time by clinical measure with 95% CI. CD-RISC-10: Conner-Davidson Resilience Scale-10; GAD-7: Generalized Anxiety Disorder scale-7; PHQ-9: Patient Health Questionnaire-9; WSAS: Work and Social Adjustment Scale.



**Table 3.** Program adherence metric effects on outcome measures.

Adherence metric	PHQ-9 <sup>a</sup>		GADS-7 <sup>b</sup>		WSAS <sup>c</sup>		CD-RISC-10 <sup>d</sup>		PHQ-9, item 9 <sup>e</sup>	
	<i>F</i> test ( <i>df</i> )	<i>P</i> value (adj) <sup>f</sup>	<i>F</i> test ( <i>df</i> )	<i>P</i> value (adj) <sup>f</sup>	<i>F</i> test ( <i>df</i> )	<i>P</i> value (adj) <sup>f</sup>	<i>F</i> test ( <i>df</i> )	<i>P</i> value (adj) <sup>f</sup>	$\chi^2$ ( <i>df</i> )	<i>P</i> value (adj) <sup>g</sup>
Logins	7.89 (1)	.02	3.62 (1)	.23	1.00 (1)	>.99	1.09 (1)	>.99	1.18 (1)	>.99
Page views	31.3 (1)	<.001	26.95 (1)	<.001	7.72 (1)	.02	22.81 (1)	<.001	4.83 (1)	.11
Lessons completed	41.42 (1)	<.001	45.65 (1)	<.001	14.72 (1)	<.001	31.69 (1)	<.001	4.57 (1)	.13

<sup>a</sup>Patient Health Questionnaire-9 (PHQ-9) score range = 0 to 27.

<sup>b</sup>Generalized Anxiety Disorder Scale-7 (GAD-7) score range = 0 to 21.

<sup>c</sup>Work and Social Adjustment Scale (WSAS) score range = 0 to 40.

<sup>d</sup>Connor-Davidson Resilience Scale (CD-RISC 10) score range = 0 to 40.

<sup>e</sup>PHQ-9 Item 9, Suicidal Ideation, score range = 0 to 3.

<sup>f</sup>*P* value was associated with the test (*F* test, Type III Tests of Fixed Effects) of the overall time period (week) difference. *P* values adjusted by the Bonferroni method (adj) are shown.

<sup>g</sup>*P* value was associated with the test ( $\chi^2$ , Type 3 generalized estimating equations analysis) of the overall time period (week) difference. *P* values are adjusted by the Bonferroni method (adj) are shown.

## Discussion

### Principal Results

This study evaluated the long-term (6- and 12-month) outcomes of a fully automated self-guided (no supportive contacts by email, text, telephone, or face-to-face) video-centric iCBT intervention called Thrive in a rural US community. We believe this is the first study to assess the long-term impacts of an iCBT

program within an adult population in rural United States. These analyses focused on participants receiving immediate access to the Thrive intervention. Over the course of 8 weeks, over a third of participants with 8-week data achieved remission, and over half of those maintained remission at 6 months and 12 months. A very low relapse rate was one of the noteworthy observations of this study. In regard to long-term outcomes of primary and secondary measures, study findings demonstrated mean improvements in depression and anxiety symptoms, work and

social functioning, and resilience from baseline to 8 weeks. These improvements were sustained at 6 and 12 months. Comparable sustained improvements were also observed with decreased percentages of participants reporting suicidal ideation.

In regard to our adherence analyses, the number of page views and the number of lessons completed most consistently predicted greater sustained positive effects on all outcome measures. Both page views and lessons completed are reasonable markers to assess the program adherence with self-guided iCBT interventions like Thrive, as they indicate the extent to which the users progress through the program. In contrast, the number logins and progress through the program will expectedly vary because some users tend to spend more blocks of time in the program compared to others.

### Comparison With Prior Work

There is limited evidence on the long-term impacts of self-guided iCBT interventions. To our knowledge, only a few studies have examined long-term effects on depression. Mira and colleagues' study [19] with adults residing in Spain found similar 12-month sustained effects of an iCBT intervention. Clarke and colleagues' study [20] examined 8-month effects of an iCBT intervention on depression among members of the US Pacific Northwest Kaiser Permanente organization.

Substantial evidence exists regarding CBT's short- and long-term effectiveness for depression and anxiety [31-33]. A growing number of studies report evidence of equivalent effectiveness of iCBT compared with clinician-delivered CBT [7]. Lorenzo-Luaces and colleagues' [34] meta-analysis concluded comparable effectiveness of self-guided iCBT with that of antidepressants and in-person psychotherapy. In our trial, Thrive's 8-week RCT between group (intervention vs control) effect size of 0.63 compared well with "mostly 8-week clinical trials of antidepressants for adults with unipolar major depression" between group (antidepressant vs placebo) effect sizes of 0.27 (Hamilton Depression Rating Scale) and 0.30 (Montgomery-Asberg Depression Rating Scale) in 109 antidepressant medication RCTs [35]. For longer-term outcomes, 12-month relapse rates from remission for Thrive at 8 weeks and STAR\*D Phase I citalopram at 12 weeks [36] are 14.2% versus 33.5%, respectively. This comparison includes only Thrive participants with PHQ-9 baseline score 10, consistent with a diagnosis of major depressive disorder.

Adherence in iCBT needs further exploration. In Beintner and colleagues' review [21], most studies (85%) reported at least 1 adherence indicator, yet adherence metrics varied widely across this literature. The 10 most commonly reported adherence metrics were as follows: full intervention completion; completion of a minimum number of sessions/modules; average number of completed sessions/modules; specified point of discontinuation of the intervention ("dropout"); dropout without specifying a time point; number of participants who were randomized to an intervention group but never logged on; number of times a participant logged on to access the intervention program; total time spent on the program; number of entries into a diary; and number of messages sent to a coach. Recognizing that the specific metric(s) chosen will need to be appropriate to the specific components or delivery formats of

online interventions, Beintner and colleagues [21] further recommend the use of multiple adherence metrics, which all studies should report on adherence, providing detailed information on its operationalization.

In our study, we analyzed data on the number of logins, page views, and lessons completed. Page views and lessons completed were the consistent significant predictors of our outcome measures with the exception of suicidal ideation. Cuijpers and colleagues' [33] meta-analysis of CBT depression studies assessed the number of sessions (comparable to lessons completed in Thrive) as an adherence measure to determine a dose-response effect. In contrast to our findings, they found no significant relationship with study effect sizes. Given the relative infancy of iCBT platforms, it is imperative that future studies critically devise relevant adherence metrics that fit the type of medium.

### Limitations

Our findings need to be considered in light of several limitations. As commonly observed in iCBT studies, assessment completion rates were low with 50.3% and 51.4% of participants completing assessments at 6 and 12 months, respectively. Thus, our results may be skewed due to underlying responder biases. Within-group analyses are limited in that there is no control group with which to compare findings. Relying solely on self-assessments, a common practice in iCBT studies, is a potential weakness of the study; however, the use of validated, widely used instruments largely addressed this issue. The PHQ-9, GAD-7, and WSAS measures correlate well with clinician-administered instruments [24,25,37]; furthermore, they have been shown to be sensitive to treatment effects [26,38,39]. Additionally, self-assessments may underestimate the effect of iCBT compared to clinician-administered assessments [40]. As a community-based trial, our findings cannot be generalized to health care settings. In regard to adherence metrics, we limited our analyses to include the number of logins, page views, and lessons completed. Our study was not originally designed as a dose-response analysis; and therefore, our findings are limited to our post hoc analyses of adherence.

### Conclusions

To our knowledge, this study is the first to demonstrate both short- and long-term positive impacts of a self-guided iCBT intervention on depression and anxiety symptoms, work and social functioning, and resilience among rural adults in the United States. iCBT interventions, such as Thrive, have the potential to provide help and clinical benefit to people with substantial barriers to traditional forms of care; they also present a cost-effective alternative [41] with additional benefits of confidentiality that may not be possible for those seeking traditional care in small rural communities. That iCBT interventions have the potential for greater accessibility, privacy, and affordability in many rural areas compared to in-person psychological treatment is particularly encouraging. Further research is warranted to identify effective and cost-effective dissemination strategies for expanded reach and to understand adoption patterns of iCBT and other internet-based therapies in

rural American communities for ensuring optimal uptake of these promising interventions [42].

## Acknowledgments

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers P20GM103474, U54GM115371, and 5P20GM104417. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors are grateful to Dr Bill Bryan for his tireless efforts to publicize Thrive across the state of Montana and Dr Paul Crits-Christoph of the University of Pennsylvania for his guidance.

## Conflicts of Interest

JG previously held a financial interest in Waypoint Health Innovations, which developed the Thrive intervention evaluated in this work. Waypoint Health Innovations pays him a royalty based on revenue from Thrive use. He no longer has a direct financial interest in Waypoint Health Innovations but does retain a small interest in Waypoint Health Innovations through Healthcare Technology Systems where he is CEO and a shareholder. He is also a consultant to Waypoint on projects outside of the grant supporting this study. The terms of JG's financial relationship with Waypoint Health Innovations have been reviewed by Montana State University, and his involvement with this research project has been approved in accordance with its conflict of interest policies.

### Multimedia Appendix 1

CONSORT (Consolidated Standards of Reporting Trials) flowchart.

[PDF File (Adobe PDF File), 103 KB - [jmir\\_v22i10e21336\\_app1.pdf](#)]

### Multimedia Appendix 2

Safety Protocol.

[PDF File (Adobe PDF File), 114 KB - [jmir\\_v22i10e21336\\_app2.pdf](#)]

### Multimedia Appendix 3

Consort-eHealth.

[PDF File (Adobe PDF File), 438 KB - [jmir\\_v22i10e21336\\_app3.pdf](#)]

## References

1. Chand S, Ravi C, Chakkamparambil B, Prasad A, Vora A. CBT for depression: What the evidence says. *Current Psychiatry* 2018;17(9):14-23.
2. Deacon BJ, Abramowitz JS. Cognitive and behavioral treatments for anxiety disorders: a review of meta-analytic findings. *J Clin Psychol* 2004 Apr;60(4):429-441. [doi: [10.1002/jclp.10255](#)] [Medline: [15022272](#)]
3. Selmi PM, Klein MH, Greist JH, Sorrell SP, Erdman HP. Computer-administered cognitive-behavioral therapy for depression. *Am J Psychiatry* 1990 Jan;147(1):51-56. [doi: [10.1176/ajp.147.1.51](#)] [Medline: [2403473](#)]
4. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms: A Meta-analysis of Individual Participant Data. *JAMA Psychiatry* 2017 Apr 01;74(4):351-359. [doi: [10.1001/jamapsychiatry.2017.0044](#)] [Medline: [28241179](#)]
5. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. *Ann Behav Med* 2006 Dec;32(3):254-258. [doi: [10.1207/s15324796abm3203\\_12](#)] [Medline: [17107299](#)]
6. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. *PLoS One* 2010;5(10):e13196 [FREE Full text] [doi: [10.1371/journal.pone.0013196](#)] [Medline: [20967242](#)]
7. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. *World Psychiatry* 2014 Oct;13(3):288-295 [FREE Full text] [doi: [10.1002/wps.20151](#)] [Medline: [25273302](#)]
8. Schure MB, Howard M, Bailey SJ, Bryan B, Greist J. Exploring Perceptions of a Computerized Cognitive Behavior Therapy Program in a U.S. Rural Western State. *Rural Ment Health* 2018 Jul;42(3-4):174-183. [doi: [10.1037/rmh0000102](#)] [Medline: [30740195](#)]
9. Schure MB, Lindow JC, Greist JH, Nakonezny PA, Bailey SJ, Bryan WL, et al. Use of a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention in a Community Population of Adults With Depression Symptoms: Randomized Controlled Trial. *J Med Internet Res* 2019 Nov 18;21(11):e14754 [FREE Full text] [doi: [10.2196/14754](#)] [Medline: [31738173](#)]



10. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health* 2017 Jun 06;4(2):e19 [FREE Full text] [doi: [10.2196/mental.7785](https://doi.org/10.2196/mental.7785)] [Medline: [28588005](https://pubmed.ncbi.nlm.nih.gov/28588005/)]
11. Robert WJF. Life in Rural America: Part II. Harvard School of Public Health. 2019. URL: <https://www.rwjf.org/en/library/research/2019/05/life-in-rural-america--part-ii.html> [accessed 2020-06-01]
12. Scheil-Adlung X. Global evidence on inequities in rural health protection: new data on rural deficits in health coverage for 174 countries. World Health Organization. 2015. URL: <https://reliefweb.int/report/world/global-evidence-inequities-rural-health-protection-new-data-rural-deficits-health> [accessed 2020-05-02]
13. Titov N, Fogliati VJ, Staples LG, Gandy M, Johnston L, Wootton B, et al. Treating anxiety and depression in older adults: randomised controlled trial comparing guided, self-guided internet-delivered cognitive-behavioural therapy. *BJPsych Open* 2016 Jan;2(1):50-58 [FREE Full text] [doi: [10.1192/bjpo.bp.115.002139](https://doi.org/10.1192/bjpo.bp.115.002139)] [Medline: [27703754](https://pubmed.ncbi.nlm.nih.gov/27703754/)]
14. Mental Health Care Health Professional Shortage Areas (HPSAs). Kaiser Family Foundation. 2018 Dec 21. URL: <https://www.kff.org/other/state-indicator/mental-health-care-health-professional-shortage-areas-hpsas> [accessed 2019-12-05]
15. Larson JE, Corrigan PW. Psychotherapy for self-stigma among rural clients. *J Clin Psychol* 2010 May;66(5):524-536. [doi: [10.1002/jclp.20679](https://doi.org/10.1002/jclp.20679)] [Medline: [20222119](https://pubmed.ncbi.nlm.nih.gov/20222119/)]
16. Heitkamp T, Nielsen S, Schroeder S. Mountain Health Technology Transfer Network. Promoting positive mental health in rural schools. 2019 Aug 12. URL: <https://mhhtcnetwork.org/sites/default/files/2019-08/promoting-positive-mental-health-in-rural-schools.pdf> [accessed 2020-08-04]
17. De Silva T, Prakash A, Yarlagadda S, Johns MD, Sandy K, Hansen V, et al. General practitioners' experiences and perceptions of mild moderate depression management and factors influencing effective service delivery in rural Australian communities: a qualitative study. *Int J Ment Health Syst* 2017;11:54 [FREE Full text] [doi: [10.1186/s13033-017-0159-x](https://doi.org/10.1186/s13033-017-0159-x)] [Medline: [28936233](https://pubmed.ncbi.nlm.nih.gov/28936233/)]
18. Hull MJ, Fennell KM, Vallury K, Jones M, Dollman J. A comparison of barriers to mental health support-seeking among farming and non-farming adults in rural South Australia. *Aust J Rural Health* 2017 Dec;25(6):347-353. [doi: [10.1111/ajr.12352](https://doi.org/10.1111/ajr.12352)] [Medline: [28618088](https://pubmed.ncbi.nlm.nih.gov/28618088/)]
19. Mira A, Bretón-López J, García-Palacios A, Quero S, Baños RM, Botella C. An Internet-based program for depressive symptoms using human and automated support: a randomized controlled trial. *Neuropsychiatr Dis Treat* 2017;13:987-1006 [FREE Full text] [doi: [10.2147/NDT.S130994](https://doi.org/10.2147/NDT.S130994)] [Medline: [28408833](https://pubmed.ncbi.nlm.nih.gov/28408833/)]
20. Clarke G, Reid E, Eubanks D, O'Connor E, DeBar LL, Kelleher C, et al. Overcoming depression on the Internet (ODIN): a randomized controlled trial of an Internet depression skills intervention program. *J Med Internet Res* 2002 Dec;4(3):E14 [FREE Full text] [doi: [10.2196/jmir.4.3.e14](https://doi.org/10.2196/jmir.4.3.e14)] [Medline: [12554545](https://pubmed.ncbi.nlm.nih.gov/12554545/)]
21. Beintner I, Vollert B, Zarski A, Bolinski F, Musiat P, Görlich D, et al. Adherence Reporting in Randomized Controlled Trials Examining Manualized Multisession Online Interventions: Systematic Review of Practices and Proposal for Reporting Standards. *J Med Internet Res* 2019 Aug 15;21(8):e14181 [FREE Full text] [doi: [10.2196/14181](https://doi.org/10.2196/14181)] [Medline: [31414664](https://pubmed.ncbi.nlm.nih.gov/31414664/)]
22. Thrive. Waypoint Health Innovations. 2018. URL: <https://waypointhealth.com/thrive/> [accessed 2020-07-01]
23. Feeling down, anxious, or stressed? Free online help is here for Montanans. Thrive for Montana. 2018 Feb 28. URL: <https://thriveformontana.com> [accessed 2020-07-01]
24. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001 Sep;16(9):606-613 [FREE Full text] [doi: [10.1046/j.1525-1497.2001.016009606.x](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
25. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006 May 22;166(10):1092-1097. [doi: [10.1001/archinte.166.10.1092](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]
26. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *Br J Psychiatry* 2002 May;180:461-464. [doi: [10.1192/bjp.180.5.461](https://doi.org/10.1192/bjp.180.5.461)] [Medline: [11983645](https://pubmed.ncbi.nlm.nih.gov/11983645/)]
27. Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the Connor-davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. *J Trauma Stress* 2007 Dec;20(6):1019-1028. [doi: [10.1002/jts.20271](https://doi.org/10.1002/jts.20271)] [Medline: [18157881](https://pubmed.ncbi.nlm.nih.gov/18157881/)]
28. Wolfinger R, Chang M. Comparing the SAS GLM and MIXED procedures for repeated measures. In: Comparing the SAS GLMMIXED procedures for repeated measures. Presented at: Proceedings of the Twentieth Annual SAS Users Group Conference; 1998 Presented at: Twentieth Annual SAS User Group; 1998; Cary, NC.
29. Little RJ. Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association* 1995 Sep;90(431):1112. [doi: [10.2307/2291350](https://doi.org/10.2307/2291350)]
30. Rubin DB. Inference and Missing Data. *Biometrika* 1976 Dec;63(3):581-592. [doi: [10.2307/2335739](https://doi.org/10.2307/2335739)]
31. Hollon SD, Stewart MO, Strunk D. Enduring effects for cognitive behavior therapy in the treatment of depression and anxiety. *Annu Rev Psychol* 2006;57:285-315. [doi: [10.1146/annurev.psych.57.102904.190044](https://doi.org/10.1146/annurev.psych.57.102904.190044)] [Medline: [16318597](https://pubmed.ncbi.nlm.nih.gov/16318597/)]
32. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin Psychol Rev* 2006 Jan;26(1):17-31. [doi: [10.1016/j.cpr.2005.07.003](https://doi.org/10.1016/j.cpr.2005.07.003)] [Medline: [16199119](https://pubmed.ncbi.nlm.nih.gov/16199119/)]
33. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Can J Psychiatry* 2013 Jul;58(7):376-385. [doi: [10.1177/070674371305800702](https://doi.org/10.1177/070674371305800702)] [Medline: [23870719](https://pubmed.ncbi.nlm.nih.gov/23870719/)]

34. Lorenzo-Luaces L, Johns E, Keefe JR. The Generalizability of Randomized Controlled Trials of Self-Guided Internet-Based Cognitive Behavioral Therapy for Depressive Symptoms: Systematic Review and Meta-Regression Analysis. *J Med Internet Res* 2018 Nov 09;20(11):e10113 [FREE Full text] [doi: [10.2196/10113](https://doi.org/10.2196/10113)] [Medline: [30413400](https://pubmed.ncbi.nlm.nih.gov/30413400/)]
35. Hengartner MP, Jakobsen JC, Sørensen A, Plödel M. Efficacy of new-generation antidepressants assessed with the Montgomery-Asberg Depression Rating Scale, the gold standard clinician rating scale: A meta-analysis of randomised placebo-controlled trials. *PLoS One* 2020;15(2):e0229381 [FREE Full text] [doi: [10.1371/journal.pone.0229381](https://doi.org/10.1371/journal.pone.0229381)] [Medline: [32101579](https://pubmed.ncbi.nlm.nih.gov/32101579/)]
36. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR\*D report. *Am J Psychiatry* 2006 Nov;163(11):1905-1917. [doi: [10.1176/ajp.2006.163.11.1905](https://doi.org/10.1176/ajp.2006.163.11.1905)] [Medline: [17074942](https://pubmed.ncbi.nlm.nih.gov/17074942/)]
37. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Med Care* 2008 Mar;46(3):266-274. [doi: [10.1097/MLR.0b013e318160d093](https://doi.org/10.1097/MLR.0b013e318160d093)] [Medline: [18388841](https://pubmed.ncbi.nlm.nih.gov/18388841/)]
38. Zimmerman M, Walsh E, Friedman M, Boerescu DA, Attiullah N. Are self-report scales as effective as clinician rating scales in measuring treatment response in routine clinical practice? *J Affect Disord* 2018 Jan 01;225:449-452. [doi: [10.1016/j.jad.2017.08.024](https://doi.org/10.1016/j.jad.2017.08.024)] [Medline: [28858659](https://pubmed.ncbi.nlm.nih.gov/28858659/)]
39. Kertz S, Bigda-Peyton J, Bjorgvinsson T. Validity of the Generalized Anxiety Disorder-7 scale in an acute psychiatric sample. *Clin Psychol Psychother* 2013;20(5):456-464. [doi: [10.1002/cpp.1802](https://doi.org/10.1002/cpp.1802)] [Medline: [22593009](https://pubmed.ncbi.nlm.nih.gov/22593009/)]
40. Cuijpers P, Li J, Hofmann SG, Andersson G. Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: a meta-analysis. *Clin Psychol Rev* 2010 Aug;30(6):768-778. [doi: [10.1016/j.cpr.2010.06.001](https://doi.org/10.1016/j.cpr.2010.06.001)] [Medline: [20619943](https://pubmed.ncbi.nlm.nih.gov/20619943/)]
41. Thase M, McCrone P, Barrett M, Eells T, Wisniewski S, Balasubramani G, et al. Improving Cost-effectiveness and Access to Cognitive Behavior Therapy for Depression: Providing Remote-Ready, Computer-Assisted Psychotherapy in Times of Crisis and Beyond. *Psychother Psychosom* 2020;89(5):307-313 [FREE Full text] [doi: [10.1159/000508143](https://doi.org/10.1159/000508143)] [Medline: [32396917](https://pubmed.ncbi.nlm.nih.gov/32396917/)]
42. Taylor CB, Ruzek JI, Fitzsimmons-Craft EE, Sadeh-Sharvit S, Topococo N, Weissman RS, et al. Using Digital Technology to Reduce the Prevalence of Mental Health Disorders in Populations: Time for a New Approach. *J Med Internet Res* 2020 Jul 24;22(7):e17493 [FREE Full text] [doi: [10.2196/17493](https://doi.org/10.2196/17493)] [Medline: [32706665](https://pubmed.ncbi.nlm.nih.gov/32706665/)]

## Abbreviations

**CBT:** cognitive behavior therapy  
**CD-RISC-10:** Conner-Davidson Resilience Scale-10  
**CONSORT:** Consolidated Standards of Reporting Trials  
**GAD-7:** Generalized Anxiety Disorder scale-7  
**iCBT:** internet-based cognitive behavior therapy  
**IRB:** institutional review board  
**PHQ-9:** Patient Health Questionnaire-9  
**RCT:** randomized controlled trial  
**WLC:** waitlist control  
**WSAS:** Work and Social Adjustment Scale

*Edited by G Eysenbach; submitted 11.06.20; peer-reviewed by S Pit, MDG Pimentel, K Mathiasen; comments to author 01.07.20; revised version received 14.08.20; accepted 02.09.20; published 02.10.20.*

### *Please cite as:*

Schure M, McCrory B, Tuchscherer Franklin K, Greist J, Weissman RS  
 Twelve-Month Follow-Up to a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention for Rural Adults With Depression Symptoms: Single-Arm Longitudinal Study  
*J Med Internet Res* 2020;22(10):e21336  
 URL: <https://www.jmir.org/2020/10/e21336>  
 doi: [10.2196/21336](https://doi.org/10.2196/21336)  
 PMID: [33006561](https://pubmed.ncbi.nlm.nih.gov/33006561/)

©Mark Schure, Bernadette McCrory, Kathryn Tuchscherer Franklin, John Greist, Ruth Striegel Weissman. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 02.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted

use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Mechanisms of Action of a Web-Based Intervention With Health Professional Support to Increase Adherence to Nebulizer Treatments in Adults With Cystic Fibrosis: Qualitative Interview Study

Sarah J Drabble<sup>1</sup>, PhD; Alicia O'Cathain<sup>1</sup>, PhD; Alexander J Scott<sup>1</sup>, PhD; Madelynne A Arden<sup>2</sup>, PhD; Samuel Keating<sup>1</sup>, BSc; Marlene Hutchings<sup>3</sup>, BSc; Chin Maguire<sup>1</sup>, MPH; Martin Wildman<sup>3</sup>, MBChB, PhD

<sup>1</sup>School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom

<sup>2</sup>Centre for Behavioural Science and Applied Psychology, Sheffield Hallam University, Sheffield, United Kingdom

<sup>3</sup>Sheffield Adult CF Centre, Sheffield Teaching Hospitals, Sheffield, United Kingdom

**Corresponding Author:**

Sarah J Drabble, PhD

School of Health and Related Research

University of Sheffield

Regent Court

30 Regent St

Sheffield, S1 4DA

United Kingdom

Phone: 44 07711973986

Email: [s.j.drabble@sheffield.ac.uk](mailto:s.j.drabble@sheffield.ac.uk)

## Abstract

**Background:** Adherence to nebulizer treatments in adults with cystic fibrosis (CF) is often low. A new complex intervention to help adults with CF increase their adherence to nebulizer treatments was tested in a pilot randomized controlled trial (RCT) in 2 UK CF centers. Patients used a nebulizer with electronic monitoring capabilities that transferred data automatically to a digital platform (CFHealthHub) to monitor adherence over time and to a tailored website to display graphs of adherence data and educational and problem-solving information about adherence. A trained interventionist helped patients identify ways to increase their adherence.

**Objective:** This study aims to explore the mechanisms of action underpinning the intervention.

**Methods:** A qualitative interview study was conducted concurrently with a pilot RCT. In total, 25 semistructured interviews were conducted with 3 interventionists at 2 time points, 14 patients in the intervention arm of the trial, and 5 members of the multidisciplinary teams offering wider care to patients. A framework approach was used for the analysis.

**Results:** The intervention was informed by a theoretical framework of behavior change. There was evidence of the expected behavior change mechanisms of action. There was also evidence of additional mechanisms of action associated with effective telehealth interventions for self-management support: relationships, visibility, and fit. Patients described how building a relationship with the interventionist through face-to-face visits with someone who cared about them and their progress helped them to consider ways of increasing adherence to medication. Rather than seeing the visibility of adherence data to clinicians as problematic, patients found this motivating, particularly if they received praise about progress made. The intervention was tailored to individuals, but there were challenges in how the intervention fitted into some patients' busy lives when delivered through a desktop computer.

**Conclusions:** The mechanisms of action associated with effective telehealth interventions for self-management operated within this new intervention. The intervention was modified to strengthen mechanisms of action based on these findings, for example, delivery through an app accessed via mobile phones and then tested in an RCT in 19 UK CF centers.

**Trial Registration:** International Standard Randomized Controlled Trial Number 13076797; <http://www.isrctn.com/ISRCTN13076797>

(*J Med Internet Res* 2020;22(10):e16782) doi:[10.2196/16782](https://doi.org/10.2196/16782)

**KEYWORDS**

United Kingdom; cystic fibrosis; adherence; compliance; health behavior; psychological theory; process evaluation; qualitative research; interview; patient portals; telehealth

**Introduction****Background**

Adherence to medication for chronic conditions is a complex issue that is sometimes addressed by digital, web-based, mobile, and telehealth interventions [1-3]. We developed a new intervention involving a web-based adherence monitoring system with health professional support to help adults with cystic fibrosis (CF) increase their adherence to nebulizer treatments. Although not promoted as telehealth, the intervention has key components of web-based monitoring and web-based behavior change materials for use with the chronic condition of CF. CF is a genetic life-limiting condition affecting approximately 100,000 people worldwide [4] in which mucus builds up in the lungs, digestive system, and other organs, leading to difficulty in breathing, respiratory infections, and ultimately death. Patients with CF need to take preventative treatments such as antibiotics and mucolytics through nebulizers and airway clearance, often alongside many other treatments creating a complex, onerous treatment regime [5,6]. Adherence to nebulizer treatments is often low, with objectively measured adherence found to average 36% compared with 80% estimated by patients [7]. Low adherence to nebulizer treatments has been linked to worse health outcomes, including decreased lung function, increased pulmonary exacerbations requiring treatment with intravenous antibiotics, and higher service costs [8-11].

We tested the intervention in a feasibility study that included a pilot randomized controlled trial (RCT) and a mixed methods process evaluation. We found that a full-scale RCT was feasible with numbers of pulmonary exacerbations as the primary outcome. The pilot RCT identified that mean changes to adherence, a key secondary outcome, were 10% higher in the intervention arm (95% CI -5.2 to 25.2) [12]. The mixed methods process evaluation found that the intervention was feasible and acceptable to patients, interventionists, and health care professionals in the multidisciplinary teams (MDTs) offering wider care to patients with CF. As part of the process evaluation, we conducted qualitative interviews with patients and interventionists. One of the aims of this qualitative research was to consider the proposed mechanisms of action underpinning the intervention in practice. The mechanisms of action explain how the intervention components might work to produce an outcome [13,14] and how any theoretical model might be working within the intervention. Qualitative research can be undertaken during the development and feasibility phases of developing complex interventions [14,15] to help researchers understand how a complex intervention with multiple and interacting components might work [16]. Qualitative research within process evaluations can help researchers understand

mechanisms of action by looking for unanticipated and complex causal pathways through the intervention or unanticipated consequences of the intervention [17,18], which is useful in refining the intervention before a full-scale evaluation.

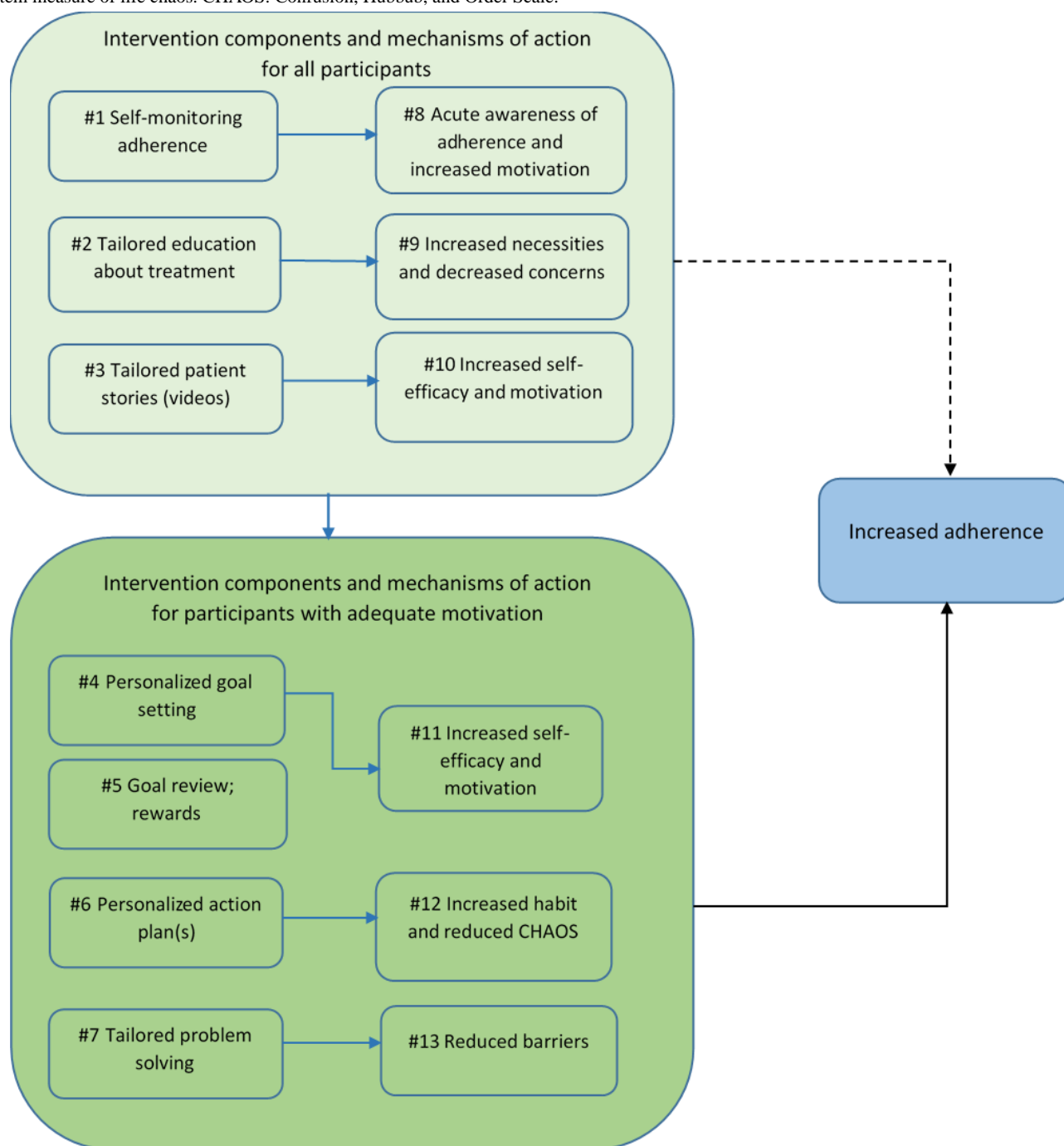
**Description of the Intervention**

The intervention is complex, with a number of interacting components [16]. A study by Arden et al (unpublished data, 2020) provides more details on the intervention and development processes. In summary, an eTrack nebulizer (PARI, Pharma GmbH) with electronic monitoring capabilities sends timestamped inhalation data to a 2net Hub (Qualcomm). This enables real-time monitoring of adherence via securely stored data on CFHealthHub servers. Adherence data are displayed on the CFHealthHub website and are visible to patients and health professionals delivering the intervention. The intervention also comprises a manualized behavior change component that is designed to help patients increase their adherence. This is underpinned by a theoretical model of behavior change (Figure 1) based on the idea that reflective motivation, in which people rationally weigh up the perceived necessity of a behavior against the perceived concerns they have about treatment, underpins adherence behavior. Therefore, some participants would need to address motivation before other strategies, such as action planning, could be successfully used. The model of behavior change is part of a larger logic model [12], informed by qualitative interviews with adults with CF about their adherence behavior [19,20] mapped to the Theoretical Domains Framework [21] and behavior change wheel [22] and developed using a person-based approach [23]. The behavior change intervention is delivered by trained health professionals (interventionists) in a series of face-to-face and telephone meetings with patients. Interventionists help patients to increase their adherence to nebulizer treatments by choosing components of the behavior change intervention to address an individual's unique issues identified through baseline questionnaires and discussions during intervention visits and placed into a toolkit area of the CFHealthHub website. The interventionists in the pilot RCT were clinicians employed to deliver the intervention. The number and location of intervention visits are tailored to patients' needs.

Figure 1 [24] shows how the model splits the behavior change components into components for all participants (including those with low motivation) and for those with adequate motivation (#1 to #7) and how those components link to mechanisms of action (#8 to #13) that lead to increased adherence. Table 1 describes the different components and the proposed mechanisms of action in more detail and how they are delivered within the intervention. The numbers in brackets in the table are linked to the numbers in Figure 1.



**Figure 1.** Theoretical model of behavior change to increase adherence in adults with cystic fibrosis. The Confusion, Hubbub, and Order Scale is a 6-item measure of life chaos. CHAOS: Confusion, Hubbub, and Order Scale.



**Table 1.** Components of the intervention and proposed mechanisms of action.

Components	Description of component and <i>proposed mechanisms of action</i> <sup>a</sup>	Delivery
<b>For all participants</b>		
Self-monitoring adherence (#1)	Patients self-monitor their adherence by reviewing graphs of adherence data by month, week, or time of the day in CFHealthHub. A simple traffic light system quickly indicates if patients have hit their agreed target for treatments (green), have done some treatments but not hit their target (amber), or have completed no treatments (red). <i>This enables patients to understand their adherence patterns leading to awareness of adherence and increased motivation to adhere</i> (#8)	<ul style="list-style-type: none"> <li>Graphs showing red, amber, and green charts of adherence data on CFHealthHub and tables showing the number of treatments completed compared with prescription data</li> </ul>
Tailored education about treatment (#2)	Interventionist identifies issues with understanding treatments and unhelpful beliefs about different aspects of treatment. <i>Patients can see the need to undergo treatment (increased necessities) and reduce any concerns they have about treatments</i> (#9)	<ul style="list-style-type: none"> <li>Baseline questionnaires and discussions with patients used by interventionists to identify issues before visits</li> <li>Education section on CFHealthHub with information about CF<sup>b</sup> and videos about how treatments work and links to other websites. Interventionists use questionnaires and discussions to identify where patients lack knowledge or have misconceptions</li> </ul>
Tailored patient stories (videos; #3)	Other patients with CF talking about how they overcame different barriers to give patients information from peers or someone like them. <i>This allows patients to see how they can find solutions to problems (increased self-efficacy) and that it is possible to change (increased motivation; #10)</i>	<ul style="list-style-type: none"> <li>Section of CFHealthHub includes <i>talking heads</i> videos that address adherence barriers. Interventionists use these with patients if they address an issue that patients experience</li> </ul>
<b>For patients with sufficient motivation</b>		
Personalized goal setting (#4)	Interventionist works with patient to set goals for adherence against prescription and life goals that may help adherence behavior. <i>This allows patients to work toward a target (increased self-efficacy) and see how adherence is linked to their life goals (increased motivation to adhere; #11)</i>	<ul style="list-style-type: none"> <li>Graphs in CFHealthHub measure actual adherence against personalized adherence goals and can be tracked by interventionists</li> <li>Interventionists work with patients during visits to set personalized adherence and life goals if appropriate</li> </ul>
Goal review, rewards (#5)	A review of progress toward goals with the patient helps to increase self-efficacy. <i>Optional notifications of achievements against progress help patients to increase motivation to adhere</i> (#11)	<ul style="list-style-type: none"> <li>Goals are reviewed during discussions between patients and interventionists and barriers are identified</li> <li>Patients receive messages on mobile phones or in CFHealthHub to provide motivation to keep going with behavior changes</li> </ul>
Personalized action plans (#6)	Patients create personalized action plans around daily routines to build adherence <i>that increases habits and reduces chaos</i> (#12)	<ul style="list-style-type: none"> <li>Action plans are made by patients with help from interventionists. Action plans are written in a section in CFHealthHub for patients to refer to</li> </ul>
Tailored problem solving (#7)	Solutions to common problems with adherence, for example, what to do when you go on holiday. <i>These allow patients to solve problems, reducing barriers to adherence</i> (#13)	<ul style="list-style-type: none"> <li>Questionnaires and discussions between interventionists and patients identify practical barriers that make adherence difficult. These are stored on a problem-solving section on CFHealthHub</li> </ul>

<sup>a</sup>Proposed mechanisms of action are shown in italics.

<sup>b</sup>CF: cystic fibrosis.

Whilst considering the proposed mechanisms of action in Table 1 in a qualitative study, we identified 3 mechanisms of action (relationships, visibility, and fit) that Vassilev et al [25] found in a realist review of effective telehealth interventions that explain how telehealth supports self-management of long-term conditions. First, effective telehealth interventions should consider how *relationships* with health professionals and peers are enabled or inhibited. Second, the intervention should enable the *visibility* of symptoms to the self and others and provide

feedback to increase knowledge, motivation, and empower patients, thus enabling reinforcement of behavior change. Finally, successful interventions should *fit*, or at least not disrupt, the everyday lives and routines of patients, including their existing skills and knowledge. The study by Vassilev et al [25] is transferable to this intervention because there are key components of telehealth, namely, the intervention has web-based monitoring and web-based behavior change materials, the intervention is concerned with self-management

(adherence to nebulizer treatments), and the condition being studied, CF, is a long-term condition. Although relationships, fit, and visibility are implicit in the stated intervention components, the ways in which they might work in the intervention as potential mechanisms of action were not fully articulated; thus, their importance to the potential effectiveness of the intervention needed further exploration. In this study, we explore the mechanisms of action operating in the intervention by confirming how the expected mechanisms of action identified in Table 1 worked in the intervention and by showing how important the mechanisms of action (relationships, visibility, and fit) identified in the review by Vassilev et al [25] of telehealth for self-management support were to facilitate the potential effectiveness of this intervention.

## Methods

### Study Design

We undertook a feasibility study in preparation for a full-sized RCT. The study consisted of a mixed methods process

evaluation undertaken concurrently with a pilot RCT in 2 UK CF centers. In total, 33 adults with CF received the intervention for 5 months [12]. The process evaluation addressed the feasibility and acceptability of the intervention and the RCT and the mechanisms of action of the intervention. The qualitative component of the process evaluation allowed exploration of the mechanisms of action at play, and this is the focus of this paper. The study received ethical approval from the London Brent Research Ethics Committee (16/LO/0356).

### Participants

We conducted 25 semistructured face-to-face interviews, with patients in the intervention arm of the RCT (n=14), interventionists (n=3 at 2 time points), and members of the wider MDT offering care to patients with CF (n=5). We interviewed more patients in RCT site 1 (n=8) than in RCT site 2 (n=6), because fewer patients agreed to be interviewed at site 2. Sociodemographic data for patient participants are shown in Table 2.

**Table 2.** Sociodemographic information for patient participants (n=14).

Characteristics	Participants, n (%)
<b>Gender</b>	
Male	9 (64)
Female	5 (36)
<b>Age (years)</b>	
≤18	1 (7)
19-25	6 (43)
31-40	3 (21)
>40	4 (29)
<b>Deprivation (IMD<sup>a,b</sup> quintile)<sup>c</sup></b>	
1	6 (43)
2	1 (7)
3	2 (14)
4	3 (21)
Missing data	2 (14)
<b>Baseline objective adherence<sup>c,d</sup></b>	
High	3 (21)
Moderate	1 (7)
Low	2 (14)
Very low	5 (36)
Missing data	3 (21)

<sup>a</sup>IMD: Index of Multiple Deprivation by Postcode [26].

<sup>b</sup>The Index measures relative deprivation by UK postal code stratifying into 5 quintiles, where 1 is the most affluent and 5 is the most deprived. By entering a postal code, it is possible to obtain an indication of the deprivation level where an individual lives.

<sup>c</sup>The percentage values for deprivation and baseline objective adherence add up to 99% due to rounding issues.

<sup>d</sup>Mean unadjusted objective adherence over the previous 6 months measured from chipped i-neb nebulizers: high (≥80.0%), moderate (50.1%-79.9%), low (25.1%-50.0%), and very low (≤25.0%) [11,27].

We interviewed all 3 interventionists. They had different backgrounds: 2 had worked as physiotherapists within the multidisciplinary health care team who provided care to patients with CF in that center and one was a psychologist who had not previously worked in a CF setting. The MDT members were from a mix of disciplines, including respiratory consultants and physiotherapists.

### Recruitment

During consent for the pilot RCT, 28 of the 33 intervention patients consented to be approached for an interview. We invited participants by letter or email and followed up by telephone or email. Of the 28 available patients, 16 consented to interview, 2 declined, 1 withdrew from the full study, 1 died during the study, we were unable to contact 7 patients, and we did not approach 1 patient because of time constraints. Of the 16 consented, 14 were interviewed and 2 patients canceled their interviews. The 3 interventionists consented to be interviewed twice during the study, toward the beginning and the end of the pilot RCT. MDT participants were recruited through the interventionists at each site. We obtained informed written consent from all participants before the interview.

### Data Collection

Interviews were conducted by SD, AS, and SK. Patients were interviewed face-to-face in their homes or the hospital depending on preference, whereas interventionists and MDT interviews were conducted at the hospital, except one that was by telephone. We used topic guides as the basis for the interviews, with questions about the acceptability of different aspects of the intervention and RCT, and explored which aspects of the intervention participants perceived to be important in increasing adherence to nebulizer treatments (mechanisms of action). Interviews were also open enough to allow patients to describe their approaches to adherence and any challenges they faced. The interviews lasted between 11 and 102 min, averaging 56 min. Interviews were digitally audio-recorded and transcribed verbatim. We removed identifying data before analysis and gave participants an identification number to maintain anonymity.

### Data Analysis

SD and AS conducted the initial coding. We initially derived a coding framework [28] based on the Theoretical Domains Framework categories related to adherence [19] and the key aspects of process evaluation (context, mechanisms, and implementation) [17]. After initial familiarization with the data, SD added the telehealth mechanisms of action by Vassilev to the coding framework, because they were evident in the data [25]. SD and AS read the transcripts and coded them against the framework using NVivo software, allowing for other themes to be identified in the data set that did not fit the framework. After the initial coding, SD worked with MA to discuss the relationship between the telehealth mechanisms by Vassilev (relationships, visibility, and fit) and the expected mechanisms of action based on behavior change theory. Trustworthiness was addressed through: researchers writing reflections after each interview, discussion of the interviews between SD, AS, and SK during data collection, discussion of the findings between SD and AC to develop the analysis, and presentations to the wider research team. We present quotes from patients in the intervention (P), interventionists (INT), and members of the MDT to provide examples of the points we are making.

### Results

In the qualitative interviews, we found evidence of the expected behavior change mechanisms of action. The mechanisms of action that appeared to work during the pilot RCT are presented in Table 3.

In a study of the mechanisms of action in telehealth interventions for self-management support, Vassilev et al [25] emphasized the importance of some mechanisms of action (relationships, visibility, and fit). These mechanisms are implicit in some of the behavior change mechanisms of action discussed earlier, but Vassilev's work identified how important they were in facilitating the potential effectiveness of this intervention. These are discussed below.

**Table 3.** Description of how behavior change components and mechanisms of action worked in the intervention.

Component and mechanism of action	Evidence	Examples of quote
<b>Self-monitoring (#1) leading to acute awareness of adherence and increased motivation to adhere (#8)</b>		
	<ul style="list-style-type: none"> <li>Patients found it motivating to see the green bars on their graphs because it was a quick and easy way to see their progress</li> <li>Some patients reported self-monitoring leading to increased awareness of adherence level although this did not always lead to increased motivation to adhere to treatment or change to adherence behavior</li> </ul>	<ul style="list-style-type: none"> <li>"I'm very much more conscientious of how much I'm doing it and almost kind of realized the importance of doing it a lot more." (P1, site 2)</li> <li>"They kind of say in theory it's a good idea and they like it but somehow they're not doing it." (INT2, site 2)</li> </ul>
<b>Tailored education about treatment (#2) leading to increased necessities and decreased concerns about doing treatments (#9)</b>		
	<ul style="list-style-type: none"> <li>Educational components such as the treatment videos were perceived by clinicians and patients as beneficial and a trusted information source</li> <li>Some patients found it difficult to translate the education into action. The pages were not accessed frequently outside meetings with interventionists</li> </ul>	<ul style="list-style-type: none"> <li>"People have said that its nice having something that you know has been prepared by [...] professionals so you know the information is accurate without it being scary." (INT1, site 2)</li> <li>"I don't know what it is but if someone tells me to do something, yes I take it on board but I'm not very good at putting that into action." (P2, site 1)</li> </ul>
<b>Tailored patient stories (videos; #3) leading to increased self-efficacy and motivation to adhere (#10)</b>		
	<ul style="list-style-type: none"> <li>The <i>talking heads</i> videos did not appear to increase self-efficacy and motivation (#30) for most patients in the sample because they feared comparison with others with CF</li> <li>Interventionists sometimes did not share the videos because they found it difficult to know the content of each video in detail and were concerned about sharing videos that could upset patients</li> </ul>	<ul style="list-style-type: none"> <li>"I don't need to listen to somebody feeling sorry for themselves." (P5, site 1)</li> <li>"Some of them said about the videos of people 'I've not looked because I don't want to see and compare myself to that person'" (INT1, site 1)</li> </ul>
<b>Personalized goal setting (#4) leading to increased self-efficacy and motivation to adhere (#11)</b>		
	<ul style="list-style-type: none"> <li>Some participants found it motivating to set targets that they could work to achieve</li> <li>Some patients preferred to set lower, achievable goals</li> <li>Some low adherers wanted to set unrealistically high goals that were unachievable</li> <li>Some high adherers did not set goals</li> </ul>	<ul style="list-style-type: none"> <li>"Again, the graphs and stats and things like that [...] motivate you, keep you in the right place." (P4, site 2)</li> <li>"They know they are clearly struggling to do what's required so some people were quite happy to be told they could do less [than 100%]." (INT2, site 2)</li> <li>"The one I did this morning, we put her at 100% and I said that's quite high and she was like, no, her words were like, all or nothing." (INT1, site 2)</li> <li>"I do manage [treatments] practically twice a day so [the interventionist] didn't really set me any goals." (P1, site 2)</li> </ul>
<b>Goal review, rewards (#5) leading to increased self-efficacy and motivation to adhere (#11)</b>		
	<ul style="list-style-type: none"> <li>Notifications were unavailable during the pilot, but some patients thought they would help them adhere</li> </ul>	<ul style="list-style-type: none"> <li>"I want to receive reminders—target achieved." (P3, site 2)</li> </ul>
<b>Personalized action plans (#6) leading to increased habits and reduced chaos (#12)</b>		
	<ul style="list-style-type: none"> <li>Interventionists and some patients described how action plans were beneficial for a minority of patients to create a habit or routine, but some patients disliked them because they did not want to form habits</li> </ul>	<ul style="list-style-type: none"> <li>"It [has] helped because now I'm starting to think of things that I can link with [...] it's making me make a conscious effort toward helping my health." (P2, site 2)</li> <li>"My life is up and down you know and saying for every 5 days a week for example I'm going to do this at this time it don't work for me at all." (P2, site 1)</li> </ul>



Component and mechanism of action	Evidence	Examples of quote
	<ul style="list-style-type: none"> <li>Patients and interventionists perceived that the action plans could sometimes feel simplistic</li> </ul>	<ul style="list-style-type: none"> <li>"[A patient] said to be honest I did feel like you know, I'm not a child, that was her reaction but she was quite nice about it" (INT 2, site 2)</li> </ul>
<b>Tailored problem solving (#7) leading to reduced barriers to adherence (#13)</b>		
	<ul style="list-style-type: none"> <li>Most patients interviewed did not access the problem-solving part of the website outside the meetings with interventionists but found the resources useful if they did encounter a problem</li> </ul>	<ul style="list-style-type: none"> <li>"If you're getting in a mess with the equipment, bits and bobs like that—the few things that I investigated on that were really quite helpful." (P4, site 2)</li> </ul>

## Relationships: Building a Relationship With the Interventionist

A trained interventionist delivered the intervention through a series of tailored meetings where patients and interventionists discussed all aspects of adherence behavior. The way the interventionists communicated was intended to encourage open and honest communication about adherence, from which a realistic understanding of the barriers that patients faced in their daily lives could emerge. The importance of the style of communication was addressed in the intervention manual and during the interventionists' training program. In the pilot RCT, patients and interventionists emphasized the importance of these interactions between interventionists and patients in helping them to improve their adherence. In particular, they valued where the interactions occurred and how interventionists communicated with patients.

Meetings in a patient's home were face-to-face, lasting up to an hour. These meetings were highly valued by some patients because patients perceived that the time given and effort made to travel to a patient's home by the interventionist were indications of their interest in the patient's adherence:

*I feel like there's a lot more care in that they're more interested in the patient.* [P8, site 1]

The fact that visits were home-based was also important to some patients because this venue provided a safer space in which patients had time to talk in-depth about adherence. Interventionists also appreciated being able to find out about the lives of patients and some of the barriers they faced by seeing their living conditions. For some patients, this was an important first step in reflecting on their adherence behavior and trying to understand and identify potential barriers. It also demonstrated to patients the importance of focusing on their adherence:

*Physically coming to your home to talk to you about it, makes you think right, okay [...] I'm focused here.*" [P3, site 1]

Face-to-face meetings also gave the opportunity for informal talk, which helped to establish a rapport between patients and interventionists. One interventionist felt that the script-like nature of the intervention could prevent relationship building, so they tried to make the intervention visits flow by including *general chit-chat* and letting patients make conversation at their

own pace. This contributed to patients feeling like the interventionists displayed a caring attitude toward them:

*She's really nice and I think she's very caring, for me to see her face-to-face has helped.* [P8, site 1]

This caring attitude was an important aspect in building a relationship that had the potential to help the patients increase their adherence by allowing them to be open about the barriers they faced to adherence.

## Visibility: Self and Others Monitoring Adherence

Visibility was a key aspect of the intervention. Self-monitoring of real-time adherence data in the form of graphs and tables against personalized goals, in combination with open discussions about adherence behavior (see *Relationships: Building a Relationship With the Interventionist*), was intended to raise the visibility of adherence for patients and clinicians by aiding memory about specific instances of adherence or nonadherence. Discussions could then focus on identifying barriers to adherence, which could be addressed to improve self-efficacy. On the other hand, although monitoring of data by the multidisciplinary health care team was part of the intervention, it was not planned as a central component of the intervention and was not linked to a mechanism of action to increase adherence directly. In the pilot RCT, it was apparent that visibility to the self was an important motivator to improve adherence while visibility to others, that is, people they felt accountable to, could also be a motivator.

In the pilot RCT, we found that visibility to the self-operated as expected through self-monitoring (Table 3). Visibility to others operated when the intervention allowed the interventionist, and with patient consent, their multidisciplinary health care team, to see adherence data. Although most patients we interviewed chose to share their data with their health care team, some patients did not always like being monitored by others. Indeed, some members of the multidisciplinary health care team worried that the intervention could be perceived as unwelcome, described as *Big-Brotherish* by one interviewee (MDT1, site 2), involving the health care team checking up on patients or telling them off.

Patients who chose to share their data described how others monitoring adherence could make them accountable to someone else in their health care team. They also described how others' monitoring motivated them to adhere through a desire to look good in front of others:

*If I go to click and I know that I haven't done it, I know that my doctor knows that I haven't done it. So I suppose in a weird way that's kind of how I'm motivated.* [P4, site 1]

Offering praise to patients to increase their adherence could further motivate patients. For example, although this was not specified as part of the intervention, one of the interventionists would text patients with praise or encouragement to keep going if graphs showed improvements:

*I just periodically look at the tables to see how they're doing and then I text them a response of what I've seen and that's been quite nice you know, they've sort of liked that, I'll say "oh I've just checked and you're doing really well" you know "keep it up" sort of thing.* [INT1, site 2]

*Patient 3 (site 2): I like knowing that I'm being, not monitored and stuff, but because of when that text from [the interventionist] she was like "you're doing absolutely amazing," and because it is hard and [...] bringing in the reality of illness every day and then saying "oh you're doing amazing."*

*Interviewer: Mm so that extra support as well that someone is keeping an eye out*

*Patient 3: Yeah, that you're not alone.*

### Fit: A Better Fit for Others?

The intervention was developed to be tailored to individual patients so that the education about treatments, setting goals and action plans, and problem solving fitted patients' specific needs. Particular content identified by the interventionists could be added to the toolkit area of the website, offering a tailored space for each patient. In addition, the intervention was tailored to patient circumstances by offering options for interventionist visits in the patient's home, hospital, or by telephone.

In the pilot RCT, patients valued this tailoring and believed that it facilitated adherence improvement because not every patient had the same experience or background, so it was important that the intervention allowed patients to focus on the aspects that were most useful to them:

*I think there are different age groups, there's different educational backgrounds, it's different people with different levels of intelligence. So you have to kind of throw it in there that covers all those people and then they opt in and out of what suits them best.* [P3, site 1]

Although patients appreciated the ability to tailor the intervention, it was important to them that this tailoring was patient-led, that is, tailored to the patient's needs rather than the interventionist suggesting or even dictating what targets and goals should be set by the patient:

*Sometimes going with what the patient feels, y'know if you think it might be a good idea for them to have an action plan or a coping plan and some of them just don't really want to do that, even though you think it might be valuable for them.* [INT1, site 2]

One patient with high baseline adherence appreciated how the interventionist recognized that they did not need to set a goal to improve adherence, that is, they did not believe they needed some aspects of the intervention:

*She said there wasn't really a point because I was already so good at doing them.* [P1, site 2]

However, there were problems around fit with patients' perceived needs and lives, with some of the patients seeing the intervention as a better fit for others rather than themselves. Although most patients in the sample believed the intervention was useful, some patients with baseline high adherence did not see the need for setting action plans, but believed that those struggling with adherence would benefit from these. Some patients also perceived that education, problem solving, and video components were more useful for other people, such as younger or newly diagnosed patients with less knowledge:

*Well it would be for, I think for younger people but obviously I know what CF is, I know the ins and outs of it.* [P1, site 2]

Importantly, some patients believed that the intervention did not fit their lifestyles because they were busy and could not access the intervention away from their desktop. For example, although the toolkit area was appreciated by patients, interventionists noticed that they were unlikely to look at the materials outside the meetings. Patients described not having time to log into the website and preferring an app to access the website flexibly when away from home. There was also an issue around fit with priorities in life. For one patient, adherence was low down their list of priorities compared with work, although there were signs that the intervention had started to help them challenge that belief:

*Work is a helluva lot more important and I don't see my health as important as it should be, so my treatments are probably quite low on my ranking of things [...], but obviously I need to start changing that because it's not turning out well.* [P2, site 2]

There was a view among health professionals that the intervention fitted the needs and priorities of patients with moderate baseline adherence rates because high adherers were adhering without the intervention, and low adherers, who perhaps had the most need, were difficult to engage:

*The girl who is the worst adherer out of all those who's got the chaotic life was the one who thought there was absolutely no need to do any plans. But I think it's almost like her lifestyle is that chaotic that it's quite difficult to make a plan. She really didn't see the need which I thought was a shame because I thought it might have been helpful to her.* [INT2, site 2]

## Discussion

### Principal Findings

In this study, we described how components and mechanisms of action for behavior change planned within a new intervention potentially helped patients to increase adherence to nebulizer

treatments. We also described how mechanisms of action for effective telehealth interventions for self-management support (relationships, visibility, and fit) [25] were operationalized and valued by patients using the intervention. Patients described how building a relationship with the interventionist through face-to-face visits in the home with someone who cared about them and their progress helped them to consider ways of increasing adherence to medication. Rather than seeing the visibility of adherence data to clinicians as problematic, patients found this motivating because it made them accountable to someone else, particularly if they received praise about progress made. Although the intervention was tailored to individuals, there were challenges in how the intervention fitted into some patients' busy lives, particularly when delivered through a desktop computer. The intervention was perceived as a better fit for other patients, for example, those who were younger or less knowledgeable. Interventionists identified a middle group of patients with moderate baseline adherence who might benefit more from the intervention.

### Comparison With the Literature

Sitting down face-to-face with an interventionist provided participants with someone who they felt cared about them, providing motivation to increase their adherence. This echoes previous research that has shown how the human aspect of a telehealth intervention helped to build rapport and trust that could help a patient to change their behavior [29]. Although other research has found that self-management may be facilitated by clinic visits [30], in this study, interventionists and patients preferred home visits because they showed that the interventionist was invested in their care, put them at ease, and allowed them to focus on adherence.

The intervention worked through making adherence visible to patients, thereby empowering them to change behaviors. In terms of visibility to others, our study supported research that health care practitioners can be more concerned about negative aspects of surveillance than patients [25], although there was some evidence that those wishing to keep adherence private opted out or withdrew from the pilot RCT. Mohr et al [31] coined the term *supportive accountability* to describe how a trustworthy individual with expertise can support engagement with digital interventions [32]. In our study, this supportive accountability went beyond facilitating engagement and appeared to facilitate outcomes directly. Patients in our study appreciated praise from clinicians when they improved their adherence. Praise is a part of rewards that can facilitate adherence [19,30] by building motivation and self-efficacy [33] and by providing a supportive environment for behavior change [34]. Research has suggested that rewards may work best in combination with identifying barriers and problem solving [35], which both occurred in this intervention.

Vassilev et al [25] suggested that a fit to everyday needs, skills, and daily life was important for intervention success. Research

suggests that the fit may help to increase autonomy and thus motivation to change behavior [32]. Our finding that the intervention sometimes did not fit into people's busy lives is supported by research exploring barriers to adherence in CF where people often cite being too busy to adhere to treatments [19,36-45] and social and work demands creating competing priorities [19,36,46]. This highlights the importance of offering the intervention on an app to enable patients to access adherence data away from their desktop. Some patients also perceived that the educational aspects of the intervention fit better with the needs of younger or newly diagnosed patients who had less knowledge of nebulizer treatments. Research with adolescents and young adults has suggested that they may face similar barriers to adherence as adults [36,38-40,42]. Although the educational aspects of the intervention, such as the videos of how treatments work, may have the potential for use with that group or for those who receive a later diagnosis, we would argue that education alone is unlikely to be sufficient to increase adherence.

We also found perceptions that the intervention fits some groups better than others based on their adherence level with a group of moderate adherers seeming to benefit and engage more than low or high adherers. This could support research into grouping patients by adherence level [11,27], which may have implications for health outcomes [11].

### Strengths and Limitations

This study had some limitations. The sample of patient participants at one of the sites was small because fewer patients agreed to be approached for interview when they were recruited for the pilot RCT. Despite this, we have provided an understanding of patients' experiences of a new intervention and how the mechanisms of action operated in practice.

### Implications

We modified the intervention based on the results of the wider process evaluation, including the insights raised here [12]. Key implications for the intervention were that the interventionist was a very important aspect of the intervention; there needed to be a feedback mechanism to reward patients for meeting their targets in order to maintain motivation, and that the intervention needed to fit into busy lives, highlighting the need for an app that was produced for use in a future full-scale RCT.

### Conclusions

This behavior change intervention with telehealth components had the expected behavior change mechanisms of action and mechanisms of action associated with effective telehealth interventions for self-management support. The intervention was modified to strengthen mechanisms of action based on these findings, for example, delivery through an app accessed via mobile phones, which is ready for testing in an RCT in 19 UK CF centers.

## Acknowledgments

The authors wish to thank the trial participants, the staff at the 2 CF centers, and particularly the interviewees for their time and support. The authors gratefully acknowledge the input and support of Daniel Beever as the patient and public involvement co-applicant on the wider study, and the administrative and other support from staff in the Clinical Trials Research Unit at the University of Sheffield, United Kingdom, including Hannah Cantrill, Helen Wakefield, and Professor Danny Hind. The views expressed in this paper are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health and Social Care. This work was supported by the National Institute for Health Research Grants for Applied Research Program (Grant No: RP-PG1212-20015).

## Conflicts of Interest

MW has received funding for travel to meetings with PARI Pharma GmbH in Europe. The other authors have no conflicts of interest to declare.

## References

1. Sieben A, van Onzenoort HA, van Laarhoven KJ, Bredie SJ. A multifaceted nurse- and web-based intervention for improving adherence to treatment in patients with cardiovascular disease: rationale and design of the mirror trial. *JMIR Res Protoc* 2016 Sep 13;5(3):e187 [FREE Full text] [doi: [10.2196/resprot.5750](https://doi.org/10.2196/resprot.5750)] [Medline: [27624877](https://pubmed.ncbi.nlm.nih.gov/27624877/)]
2. Qian W, Lam TT, Lam HH, Li C, Cheung YT. Telehealth interventions for improving self-management in patients with hemophilia: scoping review of clinical studies. *J Med Internet Res* 2019 Jul 10;21(7):e12340 [FREE Full text] [doi: [10.2196/12340](https://doi.org/10.2196/12340)] [Medline: [31293241](https://pubmed.ncbi.nlm.nih.gov/31293241/)]
3. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. *JMIR Mhealth Uhealth* 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: [10.2196/13608](https://doi.org/10.2196/13608)] [Medline: [31512580](https://pubmed.ncbi.nlm.nih.gov/31512580/)]
4. Cystic Fibrosis FAQs. Cystic Fibrosis Trust - Fighting for a Life Unlimited. 2019. URL: <https://www.cysticfibrosis.org.uk/what-is-cystic-fibrosis/faqs#How> [accessed 2019-10-08]
5. Sawicki GS, Sellers DE, Robinson WM. High treatment burden in adults with cystic fibrosis: challenges to disease self-management. *J Cyst Fibros* 2009 Mar;8(2):91-96 [FREE Full text] [doi: [10.1016/j.jcf.2008.09.007](https://doi.org/10.1016/j.jcf.2008.09.007)] [Medline: [18952504](https://pubmed.ncbi.nlm.nih.gov/18952504/)]
6. Sawicki GS, Tiddens H. Managing treatment complexity in cystic fibrosis: challenges and opportunities. *Pediatr Pulmonol* 2012 Jun;47(6):523-533. [doi: [10.1002/ppul.22546](https://doi.org/10.1002/ppul.22546)] [Medline: [22467341](https://pubmed.ncbi.nlm.nih.gov/22467341/)]
7. Daniels T, Goodacre L, Sutton C, Pollard K, Conway S, Peckham D. Accurate assessment of adherence: self-report and clinician report vs electronic monitoring of nebulizers. *Chest* 2011 Aug;140(2):425-432. [doi: [10.1378/chest.09-3074](https://doi.org/10.1378/chest.09-3074)] [Medline: [21330381](https://pubmed.ncbi.nlm.nih.gov/21330381/)]
8. Smith S, Rowbotham N, Regan K. Inhaled anti-pseudomonal antibiotics for long-term therapy in cystic fibrosis. *Cochrane Database Syst Rev* 2018 Mar 30;3:CD001021. [doi: [10.1002/14651858.CD001021.pub3](https://doi.org/10.1002/14651858.CD001021.pub3)] [Medline: [29607494](https://pubmed.ncbi.nlm.nih.gov/29607494/)]
9. Yang C, Montgomery M. Dornase alfa for cystic fibrosis. *Cochrane Database Syst Rev* 2018 Sep 6;9:CD001127 [FREE Full text] [doi: [10.1002/14651858.CD001127.pub4](https://doi.org/10.1002/14651858.CD001127.pub4)] [Medline: [30187450](https://pubmed.ncbi.nlm.nih.gov/30187450/)]
10. Briesacher BA, Quittner AL, Saiman L, Sacco P, Fouayzi H, Quittell LM. Adherence with tobramycin inhaled solution and health care utilization. *BMC Pulm Med* 2011 Jan 20;11:5 [FREE Full text] [doi: [10.1186/1471-2466-11-5](https://doi.org/10.1186/1471-2466-11-5)] [Medline: [21251275](https://pubmed.ncbi.nlm.nih.gov/21251275/)]
11. Eakin MN, Bilderback A, Boyle MP, Mogayzel PJ, Riekert KA. Longitudinal association between medication adherence and lung health in people with cystic fibrosis. *J Cyst Fibros* 2011 Jul;10(4):258-264 [FREE Full text] [doi: [10.1016/j.jcf.2011.03.005](https://doi.org/10.1016/j.jcf.2011.03.005)] [Medline: [21458391](https://pubmed.ncbi.nlm.nih.gov/21458391/)]
12. Hind D, Drabble SJ, Arden MA, Mandefield L, Waterhouse S, Maguire C, et al. Supporting medication adherence for adults with cystic fibrosis: a randomised feasibility study. *BMC Pulm Med* 2019 Apr 11;19(1):77 [FREE Full text] [doi: [10.1186/s12890-019-0834-6](https://doi.org/10.1186/s12890-019-0834-6)] [Medline: [30975206](https://pubmed.ncbi.nlm.nih.gov/30975206/)]
13. O'Cathain A, Hoddinott P, Lewin S, Thomas KJ, Young B, Adamson J, et al. Maximising the impact of qualitative research in feasibility studies for randomised controlled trials: guidance for researchers. *Pilot Feasibility Stud* 2015;1:32 [FREE Full text] [doi: [10.1186/s40814-015-0026-y](https://doi.org/10.1186/s40814-015-0026-y)] [Medline: [27965810](https://pubmed.ncbi.nlm.nih.gov/27965810/)]
14. O'Cathain A, Thomas K, Drabble S, Rudolph A, Goode J, Hewison J. Maximising the value of combining qualitative research and randomised controlled trials in health research: the qualitative research in trials (QUART) study--a mixed methods study. *Health Technol Assess* 2014 Jun;18(38):1-197, v [FREE Full text] [doi: [10.3310/hta18380](https://doi.org/10.3310/hta18380)] [Medline: [24914457](https://pubmed.ncbi.nlm.nih.gov/24914457/)]
15. O'Cathain A, Thomas KJ, Drabble SJ, Rudolph A, Hewison J. What can qualitative research do for randomised controlled trials? A systematic mapping review. *BMJ Open* 2013 Jun 20;3(6):- [FREE Full text] [doi: [10.1136/bmjopen-2013-002889](https://doi.org/10.1136/bmjopen-2013-002889)] [Medline: [23794542](https://pubmed.ncbi.nlm.nih.gov/23794542/)]
16. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and Evaluating Complex Interventions. Medical Research Council - UK Research and Innovation. 2008. URL: [www.mrc.ac.uk/complexinterventionsguidance](http://www.mrc.ac.uk/complexinterventionsguidance) [accessed 2018-11-27]



17. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. *Br Med J* 2015 Mar 19;350:h1258. [doi: [10.1136/bmj.h1258](https://doi.org/10.1136/bmj.h1258)] [Medline: [25791983](https://pubmed.ncbi.nlm.nih.gov/25791983/)]
18. Moore G, Audrey S, Barker M, Bond L, Bonell C, Cooper C, et al. Process evaluation in complex public health intervention studies: the need for guidance. *J Epidemiol Community Health* 2014 Mar;68(2):101-102 [FREE Full text] [doi: [10.1136/jech-2013-202869](https://doi.org/10.1136/jech-2013-202869)] [Medline: [24022816](https://pubmed.ncbi.nlm.nih.gov/24022816/)]
19. Arden MA, Drabble S, O'Cathain A, Hutchings M, Wildman M. Adherence to medication in adults with cystic fibrosis: an investigation using objective adherence data and the theoretical domains framework. *Br J Health Psychol* 2019 May;24(2):357-380 [FREE Full text] [doi: [10.1111/bjhp.12357](https://doi.org/10.1111/bjhp.12357)] [Medline: [30825258](https://pubmed.ncbi.nlm.nih.gov/30825258/)]
20. Drabble SJ, O'Cathain A, Arden MA, Hutchings M, Beever D, Wildman M. When is forgetting not forgetting? A discursive analysis of differences in forgetting talk between adults with cystic fibrosis with different levels of adherence to nebulizer treatments. *Qual Health Res* 2019 Dec;29(14):2119-2131 [FREE Full text] [doi: [10.1177/1049732319856580](https://doi.org/10.1177/1049732319856580)] [Medline: [31303116](https://pubmed.ncbi.nlm.nih.gov/31303116/)]
21. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implement Sci* 2012 Apr 24;7:37 [FREE Full text] [doi: [10.1186/1748-5908-7-37](https://doi.org/10.1186/1748-5908-7-37)] [Medline: [22530986](https://pubmed.ncbi.nlm.nih.gov/22530986/)]
22. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011 Apr 23;6:42 [FREE Full text] [doi: [10.1186/1748-5908-6-42](https://doi.org/10.1186/1748-5908-6-42)] [Medline: [21513547](https://pubmed.ncbi.nlm.nih.gov/21513547/)]
23. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. *J Med Internet Res* 2015 Jan 30;17(1):e30. [doi: [10.2196/jmir.4055](https://doi.org/10.2196/jmir.4055)] [Medline: [25639757](https://pubmed.ncbi.nlm.nih.gov/25639757/)]
24. Wong MD, Sarkisian CA, Davis C, Kinsler J, Cunningham WE. The association between life chaos, health care use, and health status among HIV-infected persons. *J Gen Intern Med* 2007 Sep;22(9):1286-1291 [FREE Full text] [doi: [10.1007/s11606-007-0265-6](https://doi.org/10.1007/s11606-007-0265-6)] [Medline: [17597350](https://pubmed.ncbi.nlm.nih.gov/17597350/)]
25. Vassilev I, Rowsell A, Pope C, Kennedy A, O'Cathain A, Salisbury C, et al. Assessing the implementability of telehealth interventions for self-management support: a realist review. *Implement Sci* 2015 Apr 24;10:59. [doi: [10.1186/s13012-015-0238-9](https://doi.org/10.1186/s13012-015-0238-9)] [Medline: [25906822](https://pubmed.ncbi.nlm.nih.gov/25906822/)]
26. English Indices of Deprivation 2019. Government of UK. London; 2015. URL: <http://imd-by-postcode.opendatacommunities.org/> [accessed 2018-08-23]
27. Hoo Z, Campbell M, Curley R, Wildman M. An empirical method to cluster objective nebulizer adherence data among adults with cystic fibrosis. *Patient Prefer Adherence* 2017;11:631-642 [FREE Full text] [doi: [10.2147/PPA.S131497](https://doi.org/10.2147/PPA.S131497)] [Medline: [28392678](https://pubmed.ncbi.nlm.nih.gov/28392678/)]
28. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, editors. *Analysing Qualitative Data*. London, UK: Routledge; 1994:173-194.
29. O'Cathain A, Drabble SJ, Foster A, Horspool K, Edwards L, Thomas C, et al. Being human: a qualitative interview study exploring why a telehealth intervention for management of chronic conditions had a modest effect. *J Med Internet Res* 2016 Jun 30;18(6):e163 [FREE Full text] [doi: [10.2196/jmir.5879](https://doi.org/10.2196/jmir.5879)] [Medline: [27363434](https://pubmed.ncbi.nlm.nih.gov/27363434/)]
30. George M, Rand-Giovannetti D, Eakin MN, Borrelli B, Zettler M, Riekert KA. Perceptions of barriers and facilitators: self-management decisions by older adolescents and adults with CF. *J Cyst Fibros* 2010 Dec;9(6):425-432 [FREE Full text] [doi: [10.1016/j.jcf.2010.08.016](https://doi.org/10.1016/j.jcf.2010.08.016)] [Medline: [20846910](https://pubmed.ncbi.nlm.nih.gov/20846910/)]
31. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *J Med Internet Res* 2011 Mar 10;13(1):e30 [FREE Full text] [doi: [10.2196/jmir.1602](https://doi.org/10.2196/jmir.1602)] [Medline: [21393123](https://pubmed.ncbi.nlm.nih.gov/21393123/)]
32. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. *Am J Prev Med* 2016 Nov;51(5):833-842. [doi: [10.1016/j.amepre.2016.06.015](https://doi.org/10.1016/j.amepre.2016.06.015)] [Medline: [27745683](https://pubmed.ncbi.nlm.nih.gov/27745683/)]
33. Gambling T, Long AF. The realisation of patient-centred care during a 3-year proactive telephone counselling self-care intervention for diabetes. *Patient Educ Couns* 2010 Aug;80(2):219-226. [doi: [10.1016/j.pec.2009.11.007](https://doi.org/10.1016/j.pec.2009.11.007)] [Medline: [20006458](https://pubmed.ncbi.nlm.nih.gov/20006458/)]
34. Markland D, Ryan R, Tobin V, Rollnick S. Motivational interviewing and self-determination theory. *J Soc Clin Psychol* 2005 Sep;24(6):811-831 [FREE Full text] [doi: [10.1521/jscp.2005.24.6.811](https://doi.org/10.1521/jscp.2005.24.6.811)]
35. van Genugten L, Dusseldorp E, Webb TL, van Empelen P. Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? A meta-analysis. *J Med Internet Res* 2016 Jun 7;18(6):e155 [FREE Full text] [doi: [10.2196/jmir.4218](https://doi.org/10.2196/jmir.4218)] [Medline: [27268104](https://pubmed.ncbi.nlm.nih.gov/27268104/)]
36. Sawicki GS, Heller KS, Demars N, Robinson WM. Motivating adherence among adolescents with cystic fibrosis: youth and parent perspectives. *Pediatr Pulmonol* 2015 Mar;50(2):127-136 [FREE Full text] [doi: [10.1002/ppul.23017](https://doi.org/10.1002/ppul.23017)] [Medline: [24616259](https://pubmed.ncbi.nlm.nih.gov/24616259/)]
37. Hoo ZH, Boote J, Wildman MJ, Campbell MJ, Gardner B. Determinants of objective adherence to nebulised medications among adults with cystic fibrosis: an exploratory mixed methods study comparing low and high adherers. *Heal Psychol Behav Med* 2017 Jun 15;5(1):299-316. [doi: [10.1080/21642850.2017.1338958](https://doi.org/10.1080/21642850.2017.1338958)]



38. Dziuban EJ, Saab-Abazeed L, Chaudhry SR, Streetman DS, Nasr SZ. Identifying barriers to treatment adherence and related attitudinal patterns in adolescents with cystic fibrosis. *Pediatr Pulmonol* 2010 May;45(5):450-458. [doi: [10.1002/ppul.21195](https://doi.org/10.1002/ppul.21195)] [Medline: [20425852](https://pubmed.ncbi.nlm.nih.gov/20425852/)]
39. Modi A, Quittner A. Barriers to treatment adherence for children with cystic fibrosis and asthma: what gets in the way? *J Pediatr Psychol* 2006 Sep;31(8):846-858. [doi: [10.1093/jpepsy/jsj096](https://doi.org/10.1093/jpepsy/jsj096)] [Medline: [16401680](https://pubmed.ncbi.nlm.nih.gov/16401680/)]
40. John EK. Overcoming barriers to treatment adherence in adolescents with cystic fibrosis: a systematic review. *Pediatr Neonatal Care* 2016 Nov 23;5(6):- . [doi: [10.15406/jpnc.2016.05.00204](https://doi.org/10.15406/jpnc.2016.05.00204)]
41. Llorente RP, García CB, Martín JJ. Treatment compliance in children and adults with cystic fibrosis. *J Cyst Fibros* 2008 Sep;7(5):359-367 [FREE Full text] [doi: [10.1016/j.jcf.2008.01.003](https://doi.org/10.1016/j.jcf.2008.01.003)] [Medline: [18304896](https://pubmed.ncbi.nlm.nih.gov/18304896/)]
42. Bregnballe V, Schiøtz PO, Boisen KA, Pressler T, Thastum M. Barriers to adherence in adolescents and young adults with cystic fibrosis: a questionnaire study in young patients and their parents. *Patient Prefer Adherence* 2011;5:507-515 [FREE Full text] [doi: [10.2147/PPA.S25308](https://doi.org/10.2147/PPA.S25308)] [Medline: [22114464](https://pubmed.ncbi.nlm.nih.gov/22114464/)]
43. Hogan A, Bonney M, Brien J, Karamy R, Aslani P. Factors affecting nebulised medicine adherence in adult patients with cystic fibrosis: a qualitative study. *Int J Clin Pharm* 2015 Mar;37(1):86-93. [doi: [10.1007/s11096-014-0043-6](https://doi.org/10.1007/s11096-014-0043-6)] [Medline: [25432693](https://pubmed.ncbi.nlm.nih.gov/25432693/)]
44. Lask B. Non-adherence to treatment in cystic fibrosis. *J R Soc Med* 1994;87(Suppl 21):25-27 [FREE Full text] [Medline: [8201583](https://pubmed.ncbi.nlm.nih.gov/8201583/)]
45. Lomas P. Enhancing adherence to inhaled therapies in cystic fibrosis. *Ther Adv Respir Dis* 2014 Apr;8(2):39-47. [doi: [10.1177/1753465814524471](https://doi.org/10.1177/1753465814524471)] [Medline: [24594976](https://pubmed.ncbi.nlm.nih.gov/24594976/)]
46. Oddleifson DA, Sawicki GS. Adherence and recursive perception among young adults with cystic fibrosis. *Anthropol Med* 2017 Apr;24(1):65-80. [doi: [10.1080/13648470.2017.1278865](https://doi.org/10.1080/13648470.2017.1278865)] [Medline: [28292207](https://pubmed.ncbi.nlm.nih.gov/28292207/)]

## Abbreviations

**CF:** cystic fibrosis

**MDT:** multidisciplinary team

**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 24.10.19; peer-reviewed by P Lomas, A Farre, L Hynes; comments to author 23.12.19; revised version received 14.05.20; accepted 25.06.20; published 06.10.20.*

*Please cite as:*

*Drabble SJ, O'Cathain A, Scott AJ, Arden MA, Keating S, Hutchings M, Maguire C, Wildman M*

*Mechanisms of Action of a Web-Based Intervention With Health Professional Support to Increase Adherence to Nebulizer Treatments in Adults With Cystic Fibrosis: Qualitative Interview Study*

*J Med Internet Res* 2020;22(10):e16782

URL: <https://www.jmir.org/2020/10/e16782>

doi: [10.2196/16782](https://doi.org/10.2196/16782)

PMID: [32697197](https://pubmed.ncbi.nlm.nih.gov/32697197/)

©Sarah J Drabble, Alicia O'Cathain, Alexander J Scott, Madelynne A Arden, Samuel Keating, Marlene Hutchings, Chin Maguire, Martin Wildman. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Health-Related Quality of Life Improvements in Systemic Lupus Erythematosus Derived from a Digital Therapeutic Plus Tele-Health Coaching Intervention: Randomized Controlled Pilot Trial

Faiz Khan<sup>1\*</sup>, MD; Nora Granville<sup>2\*</sup>, MA; Raja Malkani<sup>3\*</sup>, BS; Yash Chathampally<sup>4\*</sup>, MD, MSc

<sup>1</sup>EVP, CityMD, Dix Hills, NY, United States

<sup>2</sup>Independent Researcher, New York, NY, United States

<sup>3</sup>Independent Researcher, Austin, TX, United States

<sup>4</sup>Harris Health, Houston, TX, United States

\* all authors contributed equally

## Corresponding Author:

Faiz Khan, MD

EVP, CityMD

101 Long Drive Court

Dix Hills, NY, 11746

United States

Phone: 1 516 426 9086

Email: [fkhan@citymd.net](mailto:fkhan@citymd.net)

## Abstract

**Background:** Systemic lupus erythematosus (SLE), a systemic autoimmune disease with no known cure, remains poorly understood and patients suffer from many gaps in care. Recent work has suggested that dietary and other lifestyle factors play an important role in triggering and propagating SLE in some susceptible individuals. However, the magnitude of influence of these triggers, how to identify pertinent triggers in individual patients, and whether removing these triggers confers clinical benefit is unknown.

**Objective:** To demonstrate that a digital therapeutic intervention, utilizing a mobile app that allows self-tracking of dietary, environmental, and lifestyle triggers, paired with telehealth coaching, added to usual care, improves quality of life in patients with SLE compared with usual care alone.

**Methods:** In this randomized controlled pilot study, adults with SLE were assigned to a 16-week digital therapeutic intervention plus usual care or usual care alone. Primary outcome measures were changes from baseline to 16 weeks on 3 validated health-related quality of life (HRQoL) tools: Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F), Brief Pain Inventory-Short Form (BPI-SF), and Lupus Quality of Life (LupusQoL).

**Results:** A total of 50 patients were randomized (23 control, 27 intervention). In per-protocol analysis, the intervention group achieved significantly greater improvement than the control group in 9 of 11 domains: FACIT-F (34% absolute improvement for the intervention group vs -1% for the control group,  $P<.001$ ), BPI-SF-Pain Interference (25% vs 0%,  $P=.02$ ), LupusQoL-Planning (17% vs 0%,  $P=.004$ ), LupusQoL-Pain (13% vs 0%,  $P=.004$ ), LupusQoL-Emotional Health (21% vs 4%,  $P=.02$ ), and LupusQoL-Fatigue (38% vs 13%,  $P<.001$ ) were significant when controlling for multiple comparisons; BPI-SF-Pain Severity (13% vs -6%,  $P=.049$ ), LupusQoL-Physical Health (17% vs 3%,  $P=.049$ ), and LupusQoL-Burden to Others (33% vs 4%,  $P=.04$ ) were significant at an unadjusted 5% significance level.

**Conclusions:** A digital therapeutic intervention that pairs self-tracking with telehealth coaching to identify and remove dietary, environmental, and lifestyle symptom triggers resulted in statistically significant, clinically meaningful improvements in HRQoL when added to usual care in patients with SLE.

**Trial Registration:** ClinicalTrials.gov NCT03426384; <https://clinicaltrials.gov/ct2/show/NCT03426384>

(*J Med Internet Res* 2020;22(10):e23868) doi:[10.2196/23868](https://doi.org/10.2196/23868)

**KEYWORDS**

systemic lupus erythematosus; digital health; digital therapeutic; autoimmunity; food as medicine; dietary intervention; health-related quality of life; lifestyle medicine; mobile health; environmental influences on autoimmunity

## Introduction

### Background

Systemic lupus erythematosus (SLE) is a multisystem, complex autoimmune disease of unclear etiology affecting at least 1.5 million Americans and 5 million worldwide [1]. The hallmark of the disease is uncontrolled inflammation in otherwise healthy tissue which can lead to organ damage and sometimes even organ failure. SLE can affect any body system—the kidneys, skin, joints, heart, lungs, gastrointestinal system, and nervous system may all become involved, and the pattern of involvement differs from patient to patient. The most common symptoms are fever, rash, profound fatigue, and joint pain and swelling. Disease activity is prone to exacerbations (called flares) alternating with periods of remission in cycles that are often unpredictable and therefore have an even more detrimental effect on quality of life.

There is no cure for SLE and universally effective treatment is not available. Current management relies on immune modulating drugs, but their side effects often increase discomfort and their use carries the risk of severe adverse events [2]. While 5-year survival rates have increased dramatically from 50% to 90% [3,4], patients with SLE still have significantly higher age-standardized mortality rates [5] and lower health-related quality of life (HRQoL) than the general population [6]. Underemployment and work disability, associated primarily with fatigue and pain, are common [7,8]. Young women are disproportionately affected, especially those of non-Caucasian ethnicity and low socioeconomic status. It has recently been reported that SLE is the leading cause of death among chronic inflammatory diseases in women aged 15-24, with death rates exceeding those of HIV and diabetes [9].

Although a clear understanding of the pathophysiology of SLE remains elusive, the recognition that an individual's DNA blueprint alone does not wholly account for disease occurrence has fueled new areas of research into the environmental and lifestyle determinants of SLE. Important to this line of inquiry is (1) the growing recognition that epigenetic alterations, such as DNA methylation, noncoding RNAs, and histone modifications, are involved in the development of autoimmune diseases [10] and (2) emerging evidence from human and animal studies that these epigenetic processes are influenced by dietary, environmental, and lifestyle factors [11-23]. Modifying these factors presents an attractive, low-risk treatment option for SLE. However, there is much work to be done to better define these potential triggers and determine if eliminating them confers clinical benefit. Complicating these efforts is the fact that SLE is an extremely heterogeneous disease. Widely variable initial presentation, disease course, organ involvement, and response to treatment complicate diagnosis, management, and clinical research efforts. An international team of experts have identified SLE heterogeneity as “the primary barrier hindering advancement” [24]. Given this heterogeneity, it is reasonable

to hypothesize that numerous dietary, environmental, and lifestyle SLE triggers exist and differ from patient to patient. Therefore, tracking and analyzing possible trigger-symptom associations require a reliable, accurate, easy-to-use method for gathering and processing a considerable amount of data. Digital therapeutics can accomplish this and may offer unique solutions to the obstacles faced in attempts to address the dietary, environmental, and lifestyle triggers of SLE.

Several digital therapeutics have already received FDA clearance, and many more are in development, to address a range of medical conditions, including prediabetes and diabetes, substance use and opioid use disorders, Alzheimer disease, obesity, hypertension, chronic back pain, attention-deficit/hyperactivity disorder, concussion, and multiple sclerosis [25-27]. Patients with SLE can use several available apps to help track symptoms and disease activity [28-30] and manage medications [31]. However, none of the existing apps have successfully addressed the relationship between dietary, environmental, and lifestyle factors and symptom severity in autoimmunity.

A digital therapeutic platform has been developed which combines self-tracking technology, analytics, and tele-health coaching to identify and remove possible dietary, environmental, or lifestyle triggers, with the goal to provide clinically meaningful improvements in symptoms and HRQoL in those with autoimmune disease. The platform is intended as an adjunct to standard of care.

The platform was developed over several years with extensive feedback from stakeholders in the autoimmune disease community. This has included discussions with patients, family members, physicians, insurance providers, foundations, patient advocacy groups, pharmaceutical companies, and even potential service providers with experience in the sector, such as contract research organizations. The goal has been to commercialize a product that serves an unmet clinical need, but also that fits into the clinical workflow, would be widely adopted, and has a pathway to reimbursement. As a digital therapeutic, the product is also able to track patient usage and engagement during the course of the program, and notifications can be sent following the program to track longer-term outcomes.

Usability and patient preferences have been carefully considered to ensure that a broad range of individuals are comfortable engaging with the smartphone interface, participating in coaching sessions, and complying with suggested interventions throughout the program. Similarly, the web portal and the health coaching protocol itself were iteratively refined through consultation with health coaches and health care providers.

This novel approach is unique in that it implicitly takes disease heterogeneity into account, leverages the growing understanding of the role environment plays in initiating and propagating autoimmune disease, and personalizes each patient's recommendations based on software data analytics.

## Objective

The objective of this study was to determine whether the addition to usual care of this digital therapeutic program—intended to identify and intervene on dietary and other lifestyle factors found by data analytics to be associated with symptom frequency and severity—improved HRQoL in patients with SLE more than usual care alone.

## Methods

### Study Population

The study enrolled adults ( $\geq 18$  years) across the United States from December 2017 to May 2018. Participants were recruited through the following online forums: Lupus Friends and Family, Flare Fighter, and Purple Wings Facebook groups; and Clara Health and Autoimmune Registry (online resources for patients interested in participating in clinical trials). Interested individuals completed a prequalification survey online, and only those individuals who passed the prequalification survey (ie, those who were not disqualified) were asked to submit medical records which were reviewed by the study principal investigator (FK) to verify a diagnosis of SLE and confirm all inclusion/exclusion criteria. Eligible individuals underwent a phone consent session and electronically signed an informed consent document if they chose to participate. Participants were assigned to either the intervention or control arm via randomized blocks of 3 to 8 individuals using a cryptographic random seed, targeting 1:1 allocation between groups. After randomization and electronic collection of baseline data, randomization groups were made known to participants.

Inclusion criteria included owning a smartphone, a threshold score for at least one of seven pain and fatigue questions, and taking a stable dose of one of more of the following drugs for 3 or more months prior to study enrollment: immunosuppressive or immunomodulating therapy (biologic or nonbiologic), immunoglobulin therapy, or 20 mg or more of prednisone (or equivalent corticosteroid). Exclusion criteria included pre-existing or incident diagnosis of cancer, pregnancy, or intention to conceive during the study period, and criteria intended to avoid confounding interpretation of changes in outcome: current or planned participation in another interventional or observational study, known plans to alter inclusion criteria medications prior to onset of study or during study, and use of pulse steroids for more than 30 days combined or during the last 4 weeks of the intervention period.

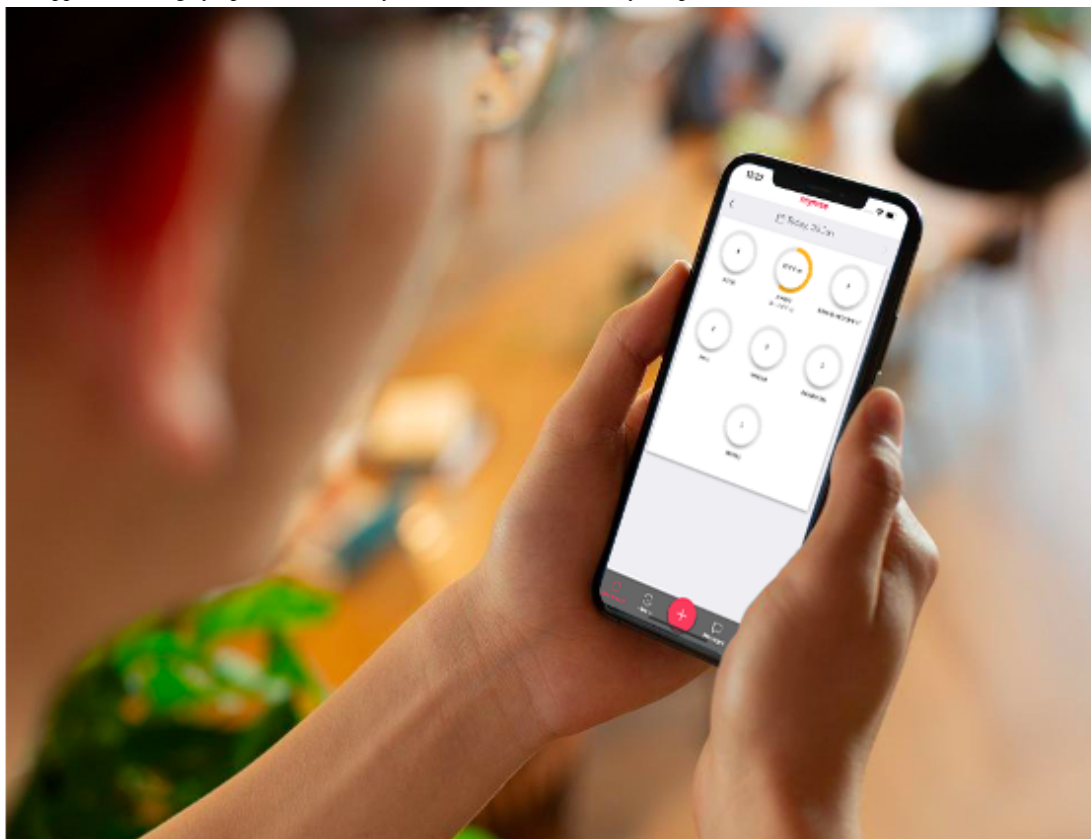
The occurrence of adverse events was assessed during the coaching sessions and by a call to every participant by the principal investigator at the conclusion of the study. Coaches were instructed to convey possible adverse events to a study team physician (FK) who would report them to the Institutional Review Board.

The study was approved by Western Institutional Review Board (CSI: NCT03426384).

### Intervention

The digital therapeutic technology has 3 key components: a smartphone iOS or Android app for the patient to track lifestyle activities (eg, diet, sleep habits, physical activity, bowel movements) and symptoms (Figure 1); software that analyzes and organizes data; and a web portal that presents all patient data to the health coach.



**Figure 1.** Mobile app for entering symptoms and dietary, environmental, and lifestyle inputs.

During weekly telehealth coaching sessions, the health coach viewed the data in the web portal and recommended interventions intended to confirm or reject suspected dietary, environmental, and lifestyle triggers. Potential trigger–symptom correlations were presented by the software and the prioritization of triggers and the decision on which intervention to suggest were performed by the health coach. Throughout the program, tracked symptoms were not a comprehensive representation of all participant's complaints but rather were chosen by the coach based on their propensity to fluctuate, thereby providing indicators which were more likely to be responsive to change. The health coach also provided nutrition and lifestyle education as needed (eg, addition of protein at breakfast, instruction on nutrient-dense foods, recommendations for stress management techniques). Over-the-counter supplementation was recommended by the health coach on a case-by-case basis as appropriate (eg, vitamin D3 if patient records showed deficiency that had not already been addressed—see [Multimedia Appendix 1](#) for full details). A single, certified health coach performed all coaching sessions for every participant. Participants randomized to the intervention group received an email with instructions on how to download the app, create a profile, and use the app to track dietary input. Throughout the study period, participants could receive technology and other support by messaging (through the app) or emailing the study team at any time. After 3–5 days of taking pictures of all the food and beverages consumed daily, participants completed an introductory telephone session to identify their symptoms and goals of the program, review initial tracking data, and receive further training on tracking of other environmental and lifestyle inputs.

Following this initial call, weekly 20–30-minute telehealth coaching sessions were scheduled for the ensuing 15 weeks. Each week, based on associations presented by the software between dietary or other tracked exposures (triggers) and symptoms, the coach suggested behavioral interventions to ameliorate symptoms (eg, eliminate dairy if a patient's joint pain appeared to flare in relation to dairy intake over the past week). The results of these iterative, weekly interventions were reviewed in subsequent sessions. Compliance with interventions was assessed by analysis of digital tracking and weekly coaching discussions. Successful interventions were maintained, whereas those which did not impact symptoms were either rejected or subjected to longer trials.

Control group participants continued usual care as recommended by their treating physician(s); were not introduced to the intervention app (or any other sham app); and received no training, coaching, or other study interventions.

Prior to entry into the study, all participants had a call with a study staff member to review a summary of the trial, the intervention procedures and schedule, potential risks and benefits, alternative treatments, and provide informed consent. Control participants completed the same battery of assessments at the same intervals as the intervention group participants. At the end of 16 weeks, control participants met with a study team member by phone during which time final assessment surveys were administered, adverse events over the prior 16 weeks were ascertained, and the opportunity for cross-over to receive the digital therapeutic intervention was offered. All surveys were completed via an HIPAA (Health Insurance Portability and Accountability Act)-compliant version of SurveyGizmo.



## Outcomes

The primary outcomes were changes between baseline and week 16 in 11 domains reflecting various aspects of HRQoL, as assessed by 3 validated patient-reported outcome measures (PROMs): (1) Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F), consisting of 13 questions aggregated into 1 domain measuring fatigue; (2) Brief Pain Inventory-Short Form (BPI-SF), consisting of 15 questions classified into 2 domains (pain severity and pain interference), and; (3) Lupus Quality of Life (LupusQoL), consisting of 34 questions classified into 8 domains (fatigue, physical health, planning, burden to others, emotional health, pain, intimate relationships, and body image). All 3 outcome measures have been previously described [32-34] and validated for use in SLE [29,35,36]. The participants were asked to complete these PROMs on a secure website prior to the start of the intervention and at weeks 4, 8, 12, and 16.

Secondary outcomes (derived from analysis of tracking data and coach dashboard information) were tracking adherence (the number of days a participant logged at least one observation into the mobile app in a 24-hour period); session adherence (the number of weekly coaching calls a participant participated in over the 16 weeks); and types and prevalence of (1) the most commonly tracked symptoms, (2) suspected triggers, and (3) interventions. These data were generated from the participant's tracking data and coaching notes.

## Adherence

Tracking adherence was calculated as the number of days (24-hour period) at least one observation (eg, symptom, food, other lifestyle input) was entered into the app divided by the number of days in the 16-week program ( $n=112$ ) to arrive at the percentage of days with tracked data (adherence of 100% indicates that the participant used the app to track more than once/day each day of the program). Coaching session adherence was calculated as the number of coaching sessions completed by the participant divided by 16 and converted to a percentage (adherence of 100% indicates that the participant completed 1 or more session/week each week of the program). Median and 25th and 75th percentile values for tracking and session adherence were then calculated for the whole group.

## Statistical Analysis

Prior to the study, the sample size was computed based on the Mann-Whitney  $U$  test to provide approximately 80% power to detect an effect size proportional to a mean difference in improvement of about 10% with a standard deviation of 10% without correcting for multiple comparisons. This effect size was chosen based on early user experience with the program as well as consideration of previously established minimally important differences for the outcome measures [35-37]. It was

determined that a sample size of 50 was sufficient to allow for attrition and still produce the needed power with the remaining participants expected to complete the study. To balance minimizing type I and type II errors, results were highlighted that were significant at an unadjusted significance level of 5% and also, due to the high level of correlation in outcomes, at a level adjusted using the Benjamini-Hochberg method to control the false detection rate at 5% for multiple comparisons.

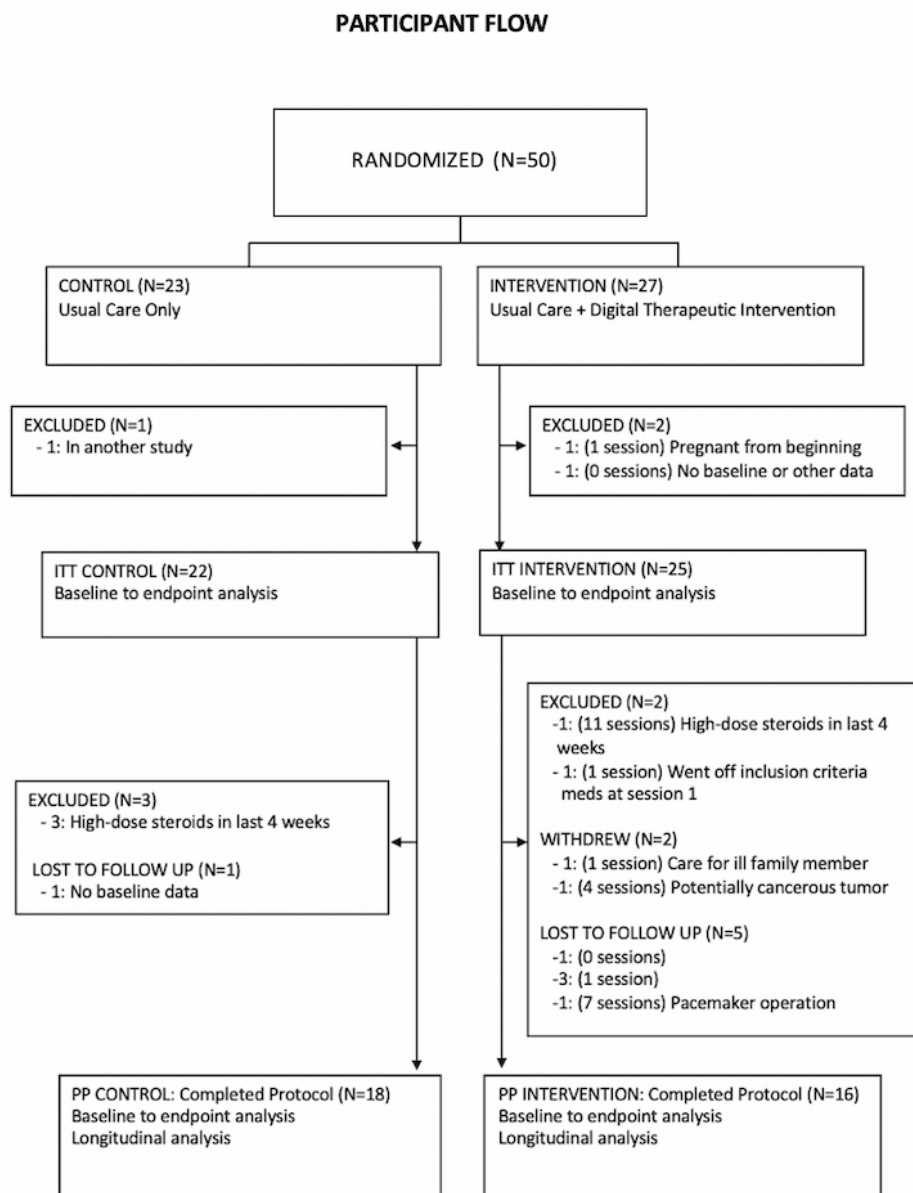
Nonparametric tests were chosen based on minimal distributional assumptions given the small sample size: Wilcoxon signed-rank test for the change within the intervention and control groups between the baseline and 16-week/end-of-program domain scores; and Mann-Whitney  $U$  test for changes in score between the intervention and control groups. Medians, 25th, and 75th percentile values are displayed as measures of central tendency and spread. All statistical analyses were performed using IBM SPSS Statistics Subscription (Build 1.0.0.1072).

Intention-to-treat (ITT) analysis included participants who met inclusion criteria at the start of the intervention period, even if they did not complete the study. Per-protocol (PP) analysis was limited to participants who completed 10 or more sessions within the 16-week study period (based on prior exploratory testing), submitted end-of-study data, and experienced no exclusions. Missing follow-up scores from participants who dropped out of the intervention group were populated with the worst observed scores for that time point, thus biasing toward the null hypothesis.

## Results

### Study Population

In total, 50 patients were enrolled, with 47 included in ITT analysis and 34 in PP analysis (Figure 2). Table 1 shows the study population baseline demographics. The control and intervention groups were similar across most categories and any differences were not expected to impact results. For the full cohort, the median age was 43; 96% (44/46) were female; 59% (27/46) of participants were Caucasian, 17% (8/46) Black or African American, and 24% (11/46) Hispanic. Of the 25 ITT intervention participants, 6 (24%) were lost to follow up after completing 0 or 1 coaching sessions (1 discontinued inclusion medication after 1 session; 1 voluntarily withdrew after 1 session to care for a sick family member; 4 were lost to follow up after completing 1 [ $n=3$  participants] or no [ $n=1$ ] coaching sessions). Of the remaining 19, 16 completed at least ten coaching sessions over 16 weeks (for a completion rate of 84%) and were included in PP analysis. Medications at study entry and baseline scores on the 3 PROMs are shown in Table 2.

**Figure 2.** Participant flow. ITT: intention to treat; PP: per protocol.

**Table 1.** Study population demographics.<sup>a</sup>

Demographics	ITT <sup>b</sup>		PP <sup>c</sup>		EPP <sup>d</sup>	
	Control	Intervention	Control	Intervention	Control	Intervention
Participants who provided baseline data, N	21	25	18	16	3	9
Age, median (25th percentile, 75th percentile)	42 (36, 50)	44 (33, 51)	42 (35, 50)	45 (35, 54)	43 (37, 59)	36 (31, 47)
<b>Ethnic background</b>						
Black or African American	2 (10)	6 (24)	2 (11)	3 (19)	0 (0)	3 (33)
Caucasian or White	12 (57)	15 (60)	9 (50)	10 (63)	3 (100)	5 (56)
Hispanic or Latino	7 (33)	4 (16)	7 (39)	3 (19)	0 (0)	1 (11)
<b>Primary language</b>						
English	20 (95)	25 (100)	17 (94)	16 (100)	3 (100)	9 (100)
Spanish	1 (5)	0 (0)	1 (6)	0 (0)	0 (0)	0 (0)
<b>Gender</b>						
Female	20 (95)	24 (96)	17 (94)	15 (94)	3 (100)	9 (100)
Male	1 (5)	1 (4)	1 (6)	1 (6)	0 (0)	0 (0)
<b>Education level</b>						
Some high school	0 (0)	1 (4)	0 (0)	0 (0)	0 (0)	1 (11)
High school	1 (5)	0 (0)	1 (6)	0 (0)	0 (0)	0 (0)
Some college/trade/technical training	4 (19)	7 (28)	3 (17)	3 (19)	1 (33)	4 (44)
Associate degree	5 (24)	6 (24)	4 (22)	5 (31)	1 (33)	1 (11)
Bachelor's degree	9 (43)	5 (20)	8 (44)	3 (19)	1 (33)	2 (22)
Master's/Professional degree	2 (10)	6 (24)	2 (11)	5 (31)	0 (0)	1 (11)
<b>Employment</b>						
Full-time paid	7 (33)	7 (28)	7 (39)	5 (31)	0 (0)	2 (22)
Part-time paid	3 (14)	4 (16)	3 (17)	3 (19)	0 (0)	1 (11)
Self-employed	1 (5)	1 (4)	0 (0)	1 (6)	1 (33)	0 (0)
Homemaker	1 (5)	0 (0)	1 (6)	0 (0)	0 (0)	0 (0)
Out of work, not currently looking	0 (0)	1 (4)	0 (0)	1 (6)	0 (0)	0 (0)
Unable to work—on disability	5 (24)	11 (44)	3 (17)	6 (38)	2 (67)	5 (56)
Unable to work—other	4 (19)	1 (4)	4 (22)	0 (0)	0 (0)	1 (11)
<b>Income Level</b>						
US \$0-US \$25,999	6 (29)	7 (28)	6 (33)	3 (19)	0 (0)	4 (44)
US \$26,000-US \$51,999	5 (24)	7 (28)	3 (17)	5 (31)	2 (67)	2 (22)
US \$52,000-US \$74,999	3 (14)	6 (24)	2 (11)	4 (25)	1 (33)	2 (22)
More than US \$75,000	7 (33)	5 (20)	7 (39)	4 (25)	0 (0)	1 (11)
<b>Relationship status</b>						
Life partner (married/other)	13 (62)	15 (60)	11 (61)	11 (69)	2 (67)	4 (44)
Single/separate/divorced/widowed	8 (38)	10 (40)	7 (39)	5 (31)	1 (33)	5 (56)

<sup>a</sup>Values are numbers (percentages) unless stated otherwise.<sup>b</sup>ITT: intention to treat.<sup>c</sup>PP: per protocol.<sup>d</sup>EPP: ITT participants who were excluded from PP.

**Table 2.** Study population inclusion medications and baseline patient-reported outcome measure scores.<sup>a</sup>

Medications and outcome measure scores	ITT <sup>b</sup>		PP <sup>c</sup>		EPP <sup>d</sup>	
	Control	Intervention	Control	Intervention	Control	Intervention
Participants who provided baseline data, N	21	25	18	16	3	9
<b>Inclusion medications<sup>e</sup></b>						
Azathioprine	2 (10)	5 (20)	1 (6)	1 (6)	1 (33)	4 (44)
Belimumab	4 (19)	7 (28)	3 (17)	4 (25)	1 (33)	3 (33)
Hydroxychloroquine	17 (81)	19 (76)	13 (72)	13 (81)	3 (100)	6 (67)
Immunoglobulin infusions	1 (5)	0 (0)	1 (6)	0 (0)	0 (0)	0 (0)
Leflunomide	0 (0)	1 (4)	0 (0)	1 (6)	0 (0)	0 (0)
Methotrexate	4 (19)	3 (12)	4 (22)	2 (13)	0 (0)	1 (11)
Mycophenolate mofetil	5 (24)	2 (8)	5 (28)	2 (13)	0 (0)	0 (0)
<b>HRQoL<sup>f</sup>, median (25th percentile, 75th percentile)</b>						
FACIT <sup>g</sup> -Fatigue	20 (14, 26)	16 (10, 23)	20 (14, 27)	20 (10, 26)	7 (4, 18)	13 (8, 19)
BPI-SF <sup>h</sup> -Pain Severity	4 (3, 5.7)	5 (3, 6)	3 (3, 5)	4 (3, 6)	5 (5, 7)	6 (3, 7)
BPI-SF-Pain Interference	5 (5, 6)	6 (4, 7)	5 (4, 6)	6 (4, 7)	7 (3, 7)	6 (6, 8)
LupusQoL <sup>i</sup> -Physical Health	50 (25, 59)	46 (28, 56)	51 (28, 65)	51 (34, 60)	25 (15, 28)	25 (12, 50)
LupusQoL-Pain	41 (16, 66)	41 (25, 66)	58 (25, 75)	50 (33, 66)	16 (8, 33)	25 (8, 50)
LupusQoL-Planning	50 (25, 75)	41 (8, 66)	62 (25, 75)	62 (25, 75)	41 (0, 41)	25 (8, 25)
LupusQoL-Intimate Relationships	50 (25, 87)	56 (25, 75)	50 (25, 75)	75 (25, 81)	68 (50, 87)	31 (25, 50)
LupusQoL-Burden to Others	25 (0, 41)	16 (0, 41)	25 (0, 58)	25 (4, 50)	0 (0, 33)	16 (0, 16)
LupusQoL-Emotional Health	54 (37, 70)	54 (29, 66)	56 (37, 70)	60 (35, 79)	50 (20, 70)	20 (16, 54)
LupusQoL-Body Image	37 (20, 56)	50 (25, 69)	31 (20, 65)	65 (18, 75)	40 (35, 50)	31 (25, 45)
LupusQoL-Fatigue	25 (6, 37)	25 (18, 31)	25 (12, 43)	28 (25, 50)	6 (0, 31)	18 (0, 18)

<sup>a</sup>Values are n (%) unless stated otherwise.<sup>b</sup>ITT: intention to treat.<sup>c</sup>PP: per protocol.<sup>d</sup>EPP: ITT participants who were excluded from PP.<sup>e</sup>Totals do not equal 100% as many patients were on multiple medications.<sup>f</sup>HRQoL: health-related quality of life.<sup>g</sup>FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F); 52-point scale with 0 (worst).<sup>h</sup>BPI-SF: Brief Pain Inventory-Short Form; 10-point scale with 0 (best).<sup>i</sup>LupusQoL: Lupus Quality of Life; 100-point scale with 0 (worst).

## Adherence

Table 3 shows tracking and coaching session adherence for the ITT and PP groups. In each group, tracking adherence exceeded 90% and coaching session adherence exceeded 80%. In the ITT group, 16/25 (64%) and 14/25 (56%) participants reached greater than 70% tracking and session adherence, respectively. In the

PP group, 16/16 (100%) and 13/16 (81%) participants reached greater than 70% tracking and session adherence, respectively. The percentage of participants achieving 70% or greater tracking and session adherence is reported based on early experience with the platform indicating that this level of engagement correlates with better outcomes.

**Table 3.** Adherence results.

Adherence	Intention to treat		Per protocol	
	Tracking %	Coaching sessions %	Tracking %	Coaching sessions %
Median (25th, 75th percentile)	91.1 (50.9, 97.3)	81.3 (25.0, 81.3)	96.9 (94.4, 99.1)	81.3 (81.3, 93.8)
Over 70% adherence, n/N (%)	16/25 (64)	14/25 (56)	16/16 (100)	13/16 (81)

### Intention-to-Treat Analysis

Within the intervention group, significant improvement over baseline was noted for FACIT-F (median of 26.0 at the end of study vs 16.0 baseline,  $P=.04$ ), LupusQoL-Burden to Others (25.0 vs 16.7,  $P=.02$ ), and LupusQoL-Fatigue (62.5 vs 25.0,  $P=.007$ ). Within the control group, LupusQoL-Burden to Others (41.7 vs 20.8,  $P=.04$ ), LupusQoL-Body Image (45.0 vs 35.0,  $P=.047$ ), and LupusQoL-Fatigue (31.3 vs 25.0,  $P=.03$ ) saw

improvement over baseline at 16 weeks. Comparing the 2 groups, although the intervention group improved more than the control group in 6 of 11 domains (FACIT-F, BPI-SF-Pain interference, LupusQoL-Pain, LupusQoL-Emotional Health, LupusQoL-Body Image, and LupusQoL-Fatigue), none of these comparisons reached statistical significance (Table 4). No significant improvements were uncovered when the Benjamini–Hochberg adjustment was applied to the significance level to account for multiple comparisons.



**Table 4.** Intention-to-treat analysis of change in FACIT, BPI-SF, and LupusQoL domain scores from baseline to end of program.<sup>a</sup>

Domain	Within group				<i>P</i> -value	Between group	
	Count	Baseline	End of program (EOP)	Change in score (EOP: Baseline)		Difference <sup>b</sup>	<i>P</i> -value
FACIT <sup>c</sup> -Fatigue (range 0-52, higher is better)						4.5	.17
Intervention	25	16.0 (9.5, 23.5)	26.0 (4.0, 44.0)	4.0 (−3.5, 21.0)	.04 <sup>f</sup>		
Control	22	19.5 (7.0, 26.3)	21.0 (10.5, 28.3)	−0.5 (−5.0, 7.3)	.75		
BPI-SF <sup>d</sup> -Pain Severity (range 0-10, lower is better)						−0.6	.73
Intervention	25	5.3 (3.0, 6.8)	5.3 (2.1, 8.3)	0.0 (−2.8, 2.3)	.76		
Control	22	4.5 (3.0, 6.6)	4.4 (2.6, 7.1)	0.6 (−1.3, 1.0)	.68		
BPI-SF-Pain Interference (range 0-10, lower is better)						−0.7	.31
Intervention	25	6.4 (4.4, 7.9)	4.7 (1.6, 9.3)	−0.6 (−3.6, 0.6)	.16		
Control	22	5.6 (4.4, 6.7)	5.1 (1.6, 7.5)	0.1 (−1.2, 1.7)	.97		
LupusQoL <sup>e</sup> -Physical Health (range 0-100, higher is better)						−3.1	.88
Intervention	25	46.9 (26.6, 56.3)	31.3 (4.7, 78.1)	0.0 (−18.8, 29.7)	.64		
Control	22	46.9 (23.4, 60.9)	40.6 (21.9, 71.1)	3.1 (−10.2, 10.2)	.82		
LupusQoL-Pain (range 0-100, higher is better)						12.5	.21
Intervention	25	41.7 (20.8, 66.7)	41.7 (0.0, 83.3)	8.3 (−20.8, 33.3)	.63		
Control	22	37.5 (14.6, 68.8)	33.3 (14.6, 66.7)	−4.2 (−16.7, 2.1)	.28		
LupusQoL-Planning (range 0-100, higher is better)						0.0	.24
Intervention	25	41.7 (8.3, 70.8)	50.0 (0.0, 91.7)	0.0 (−12.5, 25.0)	.38		
Control	22	45.8 (22.9, 75.0)	41.7 (18.8, 77.1)	0.0 (−27.1, 8.3)	.36		
LupusQoL-Burden to Others (range 0-100, higher is better)						−8.3	.92
Intervention	25	16.7 (0.0, 41.7)	25.0 (0.0, 83.3)	0.0 (0.0, 50.0)	.02 <sup>f</sup>		
Control	22	20.8 (0.0, 45.8)	41.7 (0.0, 77.1)	8.3 (0.0, 16.7)	.04 <sup>f</sup>		
LupusQoL-Intimate Relationships (range 0-100, higher is better)						−12.5	.46
Intervention	19	50.0 (25.0, 75.0)	62.5 (0.0, 87.5)	−12.5 (−25.0, 25.0)	.79		
Control	18	56.3 (25.0, 87.5)	62.5 (25.0, 100.0)	0.0 (−3.1, 12.5)	.47		
LupusQoL-Emotional Health (range 0-100, higher is better)						6.2	.37
Intervention	25	54.2 (25.0, 68.8)	75.0 (4.2, 93.8)	8.3 (−10.4, 29.2)	.30		
Control	22	52.1 (29.2, 70.8)	56.3 (37.5, 70.8)	2.1 (−12.5, 12.5)	.64		
LupusQoL-Body Image (range 0-100, higher is better)						8.1	.505
Intervention	22	41.3 (23.8, 68.8)	51.9 (0.0, 90.0)	13.1 (−30.3, 21.3)	.76		
Control	19	35.0 (20.0, 56.3)	45.0 (30.0, 70.0)	5.0 (0.0, 25.0)	.047 <sup>f</sup>		
LupusQoL-Fatigue (range 0-100, higher is better)						9.4	.22

Domain	Within group			Between group		
	Count	Baseline	End of program (EOP)	Change in score (EOP: Baseline)	<i>P</i> -value	Difference <sup>b</sup> <i>P</i> -value
Intervention	25	25.0 (15.6, 34.4)	62.5 (0.0, 81.3)	18.8 (−9.4, 43.8)	.007 <sup>f</sup>	
Control	22	25.0 (4.7, 39.1)	31.3 (15.6, 53.1)	9.4 (−6.3, 20.3)	.03 <sup>f</sup>	

<sup>a</sup>Within-group values are median (25th percentile, 75th percentile). Nonparametric tests were chosen in order to require minimal distributional assumptions given the small sample size, and medians, 25th, and 75th percentile values are displayed as measures of central tendency and spread in order to be consistent with a nonparametric analysis. The within-group *P*-value is from the Wilcoxon signed-rank test and the between-group *P*-value is from the Mann–Whitney *U* test; *P*-values themselves are unadjusted but the threshold for statistical significance is set using the Benjamini–Hochberg adjustment; LupusQoL-Intimate Relationships and LupusQoL-Body Image allow the possibility of N/A responses.

<sup>b</sup>Difference in median change (intervention – control).

<sup>c</sup>FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F); 52-point scale with 0 (worst).

<sup>d</sup>BPI-SF: Brief Pain Inventory-Short Form; 10-point scale with 0 (best).

<sup>e</sup>LupusQoL: Lupus Quality of Life; 100-point scale with 0 (worst).

<sup>f</sup>Statistically significant at an unadjusted 2-sided significance level of 5%.

## Per-Protocol Analysis

Within-group analysis of the PP intervention population revealed improvement over baseline at 16 weeks in all domains except LupusQoL-Intimate Relationships at unadjusted significance levels ( $P=.05$ ; Table 5). Adjusting for multiple comparisons, statistically significant improvement was achieved by the PP intervention group in 8 domains: FACIT-F (43.5 at the end of study vs 20.5 baseline;  $P=.001$ ), LupusQoL-Fatigue (81.3 vs 28.1;  $P<.001$ ), LupusQoL-Physical Health (71.9 vs 51.6;  $P=.02$ ), LupusQoL-Planning (83.3 vs 62.5;  $P=.008$ ), LupusQoL-Burden to Others (79.2 vs 25.0;  $P=.003$ ), LupusQoL-Emotional Health (83.3 vs 60.4;  $P=.01$ ), LupusQoL-Body Image (87.5 vs 56.3,  $P=.01$ ), and BPI-SF-Pain Interference (2.0 vs 6.3;  $P=.003$ ). The usual care PP population had significant improvement over baseline for LupusQoL-Fatigue only (34.4 vs 25.0;  $P=.028$ ).

Between-group comparisons demonstrated greater improvement in the intervention group than in the control group for every

domain. Adjusting for multiple tests, significant differences in favor of the intervention group were reached in 6 domains: FACIT-F (difference in median changes of 18.0,  $P<.001$ ), BPI-SF-Pain interference (−2.5,  $P=.02$ ), LupusQoL-Pain (12.5,  $P=.004$ ), LupusQoL-Planning (16.7,  $P=.004$ ), LupusQoL-Emotional Health (16.7,  $P=.02$ ), and LupusQoL-Fatigue (25.0,  $P<.001$ ). Three additional domains reached significance at an unadjusted level of 5%: BPI-SF-Pain Severity (−1.9,  $P=.049$ ), LupusQoL-Physical Health (14.1,  $P=.049$ ), and LupusQoL-Burden to Others (29.2,  $P=.04$ ). The magnitude of the improvements in all domains (absolute and relative) is shown in Figure 3. Significantly greater improvement was seen in the intervention group compared with the control group. Results on an absolute basis are as follows: FACIT-F (34% intervention vs −1% control,  $P<.001$ ), BPI-SF-Pain severity (13% vs −16%,  $P=.049$ ), BPI-SF-Pain interference (25% vs 0%,  $P=.02$ ), and 4 LupusQoL measures, namely, pain (13% vs 0%,  $P=.004$ ), planning (17% vs 0%,  $P=.004$ ), emotional health (21% vs 4%,  $P=.02$ ), and fatigue (38% vs 13%,  $P<.001$ ).

**Table 5.** Per-protocol analysis of change in FACIT, BPI-SF, and LupusQoL domain scores from baseline to end of program.<sup>a</sup>

Domain	Within group					Between group	
	Count	Baseline	End of program (EOP)	Change in score (EOP – Baseline)	P-value	Difference <sup>b</sup>	P-value
<b>FACIT<sup>c</sup> -Fatigue (range 0-52, higher is better)</b>						18.0	<.001 <sup>f</sup>
Intervention	16	20.5 (10.3, 26.8)	43.5 (28.5, 47.8)	17.5 (4.8, 24.0)	.001 <sup>f</sup>		
Control	18	20.5 (14.0, 27.3)	22.0 (12.5, 28.3)	-0.5 (-4.3, 7.3)	.79		
<b>BPI-SF<sup>d</sup> -Pain Severity (range 0-10, lower is better)</b>						-1.9	.049 <sup>g</sup>
Intervention	16	4.8 (3.0, 6.5)	3.3 (1.6, 5.2)	-1.3 (-3.0, 0.4)	.02 <sup>g</sup>		
Control	18	3.9 (2.9, 5.9)	3.8 (2.6, 6.4)	0.6 (-1.3, 1.0)	.68		
<b>BPI-SF-Pain Interference (range 0-10, lower is better)</b>						-2.5	.02 <sup>f</sup>
Intervention	16	6.3 (4.0, 7.5)	2.0 (0.5, 5.3)	-2.5 (-4.4, -0.2)	.003 <sup>f</sup>		
Control	18	5.4 (4.1, 6.3)	4.9 (1.6, 6.5)	0.0 (-1.2, 0.8)	.64		
<b>LupusQoL<sup>e</sup> -Physical Health (range 0-100, higher is better)</b>						14.1	.049 <sup>g</sup>
Intervention	16	51.6 (34.4, 61.7)	71.9 (37.5, 93.0)	17.2 (0.0, 35.9)	.02 <sup>f</sup>		
Control	18	51.6 (27.3, 65.6)	48.4 (26.6, 71.1)	3.1 (-7.8, 10.2)	.66		
<b>LupusQoL-Pain (range 0-100, higher is better)</b>						12.5	.004 <sup>f</sup>
Intervention	16	50.0 (33.3, 66.7)	83.3 (47.9, 89.6)	12.5 (2.1, 39.6)	.03 <sup>g</sup>		
Control	18	58.3 (22.9, 75.0)	41.7 (16.7, 66.7)	0.0 (-16.7, 2.1)	.23		
<b>LupusQoL-Planning (range 0-100, higher is better)</b>						16.7	.004 <sup>f</sup>
Intervention	16	62.5 (20.8, 75.0)	83.3 (56.3, 100.0)	16.7 (0.0, 41.7)	.008 <sup>f</sup>		
Control	18	62.5 (25.0, 75.0)	41.7 (25.0, 77.1)	0.0 (-27.1, 8.3)	.19		
<b>LupusQoL-Burden to Others (range 0-100, higher is better)</b>						29.2	.04 <sup>g</sup>
Intervention	16	25.0 (2.1, 54.2)	79.2 (31.3, 83.3)	33.3 (0.0, 58.3)	.003 <sup>f</sup>		
Control	18	25.0 (0.0, 60.4)	41.7 (0.0, 77.1)	4.2 (0.0, 16.7)	.11		
<b>LupusQoL-Intimate Relationships (range 0-100, higher is better)</b>						25.0	.12
Intervention	11	75.0 (25.0, 75.0)	87.5 (75.0, 100.0)	25.0 (-12.5, 50.0)	.06		
Control	15	62.5 (25.0, 87.5)	50.0 (25.0, 87.5)	0.0 (-12.5, 12.5)	.92		
<b>LupusQoL-Emotional Health (range 0-100, higher is better)</b>						16.7	.02 <sup>f</sup>
Intervention	16	60.4 (34.4, 81.3)	83.3 (68.8, 99.0)	20.8 (4.2, 37.5)	.01 <sup>f</sup>		
Control	18	56.3 (35.4, 71.9)	56.3 (40.6, 67.7)	4.2 (-9.4, 12.5)	.57		
<b>LupusQoL-Body Image (range 0-100, higher is better)</b>						13.8	.09
Intervention	13	56.3 (14.4, 69.4)	87.5 (68.8, 95.0)	18.8 (13.1, 46.9)	.011 <sup>f</sup>		
Control	15	31.3 (20.0, 65.0)	40.0 (25.0, 70.0)	5.0 (0.0, 23.8)	.12		
<b>LupusQoL-Fatigue (range 0-100, higher is better)</b>						25.0	<.001 <sup>f</sup>
Intervention	16	28.1 (25.0, 53.1)	81.3 (64.1, 92.2)	37.5 (21.9, 48.4)	<.001 <sup>f</sup>		
Control	18	25.0 (10.9, 43.8)	34.4 (25.0, 53.1)	12.5 (-1.6, 20.3)	.03 <sup>g</sup>		

<sup>a</sup>Within-group values are median (25th percentile, 75th percentile). Nonparametric tests were chosen in order to require minimal distributional assumptions given the small sample size. Medians, 25th, and 75th percentile values are displayed as measures of central tendency and spread. The within-group *P*-value is from the Wilcoxon signed-rank test; the between-group *P*-value is from the Mann-Whitney *U* test; *P*-values themselves are unadjusted but the threshold for statistical significance is set using the Benjamini-Hochberg adjustment. LupusQoL-Intimate Relationships and LupusQoL-Body Image allow N/A responses.

<sup>b</sup>Difference in median change (intervention – control).

<sup>c</sup>FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F); 52-point scale with 0 (worst).

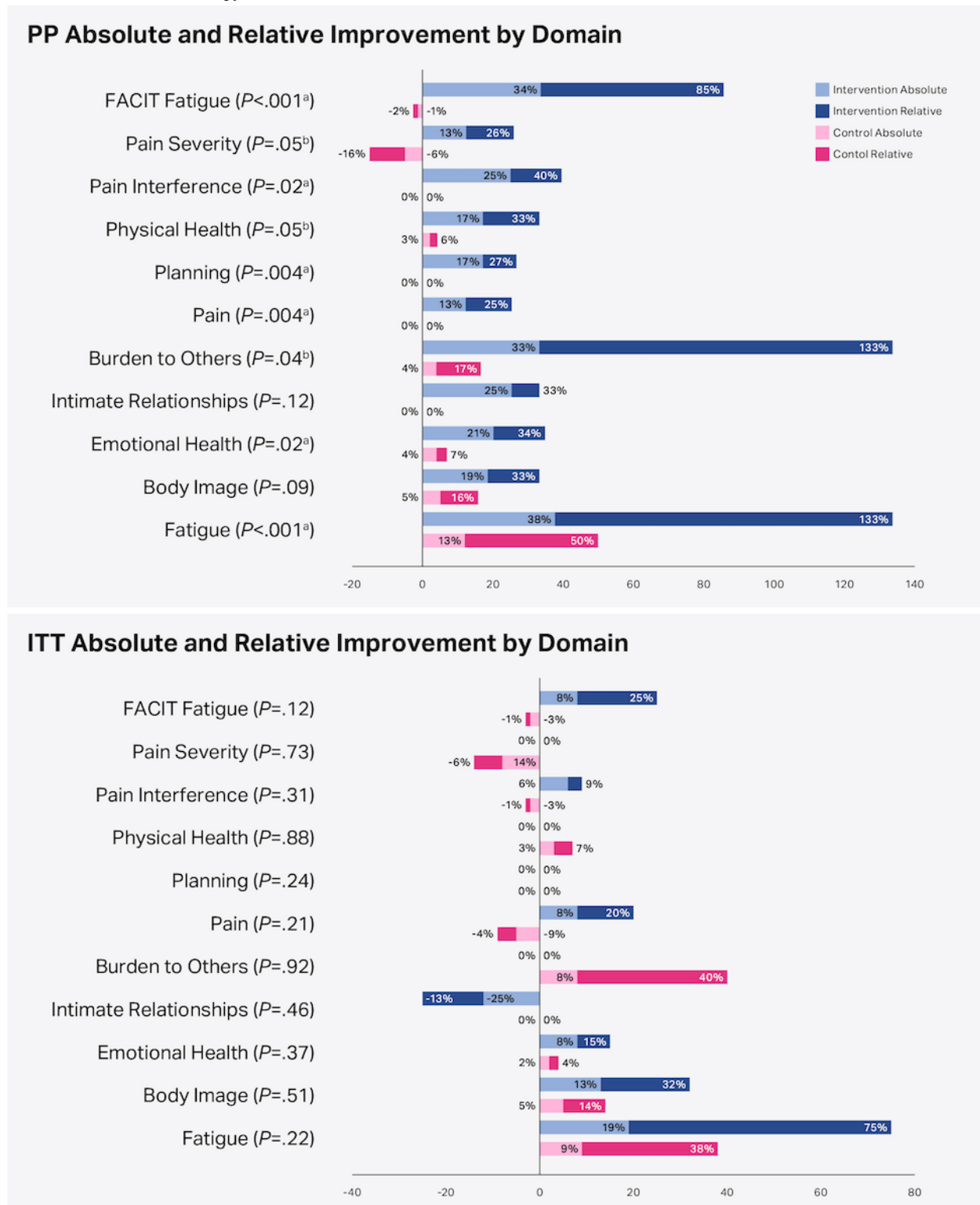
<sup>d</sup>BPI-SF: Brief Pain Inventory-Short Form; 10-point scale with 0 (best).

<sup>e</sup>LupusQoL: Lupus Quality of Life; 100-point scale with 0 (worst).

<sup>f</sup>Statistically significant after using the Benjamini–Hochberg adjustment to account for multiple comparisons.

<sup>g</sup>Statistically significant at an unadjusted 2-sided significance level of 5%.

**Figure 3.** Absolute and Relative Improvement by Domain. Absolute improvement was median change from baseline to endpoint divided by total possible domain score. Relative improvement was median change divided by the median baseline domain score. Changes in BPI-SF-pain interference and BPI-SF-pain severity are converted to positive % for consistency with other domains. *P*-values are from the Mann–Whitney *U* test comparing changes in score between intervention and control groups. *P*-values are unadjusted. Although both ITT intervention and control groups achieved significant improvement in some domains, when the groups were compared, no statistically significant differences were found. <sup>a</sup>Statistically significant after using the Benjamini–Hochberg adjustment <sup>b</sup>Statistically significant at an unadjusted two-sided significance level of 5%. FACIT: Functional Assessment of Chronic Illness Therapy; ITT: intention to treat.



### Per-Protocol Analysis of Longitudinal Change

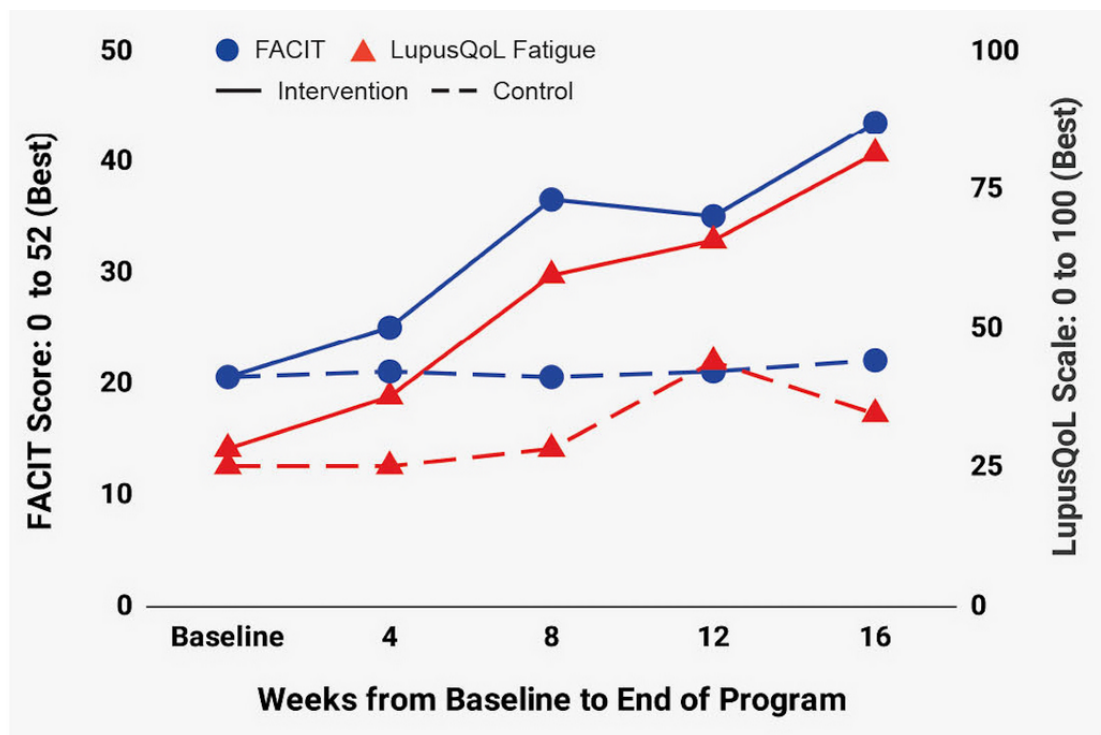
Figures 4–7 depict change over time in FACIT-F, LupusQoL, and BPI-SF in both the ITT and PP groups. Improvement in the

intervention group started in the first 4 weeks and continued through week 16 with the following exceptions: FACIT-F and LupusQoL-Fatigue had slightly higher improvement rates between weeks 4 and 8 and weeks 12 and 16 (Figure 4); a

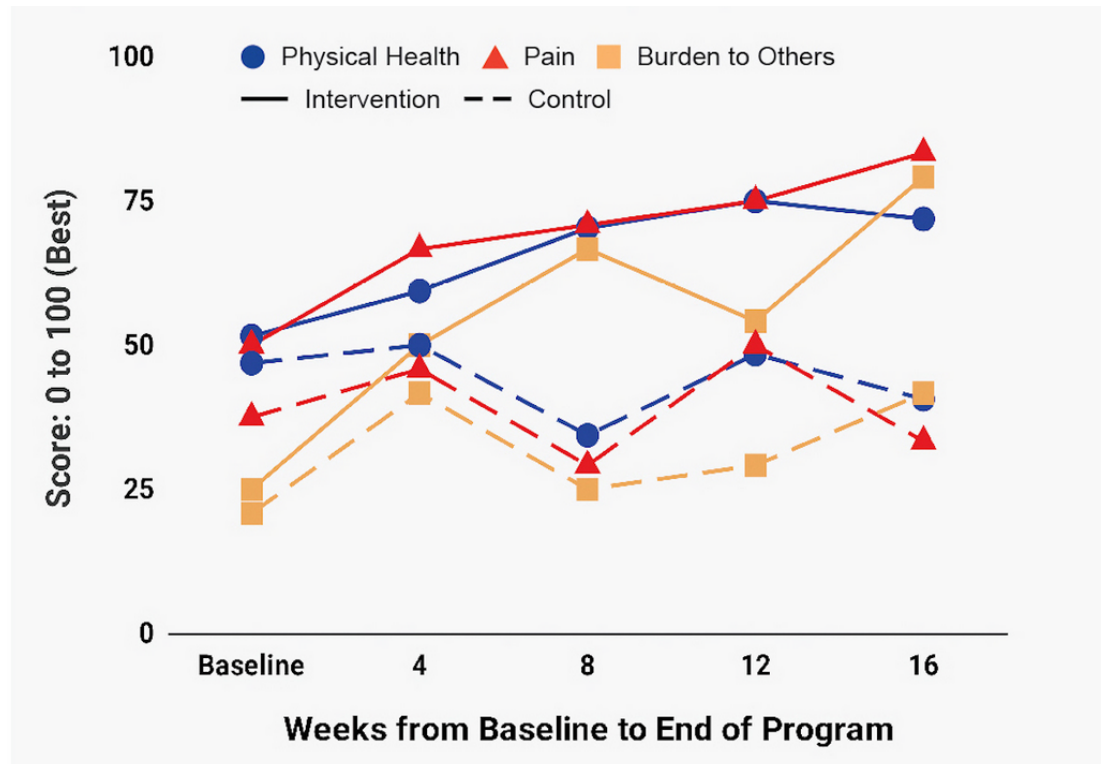


significant portion of pain reduction occurred between weeks 12 and 16 (Figure 7). The control group experienced modest improvement in LupusQoL-Fatigue at 12 weeks which diminished by week 16, whereas all other domains remained largely unchanged or deteriorated over the course of the 16 weeks.

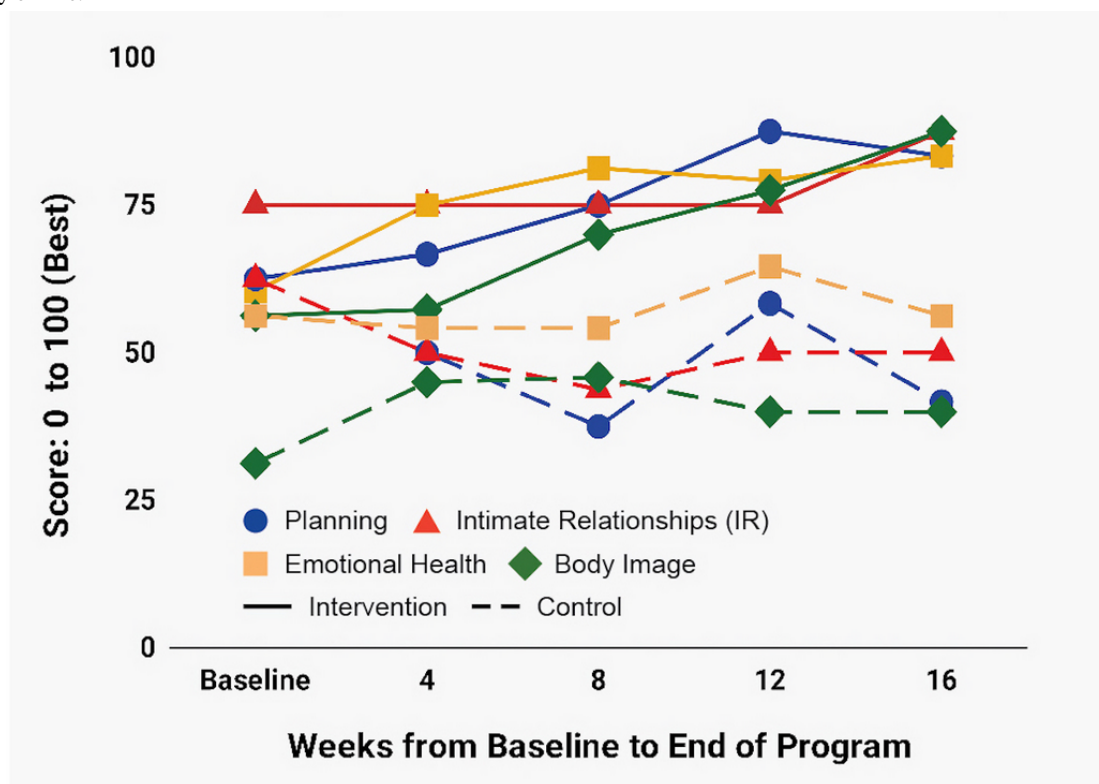
**Figure 4.** Change over time in FACIT-Fatigue and LupusQoL-Fatigue. FACIT: Functional Assessment of Chronic Illness Therapy; LupusQoL: Lupus Quality of Life.



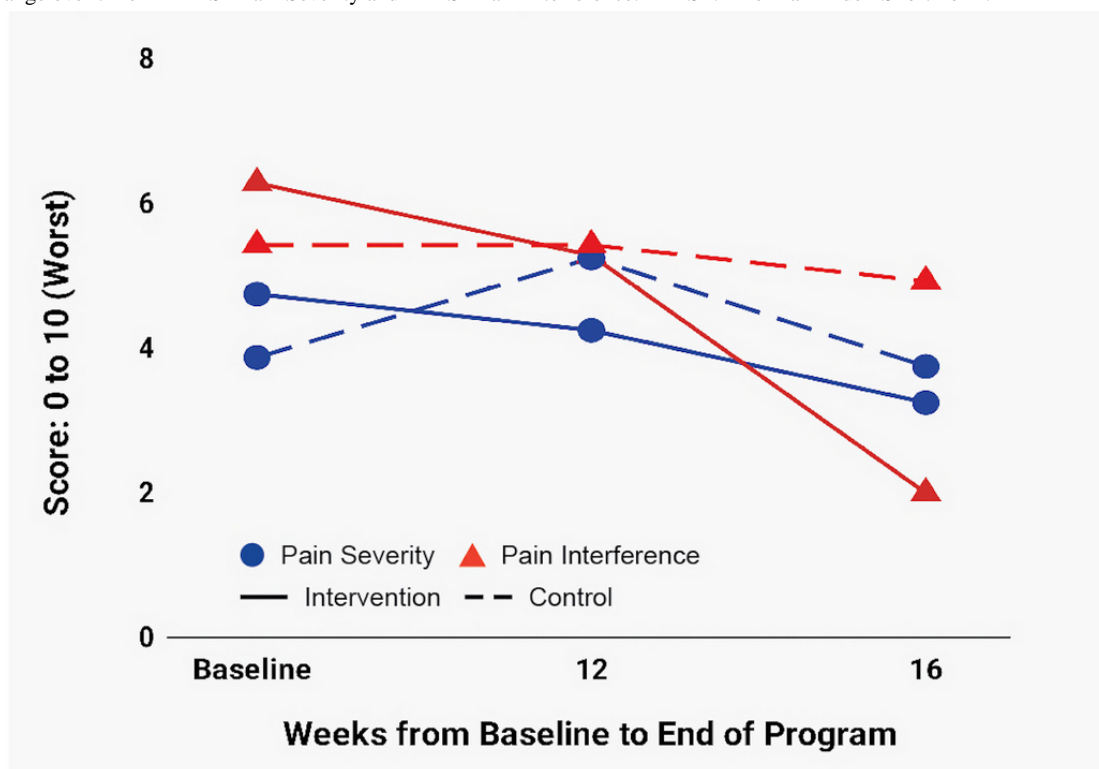
**Figure 5.** Change over time in LupusQoL-Physical Health, LupusQoL-Pain and LupusQoL-Burden to Others. LupusQoL: Lupus Quality of Life.



**Figure 6.** Change over time in LupusQoL-Planning, LupusQoL-Relationships, LupusQoL-Emotional Health and LupusQoL-Body Image. LupusQoL: Lupus Quality of Life.



**Figure 7.** Change over time in BPI-SF Pain Severity and BPI-SF Pain Interference. BPI-SF: Brief Pain Index-Short Form.



### Frequency of Tracked Symptoms, Triggers, and Interventions

Table 6 displays the top 3 participant-tracked inputs (symptoms and suspected triggers) and top 4 coach-recommended interventions in the ITT and PP groups. Among both groups,

the 3 most frequently tracked symptoms were fatigue, joint pain, and brain fog. The most common triggers in both groups were dairy, gluten, and nightshades (a family of plants that include potatoes, tomatoes, capsicum, bell peppers, eggplant, and tobacco). The most commonly recommended interventions, aside from ensuring adequate hydration, were dietary elimination

of triggers and the addition of digestive enzyme supplements, apple cider vinegar, and protein shakes.

**Table 6.** Frequency of tracked symptoms, triggers, and interventions.<sup>a</sup>

Variable	Intention to treat		Per protocol	
	n	Relative frequency (%)	n	Relative frequency (%)
<b>Interventions</b>				
Dietary elimination	17	68	16	100
Digestive enzymes	19	76	16	100
Protein shake	17	68	15	94
Apple cider vinegar	15	60	14	88
<b>Triggers</b>				
Dairy	19	76	14	88
Gluten	14	56	11	69
Nightshades	7	28	6	38
<b>Symptoms</b>				
Joint pain	6	24	6	38
Brain fog	6	24	6	38
Fatigue	7	28	5	31

<sup>a</sup>Total population sampled includes intention to treat (N=25) and per protocol (N=16).

## Adverse Events

No adverse events attributable to the intervention occurred. Four participants (3 from the control group and 1 from the intervention group) experienced SLE exacerbations requiring pulse steroids within the last 4 weeks of the study.

## Discussion

### Principal Findings

To the authors' knowledge, this is the first study to show that a digital therapeutic intervention targeting dietary, environmental, and lifestyle factors can improve HRQoL when added to usual care in patients with SLE. Participants who completed the 16-week protocol showed continuous improvement across all HRQoL domains and statistically greater improvement than those receiving usual care alone for the majority of domains. Of particular interest is the significant improvement noted in fatigue. Fatigue is one of the most debilitating symptoms reported by patients with SLE; is highly correlated with work disability [38], workplace absence (absenteeism), or impaired workplace performance (presenteeism) [39,40]; and is frustratingly recalcitrant to treatment. Although not formally assessed (and therefore only serves as a point for further exploration in future studies), qualitative data collected via coaching notes revealed that 2 participants in the PP intervention group who had been on disability at the start of the study (15 years and 3 years) felt ready for work.

In this SLE population, the most common triggers identified as correlating negatively with symptoms and leading to improvement upon elimination were all dietary—the top 3 being dairy, gluten, and nightshades. While these findings do not

provide conclusive evidence linking dairy, gluten, and nightshades to SLE, accumulating data from in vitro, animal, and human studies support the need for ongoing investigation into these potential triggers [41-43].

Elimination of food triggers identified by the program's software as aggravating symptoms was central to the therapeutic approach of this platform. In addition, a variety of low-risk, nutritional interventions not previously studied in an SLE population were frequently incorporated into the participants' personalized plans. These interventions were primarily aimed at improving the participants' numerous digestive and energy-level complaints which weigh heavily on HRQoL in SLE and included digestive enzymes, small amounts of apple cider vinegar, and protein shakes.

Progress in the care of patients with SLE has been slow, largely attributable to inherent disease characteristics as well as health care access and socioeconomic obstacles. As discussed, disease heterogeneity is perhaps the most significant obstacle to progress. Other barriers to advancement are lack of diagnostic, predictive, prognostic, and drug-response biomarkers; ineffective management of SLE due to social determinants of care in predominantly lower socioeconomic status areas; and lack of treatment adherence [24]. Nontraditional solutions to these challenges should be explored and digital therapeutics offer one such novel approach. The digital therapeutic intervention tested here focused on identifying and eliminating dietary, environmental, and lifestyle triggers of SLE as an adjunct to usual care. This approach implicitly takes disease heterogeneity into account, leverages the growing understanding of the role environment plays in initiating and propagating SLE, and personalizes each patient's recommendations supported by software data analytics. This personalized approach is especially intriguing as it applies to dietary interventions in SLE. A recent

review article assessing the significance and impact of dietary factors on SLE pathogenesis found that small and personalized improvements in diet could alter the clinical status of patients with SLE and concluded that “proper diet in SLE can help preserve the body’s homeostasis, increase the period of remission, prevent adverse effects of medication [especially systemic corticotherapy] and improve the patient’s physical and mental well-being” [44].

With on-going research, digital therapeutics may hold the key to overcoming many barriers to SLE care. The enormity of data that can be collected and analyzed via a digital therapeutic platform has the potential to help identify new SLE biomarkers. Aspects of care that prove difficult to deliver to disadvantaged populations with the traditional medical model may be made more accessible. Importantly, if larger studies validate these preliminary findings and build on this work by demonstrating improvements in disease activity measures (eg, Systemic Lupus Erythematosus Disease Activity Index [SLEDAI], SLE Responder Index [SRI]), then dietary and lifestyle interventions delivered in conjunction with a digital therapeutic device may allow for more selective and conservative use of costly, potentially dangerous immune-modulating drugs. Medication changes were not formally followed in this study but information from coaching notes revealed that 5 study participants in the PP intervention group were able to reduce or discontinue immune-modulating medications. In addition, several participants reduced or discontinued use of multiple symptom-relieving medications (including over-the-counter and prescription drugs for gastrointestinal symptoms, pain, depression, and anxiety). As medication usage was not a prespecified outcome in this trial and was not formally assessed, conclusions about the impact of this intervention on medication usage cannot be made. However, if these results are reproduced in larger studies (in which medication information is formally collected and analyzed) the implications of drug reduction alone are important. Polypharmacy is highly prevalent in SLE [45], is associated with elevated risk of adverse drug events, and was shown in a 2017 meta-analysis to be linked to increased mortality [46].

It was not possible in this small pilot trial to examine the underlying physiological mechanisms responsible for patients’ improvements. Provocative findings from several lines of research compel one to consider the effects that the collection of lifestyle modifications, particularly dietary changes, may have had on the health of the intestinal epithelium and the gut microbiome. In animal [14,47–49] and human [50,51] studies, mounting evidence points to a central role of the intestinal epithelial barrier and related diversity and function of the gut microbiome in autoimmune disease. In 2014, the National Institutes of Health (NIH) launched the Integrative Human Microbiome Project “to generate resources to permit comprehensive characterization of the human microbiota to further our understanding of how the microbiome impacts human health and disease” [51]. One of the 3 microbiome-associated conditions which are being explored is autoimmune in nature, namely, inflammatory bowel disease. As microbiome characteristics have also been implicated in SLE, it would be valuable in future studies of this digital

therapeutic to assess microbiome composition before and after the intervention.

This exploratory pilot study has many limitations and the results should be interpreted in this context. Physician-scored, validated SLE disease activity measures (such as British Isles Lupus Assessment Group or SLEDAI) were not captured, limiting the capacity to assess disease severity at baseline and change in disease activity by strict clinical criteria throughout the study. In this exploratory pilot study, limited budget and research manpower restricted the ability to pursue this depth of data collection. While inclusion of such clinical disease activity scores in future, larger studies is planned and will provide critical insights, absence of these measures should not diminish the relevance of HRQoL outcomes. The debilitating symptoms, toxicity of immune-modulating treatments, unpredictability of disease activity, and fear of serious, even life-threatening manifestations associated with SLE have a profound impact on HRQoL across multiple domains. These features are not adequately captured by clinical measures of disease activity, which previous research has shown to have poor correlation with patient assessment of HRQoL [52]. Furthermore, HRQoL has been found to be associated with treatment adherence and health care utilization in patients with SLE [53]. The PROMs utilized in this study were chosen to capture many of the diverse domains that contribute to the complex concept of HRQoL.

Selection bias may have been introduced by heavy recruitment from online SLE and other autoimmune patient websites and therefore the study group may not be representative of the general SLE population. However, the number of patients who seek online medical advice is large; continues to grow; and crosses gender, age, and socioeconomic differences [54]. Selection bias may have also been introduced by the requirement of owning a smartphone. Smartphone ownership, however, has become increasingly common across gender, race, education, and economic levels [55], hopefully minimizing this bias. But, in future studies this can be addressed by providing smartphones for those in need.

This study failed to show statistically significant between-group differences in any measured domain in the ITT analysis. These results were affected by disproportionate attrition from the intervention group early in the study (Figure 2). Six intervention participants (6/25, 24%) left the study after having completed 0 to 1 of 16 sessions, whereas only 1 control group participant was lost early. Missing data from participants who dropped out of the intervention group were populated with the worst observed scores for that time point, thus biasing toward the null hypothesis. Furthermore, while 1 participant in the intervention group did receive pulse steroids within the last 4 weeks of the study period, excluding her from PP analysis, 3 patients in the control group also received pulse steroids in this time frame. Any positive effect this treatment had on outcomes would have biased toward the null hypothesis. The attrition rate may speak to the requirements inherent in this type of intervention, namely, that participants need to be motivated and engaged with an aptitude for regular app use and an interest in attending weekly coaching sessions. In future studies, early attrition will be addressed by building a run-in period into the design to help mitigate this issue. That this intervention has shown an 83%



completion rate (participating in at least twelve of sixteen coaching sessions) in 70 autoimmune patients from a private insurance cohort is reassuring that the program has acceptable usability (internal data).

As potential adverse events were collected for the control group only at the end of the study (as opposed to the intervention group who were queried about potential adverse events on weekly coaching calls), these data may have been subject to recall bias. Future studies, which are planned to include a sham app and weekly sham coaching calls (see below), will overcome this potential bias.

There was no sham app or sham coaching in this study. Digital apps and health coaching alike are intrinsically engaging, thus vulnerable to the placebo response. It is not possible to tell to what extent HRQoL improvements were influenced by this engagement and patient expectations rather than the program interventions. Future studies should include a convincing sham app and interaction between controls and a health coach at the same frequency as that which occurs with the intervention group. Development of a sham app and sham coaching protocols are underway.

## Conclusions

In conclusion, the digital therapeutic and coaching intervention tested in this pilot trial resulted in statistically significant, clinically meaningful improvements when added to usual care, compared with usual care alone, in several measures of HRQoL (including pain and fatigue) in adult patients with SLE. The study demonstrated that an adaptive, multifaceted intervention which aims at identifying and limiting each SLE participant's specific dietary and environmental triggers can improve symptoms and HRQoL without additional pharmaceutical manipulation of the immune system. These promising results stimulate a call for a larger study that includes measurement of validated SLE disease activity measures, sham controls, analysis of the biological mechanisms that underlie the improvements, and long-term follow-up of patients to confirm sustained gains in HRQoL. Broad adoption of the intervention could assist in building a database of SLE triggers which could deepen the understanding of the etiology of this disease and potentially contribute to SLE prevention in the future. Finally, given the expected role of diet, environment, and lifestyle in other autoimmune diseases, many of which have gaps in care similar to those in SLE, studies of the intervention's application to other autoimmune conditions is warranted.

## Acknowledgments

We express our deepest gratitude to Vanessa Rudin for managing the research team and study deliverables and to Dr Vincent Campasano, who oversaw the safety of the trial. We extend our appreciation to Mette Dyhrberg (Founder, Mymee, Inc; Certified Integrative Health and Wellness Coach) for her compassionate and skilled health coaching services. We also thank Charlotta Norgaard (Lupus Friends and Family) for promoting the study to potential participants and providing feedback from her Facebook group members about the recruiting process and materials; Niecy Rivera (Flare Fighters), Aaron Jun and Lily Stairs (Clara Health), Arif Sorathia and Brett Adelman (LupusCorner), Aaron Abend (Autoimmune Registry), and Skylar Schone (MyID) for promoting the study to potential participants. And we thank the participants for having the courage to try something different. This study is supported by funds from the sponsor, Mymee, Inc.

## Authors' Contributions

All co-authors contributed substantially to the analysis and interpretation of the data and provided important intellectual input. FK and NG wrote the first draft of the manuscript with editing of analysis, content, and format by YC and RM. All authors and the statistician agreed that the accuracy and integrity of the work has been appropriately investigated and resolved, and all approved the final version of the manuscript. The corresponding author had full access to the data and had final responsibility for the decision to submit for publication. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

## Conflicts of Interest

FK and RM have received standard financial stipends from Mymee, Inc. and have no other disclosures to make. NG and YC have nothing to declare.

Multimedia Appendix 1  
Supplement Protocol.

[DOCX File, 17 KB - [jmir\\_v22i10e23868\\_app1.docx](#) ]

Multimedia Appendix 2

CONSORT EHEALTH checklist (v 1.6.2).

[PDF File (Adobe PDF File), 101 KB - [jmir\\_v22i10e23868\\_app2.pdf](#) ]

## References



1. Lupus Facts and Statistics. 2016 Oct 06. URL: <https://www.lupus.org/resources/lupus-facts-and-statistics#:~:text=The%20Lupus%20Foundation%20of%20America,mostly%20women%20of%20childbearing%20age> [accessed 2020-08-03]
2. Lightstone L, Doria A, Wilson H, Ward FL, Larosa M, Bargman JM. Can we manage lupus nephritis without chronic corticosteroids administration? *Autoimmun Rev* 2018 Jan;17(1):4-10. [doi: [10.1016/j.autrev.2017.11.002](https://doi.org/10.1016/j.autrev.2017.11.002)] [Medline: [29108828](https://pubmed.ncbi.nlm.nih.gov/29108828/)]
3. Pons-Estel GJ, Alarcón GS, Scofield L, Reinlib L, Cooper GS. Understanding the epidemiology and progression of systemic lupus erythematosus. *Semin Arthritis Rheum* 2010 Feb;39(4):257-268 [FREE Full text] [doi: [10.1016/j.semarthrit.2008.10.007](https://doi.org/10.1016/j.semarthrit.2008.10.007)] [Medline: [19136143](https://pubmed.ncbi.nlm.nih.gov/19136143/)]
4. Bernatsky S, Boivin J, Joseph L, Manzi S, Ginzler E, Gladman DD, et al. Mortality in systemic lupus erythematosus. *Arthritis Rheum* 2006 Aug;54(8):2550-2557 [FREE Full text] [doi: [10.1002/art.21955](https://doi.org/10.1002/art.21955)] [Medline: [16868977](https://pubmed.ncbi.nlm.nih.gov/16868977/)]
5. Yen EY, Shaheen M, Woo JMP, Mercer N, Li N, McCurdy DK, et al. 46-Year Trends in Systemic Lupus Erythematosus Mortality in the United States, 1968 to 2013: A Nationwide Population-Based Study. *Ann Intern Med* 2017 Dec 05;167(11):777-785 [FREE Full text] [doi: [10.7326/M17-0102](https://doi.org/10.7326/M17-0102)] [Medline: [29086801](https://pubmed.ncbi.nlm.nih.gov/29086801/)]
6. Almehed K, Carlsten H, Forsblad-d'Elia H. Health-related quality of life in systemic lupus erythematosus and its association with disease and work disability. *Scand J Rheumatol* 2010;39(1):58-62. [doi: [10.3109/03009740903124408](https://doi.org/10.3109/03009740903124408)] [Medline: [20132072](https://pubmed.ncbi.nlm.nih.gov/20132072/)]
7. Olesińska M, Saletra A. Quality of life in systemic lupus erythematosus and its measurement. *Reumatologia* 2018;56(1):45-54 [FREE Full text] [doi: [10.5114/reum.2018.74750](https://doi.org/10.5114/reum.2018.74750)] [Medline: [29686443](https://pubmed.ncbi.nlm.nih.gov/29686443/)]
8. Utset TO, Chohan S, Booth SA, Laughlin JC, Kocherginsky M, Schmitz A. Correlates of formal work disability in an urban university systemic lupus erythematosus practice. *J Rheumatol* 2008 Jun;35(6):1046-1052. [Medline: [18464309](https://pubmed.ncbi.nlm.nih.gov/18464309/)]
9. Yen EY, Singh RR. Brief Report: Lupus-An Unrecognized Leading Cause of Death in Young Females: A Population-Based Study Using Nationwide Death Certificates, 2000-2015. *Arthritis Rheumatol* 2018 Aug;70(8):1251-1255 [FREE Full text] [doi: [10.1002/art.40512](https://doi.org/10.1002/art.40512)] [Medline: [29671279](https://pubmed.ncbi.nlm.nih.gov/29671279/)]
10. Mazzone R, Zwergel C, Artico M, Taurone S, Ralli M, Greco A, et al. The emerging role of epigenetics in human autoimmune disorders. *Clin Epigenetics* 2019 Feb 26;11(1):34 [FREE Full text] [doi: [10.1186/s13148-019-0632-2](https://doi.org/10.1186/s13148-019-0632-2)] [Medline: [30808407](https://pubmed.ncbi.nlm.nih.gov/30808407/)]
11. Riscuta G, Xi D, Pierre-Victor D, Starke-Reed P, Khalsa J, Duffy L. Diet, Microbiome, and Epigenetics in the Era of Precision Medicine. *Methods Mol Biol* 2018;1856:141-156. [doi: [10.1007/978-1-4939-8751-1\\_8](https://doi.org/10.1007/978-1-4939-8751-1_8)] [Medline: [30178250](https://pubmed.ncbi.nlm.nih.gov/30178250/)]
12. Pan Y, Ke H, Yan Z, Geng Y, Asner N, Palani S, et al. The western-type diet induces anti-HMGB1 autoimmunity in Apoe(-/-) mice. *Atherosclerosis* 2016 Aug;251:31-38 [FREE Full text] [doi: [10.1016/j.atherosclerosis.2016.05.027](https://doi.org/10.1016/j.atherosclerosis.2016.05.027)] [Medline: [27240253](https://pubmed.ncbi.nlm.nih.gov/27240253/)]
13. Mu Q, Zhang H, Luo XM. SLE: Another Autoimmune Disorder Influenced by Microbes and Diet? *Front Immunol* 2015;6:608 [FREE Full text] [doi: [10.3389/fimmu.2015.00608](https://doi.org/10.3389/fimmu.2015.00608)] [Medline: [26648937](https://pubmed.ncbi.nlm.nih.gov/26648937/)]
14. Parks CG, de Souza Espindola Santos A, Barbhuiya M, Costenbader KH. Understanding the role of environmental factors in the development of systemic lupus erythematosus. *Best Pract Res Clin Rheumatol* 2017 Jun;31(3):306-320 [FREE Full text] [doi: [10.1016/j.berh.2017.09.005](https://doi.org/10.1016/j.berh.2017.09.005)] [Medline: [29224673](https://pubmed.ncbi.nlm.nih.gov/29224673/)]
15. Hardy TM, Tollefsbol TO. Epigenetic diet: impact on the epigenome and cancer. *Epigenomics* 2011 Aug;3(4):503-518 [FREE Full text] [doi: [10.2217/epi.11.71](https://doi.org/10.2217/epi.11.71)] [Medline: [22022340](https://pubmed.ncbi.nlm.nih.gov/22022340/)]
16. Young KA, Munroe ME, Harley JB, Guthridge JM, Kamen DL, Galkens GS, et al. Less than 7 hours of sleep per night is associated with transitioning to systemic lupus erythematosus. *Lupus* 2018 Aug;27(9):1524-1531 [FREE Full text] [doi: [10.1177/0961203318778368](https://doi.org/10.1177/0961203318778368)] [Medline: [29804502](https://pubmed.ncbi.nlm.nih.gov/29804502/)]
17. Gaine ME, Chatterjee S, Abel T. Sleep Deprivation and the Epigenome. *Front Neural Circuits* 2018;12:14 [FREE Full text] [doi: [10.3389/fncir.2018.00014](https://doi.org/10.3389/fncir.2018.00014)] [Medline: [29535611](https://pubmed.ncbi.nlm.nih.gov/29535611/)]
18. Hsiao Y, Chen Y, Tseng C, Wu L, Lin W, Su VY, et al. Sleep disorders and increased risk of autoimmune diseases in individuals without sleep apnea. *Sleep* 2015 Apr 01;38(4):581-586 [FREE Full text] [doi: [10.5665/sleep.4574](https://doi.org/10.5665/sleep.4574)] [Medline: [25669189](https://pubmed.ncbi.nlm.nih.gov/25669189/)]
19. Song H, Fang F, Tomasson G, Arnberg FK, Mataix-Cols D, Fernández de la Cruz L, et al. Association of Stress-Related Disorders With Subsequent Autoimmune Disease. *JAMA* 2018 Jun 19;319(23):2388-2400 [FREE Full text] [doi: [10.1001/jama.2018.7028](https://doi.org/10.1001/jama.2018.7028)] [Medline: [29922828](https://pubmed.ncbi.nlm.nih.gov/29922828/)]
20. Gudsnuik K, Champagne FA. Epigenetic influence of stress and the social environment. *ILAR J* 2012;53(3-4):279-288 [FREE Full text] [doi: [10.1093/ilar.53.3-4.279](https://doi.org/10.1093/ilar.53.3-4.279)] [Medline: [23744967](https://pubmed.ncbi.nlm.nih.gov/23744967/)]
21. Lee RS, Tamashiro K, Yang X, Purcell RH, Harvey A, Willour VL, et al. Chronic corticosterone exposure increases expression and decreases deoxyribonucleic acid methylation of Fkbp5 in mice. *Endocrinology* 2010 Sep;151(9):4332-4343 [FREE Full text] [doi: [10.1210/en.2010-0225](https://doi.org/10.1210/en.2010-0225)] [Medline: [20668026](https://pubmed.ncbi.nlm.nih.gov/20668026/)]
22. Simpson R, Kunz H, Agha N, Graff R. Exercise and the Regulation of Immune Functions. *Prog Mol Biol Transl Sci* 2015;135:355-380. [doi: [10.1016/bs.pmbts.2015.08.001](https://doi.org/10.1016/bs.pmbts.2015.08.001)] [Medline: [26477922](https://pubmed.ncbi.nlm.nih.gov/26477922/)]
23. Zhang FF, Cardarelli R, Carroll J, Zhang S, Fulda KG, Gonzalez K, et al. Physical activity and global genomic DNA methylation in a cancer-free population. *Epigenetics* 2011 Mar;6(3):293-299 [FREE Full text] [doi: [10.4161/epi.6.3.14378](https://doi.org/10.4161/epi.6.3.14378)] [Medline: [21178401](https://pubmed.ncbi.nlm.nih.gov/21178401/)]

24. Manzi S, Raymond S, Tse K, Peña Y, Anderson A, Arntsen K, et al. Global consensus building and prioritisation of fundamental lupus challenges: the ALPHA project. *Lupus Sci Med* 2019;6(1):e000342. [doi: [10.1136/lupus-2019-000342](https://doi.org/10.1136/lupus-2019-000342)] [Medline: [31413854](https://pubmed.ncbi.nlm.nih.gov/31413854/)]
25. Examples of Device Software Functions the FDA Regulates. Silver Spring, MD: Food and Drug Administration; 2019 Sep 26. URL: <https://www.fda.gov/medical-devices/device-software-functions-including-mobile-medical-applications/examples-device-software-functions-fda-regulates> [accessed 2020-08-02]
26. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *NPJ Digit Med* 2019;2(1):40 [FREE Full text] [doi: [10.1038/s41746-019-0090-4](https://doi.org/10.1038/s41746-019-0090-4)] [Medline: [30868107](https://pubmed.ncbi.nlm.nih.gov/30868107/)]
27. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. *NPJ Digit Med* 2020;3:14 [FREE Full text] [doi: [10.1038/s41746-019-0212-z](https://doi.org/10.1038/s41746-019-0212-z)] [Medline: [32047860](https://pubmed.ncbi.nlm.nih.gov/32047860/)]
28. Us in Lupus. URL: <https://www.usinlupus.com/tools/mylupuslog.html> [accessed 2020-08-02]
29. Track the Symptoms of Lupus with this Symptom Tracking App. URL: <http://careclinic.io/lupus-symptom-tracker> [accessed 2020-08-02]
30. My Lupus Log. URL: <http://myhealthapps.net/app/details/410/my-lupus-log>
31. Mango Health. URL: <https://www.mangohealth.com/> [accessed 2020-09-18]
32. McElhone K, Abbott J, Shelmerdine J, Bruce IN, Ahmad Y, Gordon C, et al. Development and validation of a disease-specific health-related quality of life measure, the LupusQoL, for adults with systemic lupus erythematosus. *Arthritis Rheum* 2007 Aug 15;57(6):972-979 [FREE Full text] [doi: [10.1002/art.22881](https://doi.org/10.1002/art.22881)] [Medline: [17665467](https://pubmed.ncbi.nlm.nih.gov/17665467/)]
33. Lai J, Beaumont JL, Ogale S, Brunetta P, Cella D. Validation of the functional assessment of chronic illness therapy-fatigue scale in patients with moderately to severely active systemic lupus erythematosus, participating in a clinical trial. *J Rheumatol* 2011 Apr;38(4):672-679. [doi: [10.3899/jrheum.100799](https://doi.org/10.3899/jrheum.100799)] [Medline: [21239746](https://pubmed.ncbi.nlm.nih.gov/21239746/)]
34. Mease PJ, Spaeth M, Clauw DJ, Arnold LM, Bradley LA, Russell IJ, et al. Estimation of minimum clinically important difference for pain in fibromyalgia. *Arthritis Care Res (Hoboken)* 2011 Jun;63(6):821-826 [FREE Full text] [doi: [10.1002/acr.20449](https://doi.org/10.1002/acr.20449)] [Medline: [21312349](https://pubmed.ncbi.nlm.nih.gov/21312349/)]
35. Izadi Z, Gandrup J, Katz PP, Yazdany J. Patient-reported outcome measures for use in clinical trials of SLE: a review. *Lupus Sci Med* 2018;5(1):e000279. [doi: [10.1136/lupus-2018-000279](https://doi.org/10.1136/lupus-2018-000279)] [Medline: [30167315](https://pubmed.ncbi.nlm.nih.gov/30167315/)]
36. Naegeli AN, Tomaszewski EL, Al Sawah S. Psychometric validation of the Brief Pain Inventory-Short Form in patients with systemic lupus erythematosus in the United States. *Lupus* 2015 Nov;24(13):1377-1383. [doi: [10.1177/0961203315588972](https://doi.org/10.1177/0961203315588972)] [Medline: [26038345](https://pubmed.ncbi.nlm.nih.gov/26038345/)]
37. McElhone K, Abbott J, Sutton C, Mullen M, Lanyon P, Rahman A, et al. Sensitivity to Change and Minimal Important Differences of the LupusQoL in Patients With Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)* 2016 Oct;68(10):1505-1513 [FREE Full text] [doi: [10.1002/acr.22850](https://doi.org/10.1002/acr.22850)] [Medline: [26816223](https://pubmed.ncbi.nlm.nih.gov/26816223/)]
38. Allaire S. Is work disability associated with systemic lupus erythematosus modifiable? *J Rheumatol* 2008 Jun;35(6):953-955. [Medline: [18528948](https://pubmed.ncbi.nlm.nih.gov/18528948/)]
39. Scofield L, Reinlib L, Alarcón GS, Cooper GS. Employment and disability issues in systemic lupus erythematosus: a review. *Arthritis Rheum* 2008 Oct 15;59(10):1475-1479 [FREE Full text] [doi: [10.1002/art.24113](https://doi.org/10.1002/art.24113)] [Medline: [18821664](https://pubmed.ncbi.nlm.nih.gov/18821664/)]
40. Connolly D, Fitzpatrick C, O'Toole L, Doran M, O'Shea F. Impact of Fatigue in Rheumatic Diseases in the Work Environment: A Qualitative Study. *Int J Environ Res Public Health* 2015 Oct 28;12(11):13807-13822 [FREE Full text] [doi: [10.3390/ijerph121113807](https://doi.org/10.3390/ijerph121113807)] [Medline: [26516896](https://pubmed.ncbi.nlm.nih.gov/26516896/)]
41. Arroyo Hornero R, Hamad I, Côté-Real B, Kleinewietfeld M. The Impact of Dietary Components on Regulatory T Cells and Disease. *Front Immunol* 2020;11:253 [FREE Full text] [doi: [10.3389/fimmu.2020.00253](https://doi.org/10.3389/fimmu.2020.00253)] [Medline: [32153577](https://pubmed.ncbi.nlm.nih.gov/32153577/)]
42. Hussain N. N282: Effect of gluten containing diet on pristane induced lupus prone mice. *Lupus Science and Medicine* 2017;4(Suppl\_1):A129-A130 [FREE Full text] [doi: [10.1136/lupus-2017-000215.282](https://doi.org/10.1136/lupus-2017-000215.282)]
43. Coucke F. Food intolerance in patients with manifest autoimmunity. *Observational study. Autoimmun Rev* 2018 Nov;17(11):1078-1080. [doi: [10.1016/j.autrev.2018.05.011](https://doi.org/10.1016/j.autrev.2018.05.011)] [Medline: [30213697](https://pubmed.ncbi.nlm.nih.gov/30213697/)]
44. Constantin M, Nita IE, Olteanu R, Constantin T, Bucur S, Matei C, et al. Significance and impact of dietary factors on systemic lupus erythematosus pathogenesis. *Exp Ther Med* 2019 Feb;17(2):1085-1090 [FREE Full text] [doi: [10.3892/etm.2018.6986](https://doi.org/10.3892/etm.2018.6986)] [Medline: [30679978](https://pubmed.ncbi.nlm.nih.gov/30679978/)]
45. Seguin D, Peschken C, Dolovich C, Grymonpre R, St JP, Tisseverasinghe A. Polypharmacy and potentially inappropriate medication use in young vs older adults with SLE (abstract). 2019 Presented at: 2019 ACR/ARP Annual Meeting; November 8-13, 2019; Atlanta, GA URL: <https://acrabstracts.org/abstract/polypharmacy-and-potentially-inappropriate-medication-use-in-young-versus-older-adults-with-sle/> [doi: [10.21203/rs.2.10158/v4](https://doi.org/10.21203/rs.2.10158/v4)]
46. Leelakanok N, Holcombe AL, Lund BC, Gu X, Schweizer ML. Association between polypharmacy and death: A systematic review and meta-analysis. *J Am Pharm Assoc (2003)* 2017;57(6):729-738.e10. [doi: [10.1016/j.japh.2017.06.002](https://doi.org/10.1016/j.japh.2017.06.002)] [Medline: [28784299](https://pubmed.ncbi.nlm.nih.gov/28784299/)]
47. Manfredo Vieira S, Hiltensperger M, Kumar V, Zegarar-Ruiz D, Dehner C, Khan N, et al. Translocation of a gut pathobiont drives autoimmunity in mice and humans. *Science* 2018 Mar 09;359(6380):1156-1161 [FREE Full text] [doi: [10.1126/science.aar7201](https://doi.org/10.1126/science.aar7201)] [Medline: [29590047](https://pubmed.ncbi.nlm.nih.gov/29590047/)]

48. He B, Hoang TK, Wang T, Ferris M, Taylor CM, Tian X, et al. Resetting microbiota by *Lactobacillus reuteri* inhibits T reg deficiency-induced autoimmunity via adenosine A2A receptors. *J Exp Med* 2017 Jan;214(1):107-123 [FREE Full text] [doi: [10.1084/jem.20160961](https://doi.org/10.1084/jem.20160961)] [Medline: [27994068](https://pubmed.ncbi.nlm.nih.gov/27994068/)]
49. Mu Q, Kirby J, Reilly CM, Luo XM. Leaky Gut As a Danger Signal for Autoimmune Diseases. *Front Immunol* 2017;8:598 [FREE Full text] [doi: [10.3389/fimmu.2017.00598](https://doi.org/10.3389/fimmu.2017.00598)] [Medline: [28588585](https://pubmed.ncbi.nlm.nih.gov/28588585/)]
50. Hevia A, Milani C, López P, Cuervo A, Arboleya S, Duranti S, et al. Intestinal dysbiosis associated with systemic lupus erythematosus. *mBio* 2014 Sep 30;5(5):e01548-e01514. [doi: [10.1128/mBio.01548-14](https://doi.org/10.1128/mBio.01548-14)] [Medline: [25271284](https://pubmed.ncbi.nlm.nih.gov/25271284/)]
51. NIH Integrative Human Microbiome Project. URL: <https://www.hmpdacc.org/ihmp/overview/> [accessed 2008-05-20]
52. Jolly M, Katz P. Systemic lupus erythematosus: The promise of PROMIS - is it ready for prime time in SLE? *Nat Rev Rheumatol* 2017 Aug;13(8):453-454. [doi: [10.1038/nrrheum.2017.100](https://doi.org/10.1038/nrrheum.2017.100)] [Medline: [28660907](https://pubmed.ncbi.nlm.nih.gov/28660907/)]
53. Mahieu M, Yount S, Ramsey-Goldman R. Patient-Reported Outcomes in Systemic Lupus Erythematosus. *Rheum Dis Clin North Am* 2016 May;42(2):253-263 [FREE Full text] [doi: [10.1016/j.rdc.2016.01.001](https://doi.org/10.1016/j.rdc.2016.01.001)] [Medline: [27133488](https://pubmed.ncbi.nlm.nih.gov/27133488/)]
54. More People Search for Health Online. *nbcnews.com*. URL: [http://www.nbcnews.com/id/3077086/t/more-people-search-health-online/#.Xzwi\\_S2ZOrJ](http://www.nbcnews.com/id/3077086/t/more-people-search-health-online/#.Xzwi_S2ZOrJ) [accessed 2020-09-28]
55. Mobile Fact Sheet. URL: <https://www.pewresearch.org/internet/fact-sheet/mobile/> [accessed 2020-09-20]

## Abbreviations

**BPI-SF:** Brief Pain Index-Short Form

**CNS:** central nervous system

**FACIT-F:** Functional Assessment of Chronic Illness Therapy-Fatigue

**HIPAA:** Health Insurance Portability and Accountability Act

**HRQoL:** health-related quality of life

**ITT:** intention to treat

**LupusQoL:** Lupus Quality of Life

**PP:** per protocol

**PROMs:** patient-reported outcome measures

**SLE:** systemic lupus erythematosus

**SLEDAI:** Systemic Lupus Erythematosus Disease Activity Index

**SRI:** SLE Responder Index

*Edited by G Eysenbach; submitted 26.08.20; peer-reviewed by W Gordon, A Azzam; comments to author 17.09.20; revised version received 28.09.20; accepted 30.09.20; published 20.10.20.*

*Please cite as:*

Khan F, Granville N, Malkani R, Chathampally Y

Health-Related Quality of Life Improvements in Systemic Lupus Erythematosus Derived from a Digital Therapeutic Plus Tele-Health Coaching Intervention: Randomized Controlled Pilot Trial

*J Med Internet Res* 2020;22(10):e23868

URL: <https://www.jmir.org/2020/10/e23868>

doi: [10.2196/23868](https://doi.org/10.2196/23868)

PMID: [33079070](https://pubmed.ncbi.nlm.nih.gov/33079070/)

©Faiz Khan, Nora Granville, Raja Malkani, Yash Chathampally. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 20.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Internet-Based Multimodal Pain Program With Telephone Support for Adults With Chronic Temporomandibular Disorder Pain: Randomized Controlled Pilot Trial

Julia Lam<sup>1,2,3</sup>, DDS; Peter Svensson<sup>1,4,5</sup>, DDS, PhD; Per Alstergren<sup>1,3,6</sup>, DDS, PhD

<sup>1</sup>Department of Orofacial Pain and Jaw Function, Faculty of Odontology, Malmö, Sweden

<sup>2</sup>Folk tandvården Skåne AB, Hässleholm and Lund, Sweden

<sup>3</sup>Scandinavian Center for Orofacial Neurosciences (SCON), Malmö, Sweden

<sup>4</sup>Scandinavian Center for Orofacial Neurosciences (SCON), Aarhus, Denmark

<sup>5</sup>Department of Orofacial Pain and Jaw Function, Faculty of Health, Aarhus University, Aarhus, Denmark

<sup>6</sup>Specialized Pain Rehabilitation, Skåne University Hospital, Lund, Sweden

**Corresponding Author:**

Julia Lam, DDS

Department of Orofacial Pain and Jaw Function

Faculty of Odontology

Malmö universitet

Malmö, 205 06

Sweden

Phone: 46 767690002

Email: [Julia.lam@mau.se](mailto:Julia.lam@mau.se)

## Abstract

**Background:** Chronic pain from temporomandibular disorders remains an undertreated condition with debate regarding the most effective treatment modalities.

**Objective:** The aim of the study was to investigate the treatment effect of an internet-based multimodal pain program on chronic temporomandibular disorder pain and evaluate the feasibility of a larger randomized controlled trial.

**Methods:** An unblinded randomized controlled pilot trial was conducted with 43 participants (34 females, 9 males; median age 27, IQR 23-37 years) with chronic temporomandibular pain. Participants were recruited within the Public Dental Health Service and randomized to intervention (n=20) or active control (n=23). The intervention comprised a dentist-assisted internet-based multimodal pain program with 7 modules based on cognitive behavior therapy and self-management principles. The control group received conventional occlusal splint therapy. Primary outcomes included characteristic pain intensity, pain-related disability, and jaw functional limitation. Secondary outcomes were depression, anxiety, catastrophizing, and stress. Outcomes were self-assessed through questionnaires sent by mail at 3 and 6 months after treatment start. Feasibility evaluation included testing the study protocol and estimation of recruitment and attrition rates in the current research setting.

**Results:** Only 49% of participants (21/43) provided data at the 6-month follow-up (internet-based multimodal pain program: n=7; control: n=14). Of the 20 participants randomized to the internet-based multimodal pain program, 14 started treatment and 8 completed all 7 modules of the program. Between-group analysis showed no significant difference for any outcome measure at 3- or 6-month follow-up—characteristic pain intensity (3 months:  $P=.58$ ; 6 months:  $P=.41$ ), pain-related disability (3 months:  $P=.51$ ; 6 months:  $P=.12$ ), jaw functional limitation (3 months:  $P=.45$ ; 6 months:  $P=.90$ ), degree of depression (3 months:  $P=.64$ ; 6 months:  $P=.65$ ), anxiety (3 months:  $P=.93$ ; 6 months:  $P=.31$ ), stress (3 months:  $P=.66$ ; 6 months:  $P=.74$ ), or catastrophizing (3 months:  $P=.86$ ; 6 months:  $P=.85$ ). Within-group analysis in the internet-based multimodal pain program group showed a significant reduction in jaw functional limitation score at the 6-month follow-up compared to baseline (Friedman:  $\chi^2=10.2$ ,  $P=.04$ ; Wilcoxon:  $z=-2.3$ ,  $P=.02$ ). In the occlusal splint group, jaw function limitation was also reduced at the 6-month follow-up (Friedman:  $\chi^2=20.0$ ,  $P=.045$ ; Wilcoxon:  $z=-2.3$ ,  $P=.02$ ), and there was a reduction in characteristic pain intensity at the 3- and 6-month follow-up (Friedman:  $\chi^2=25.1$ ,  $P=.01$ ; Wilcoxon 3 months:  $z=-3.0$ ,  $P=.003$ ; Wilcoxon 6 months:  $z=-3.3$ ,  $P=.001$ ).

**Conclusions:** This study was not able to demonstrate a difference in treatment outcome between an internet-based multimodal pain program and occlusal splint therapy in patients with chronic temporomandibular pain. However, the findings suggested that the internet-based multimodal pain program improves jaw function. The results also confirmed the treatment effect of occlusal



splint therapy for chronic temporomandibular pain. Furthermore, because of the high attrition rate, this pilot study showed that a randomized controlled trial with this design is not feasible.

**Trial Registration:** ClinicalTrials.gov NCT04363762; <https://clinicaltrials.gov/show/NCT04363762>

(*J Med Internet Res* 2020;22(10):e22326) doi:[10.2196/22326](https://doi.org/10.2196/22326)

## KEYWORDS

chronic pain; cognitive behavior therapy; combined modality therapy; facial pain; feasibility studies; health services research; internet-based intervention; occlusal splints; pilot projects; temporomandibular disorders

## Introduction

Temporomandibular disorders (TMD) are underdiagnosed and undertreated conditions in the general population [1]. TMD involves conditions affecting the temporomandibular joint and the masticatory muscles and has a prevalence of 10%-15% in the adult population [2]. Typical TMD complaints consist of pain, fatigue, and stiffness in the jaw muscles; limitations in jaw movements; and clicking or grating noises from the jaw joints [3]. Chronic TMD pain negatively affects quality of life, and in Sweden, even though the estimated treatment need is 5%-15%, only 0.5%-1.5% receive treatment in general dentistry [4,5]. The Swedish national guidelines [1] recommend a multimodal treatment approach, and common treatment options include occlusal appliances, physiotherapy, pharmacologic treatment, and behavior treatment. There are still debates regarding the most efficient and cost-effective treatment option that can easily be distributed and applied.

Internet-based intervention is an appealing modality for multimodal TMD treatment as it may not only assist dentists in providing treatment at reduced costs but may also potentially reduce traveling time, treatment costs, and waiting lists for patients. Generally, in the field of pain, there are several apps for various chronic pain conditions (eg, headache, fibromyalgia, and back pain) [6-9]. Internet-based treatment reduces pain intensity and pain interference [9,10] and internet-based cognitive behavior therapy seems to be at least as effective as a face-to-face intervention for chronic pain [11]. However, findings regarding depression and anxiety are inconsistent [12], and there are no internet-based cognitive behavior therapy studies on chronic TMD pain.

This study set out to collect data on the effectiveness of a guided internet-based multimodal pain program for adults with chronic TMD pain compared to that of conventional occlusal splint therapy. As the study progressed, it also became a measure to evaluate the feasibility of a larger randomized controlled trial.

## Methods

### Study Design

An unblinded, parallel-arm, pilot randomized controlled trial with equal allocation was conducted. The participants were randomized to the internet-based multimodal pain program or to an active control. The study was approved by the Ethics Review Board in Lund, Sweden (No. 2016/6). The examination and treatments were free of charge for the participants, and no other financial compensation was given. See [Multimedia](#)

[Appendix 1](#) for the informed consent documentation in Swedish. This study was retrospectively registered at ClinicalTrials.gov (NCT04363762).

Furthermore, this study was part of a large research project with a primary objective of studying changes in the brain in patients with chronic TMD pain after treatment. Therefore, all participants underwent magnetic resonance imaging (MRI) of the brain. The neuroimaging data will be presented elsewhere; however, the feasibility aspect of this additional examination will be addressed here.

### Participants

Participant inclusion criteria were (1) age between 18 and 75 years; (2) at least one TMD pain diagnosis such as myalgia, myofascial pain with referral, headache attributed to TMD, or arthralgia according to the Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) [13]; (3) chronic ( $\geq 3$  months) TMD pain, experienced once a week or more often, with an intensity of  $\geq 3$  (on a scale from 0 to 10); (4) access to a computer with an internet connection and a mobile phone; (5) sufficient computer literacy; and (6) Swedish language fluency. Exclusion criteria were (1) chronic inflammatory systemic diseases; (2) all psychiatric disorders except depression and anxiety due to high comorbidity; (3) occlusal splint therapy in the past 12 months; (4) ongoing extensive dental treatment; and (5) conditions precluding MRI examination.

### Procedure

Participants were recruited from a general dental care clinic (Fäladstorget, Lund) within the National Dental Care in Skåne, Sweden. To identify potential participants, all consecutive adult patients visiting for a regular dental check-up were screened with the 2Q/TMD screening tool [14,15]. If a patient was deemed eligible for the inclusion examination, oral and written information about the study were given, and a new appointment was booked for an inclusion examination. The patient was asked to provide written informed consent, if they were deemed eligible after the inclusion examination. Permuted block randomization with a fixed block size of 10 was used. At the moment of assignment, the local research coordinator blindly picked a piece of paper from an envelope with allocated treatment. Before the study start, 6 sets of opaque envelopes with 10 allocation notes each, 5 with "internet-based multimodal pain program" and 5 with "occlusal splint," were prepared. All 10 notes in one envelope had to be used before the next envelope was opened. If allocated to internet-based multimodal pain program, working material was sent to the participant by mail, and the treatment was started, with assistance provided by phone. If allocated to treatment with occlusal splint, a new



appointment was booked for delivering the splint. Follow-up was performed 3 and 6 months after treatment start with questionnaires that were mailed to the participants. Participants also underwent an MRI of the brain at baseline, after randomization but before treatment start, and 6 months after treatment start at Skåne University Hospital, Malmö, Sweden.

Participant recruitment started in April 2016 and ended in December 2018. Screening was paused during annual summer holiday periods (June-August 2016, 2017, 2018) and during February-April 2017 due to technical issues with the MRI scanner.

### Inclusion Examination

All participants underwent a standardized examination according to DC/TMD to control for eligibility and collect baseline data. The DC/TMD includes a clinical examination and questionnaires described below [13]. Demographic data and pretreatment expectancies on allocated treatment were also assessed.

### Clinical Examination

The clinical DC/TMD examination comprises a standardized assessment of pain and headache locations, jaw opening capacity, pain on mandibular movement, pain on palpation, and temporomandibular joint noises [13]. The clinical examinations were carried out by one of 2 trained dentists [16].

The DC/TMD Symptom Questionnaire was used to assess pain symptoms involving the jaw, jaw noise and locking, and headache. Data from this questionnaire were combined with findings from the clinical examination for diagnosis according to DC/TMD [13].

### Questionnaires

#### Pain

The Graded Chronic Pain Scale [17] was used to assess the subdomains of characteristic pain intensity (pain intensity for reported worst, current, and average pain) and pain-related disability (how much facial pain changed the patient's ability to participate in daily activities, social activities, and work). The questionnaire included the assessment of pain intensity and pain-related disability scored on a 0-10 numeric rating scale, modified from [17].

A full-body pain drawing was used to assess pain locations and distribution [18]. Pain distribution was categorized as local pain in the face area; regional pain in the neck area in addition to local pain; or widespread pain at any other site of the body in addition to local or regional pain.

#### Physical Functioning

The Jaw Functional Limitation Scale-8 was used to assess jaw function in the masticatory system [19]. The Oral Behaviors Checklist was used to assess the self-reported frequency of oral parafunctional behaviors during awake time [20].

### Emotional Functioning

The Patient Health Questionnaire-9 was used to assess depression, and the Patient Health Questionnaire-15 was used to assess nonspecific physical symptoms [21,22]. Anxiety was assessed with the Generalized Anxiety Disorders-7 [23], stress was assessed with the Perceived Stress Scale-10, and catastrophizing was assessed with the Pain Catastrophizing Scale [24,25].

### Patient Outcome Expectancy

A Swedish translation of the Stanford Expectations of Treatment Scale was used to assess positive and negative pretreatment expectancies of treatments [26].

### Treatment Modalities

#### Intervention

The internet-based multimodal pain program is based on cognitive behavior therapy and self-management principles that help patients to cope with chronic TMD pain. It was developed for use in general dentistry by orofacial pain specialists at the Department of Orofacial Pain and Jaw Function at Malmö University, Sweden, in collaboration with Psykologpartners AB (a Swedish provider of internet-based psychological treatment).

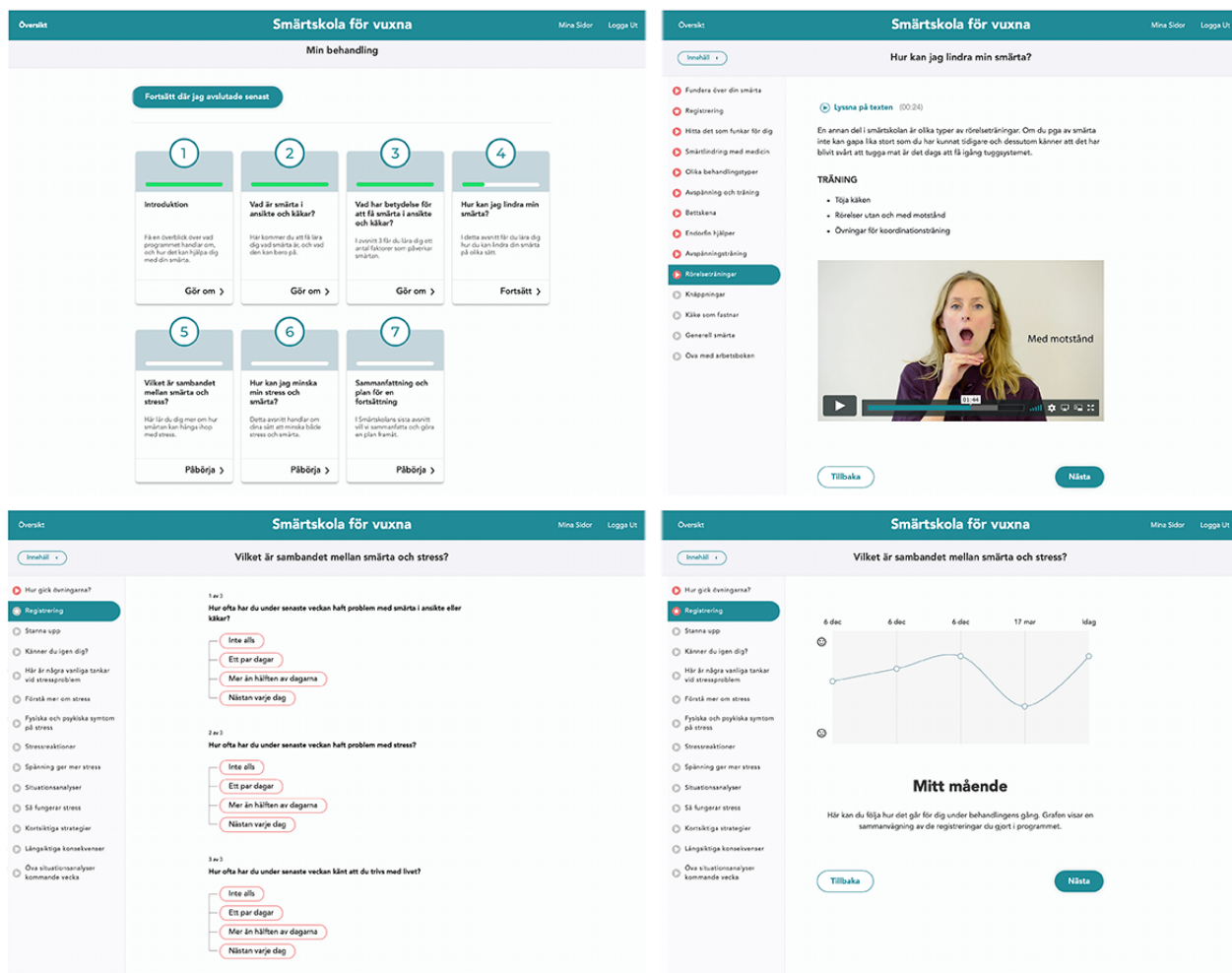
The internet-based multimodal pain program adapts face-to-face therapy to a software platform program. Access required a 2-factor authentication system. A demo of the program is available ([Multimedia Appendix 2](#)). The internet-based multimodal pain program consisted of 7 treatment modules ([Table 1](#), [Figure 1](#)) and offline activities included homework assignments in a paperback workbook. The intended treatment duration was 7 weeks: 1 module/week; 40 minutes/module online plus time for homework assignments.

The internet-based multimodal pain program was designed to be used without guidance. However, to support study adherence, telephone, email, and asynchronous chat support was provided by a dentist (JL) who had received 2-day training on internet-based cognitive behavior therapy arranged by Psykologpartners AB. At the start-up phone call, participants were guided through the functions in the program and were informed about time requirements for the treatment. Follow-up phone calls after each finished module involved individualized support and feedback. If participants were unreachable at scheduled follow-up calls, a message through the chat function in the program was sent with a prompt to get in touch. If still not heard from in the subsequent 2 weeks, the dentist tried to contact the participant by phone. If still not reached, access to the internet-based multimodal pain program was removed, and the participant was withdrawn from the study. A letter was sent to the participant with information about the withdrawal and a prompt to seek treatment from a dentist.

**Table 1.** An overview of the internet-based multimodal pain program for chronic temporomandibular disorders pain.

Module	Theme	Content	Cognitive behavior therapy component
1	Introduction	Information: introduction to the treatment program and cognitive behavior therapy, goal setting Assignment: a reflection on previous strategies of handling temporomandibular disorder pain, assessment of core values	Values and goals
2	What is face and jaw pain?	Information: etiology and epidemiology of temporomandibular disorder pain, anatomy of the masticatory system Assignment: pain drawing, identify jaw functions limited by pain	Psychoeducation and assessment
3	What affects my pain in the face and jaw?	Information: modulating and maintaining factors of temporomandibular disorder pain, pain physiology, acute and chronic pain, the link between quality of life and temporomandibular disorder pain Assignment: 7-day pain diary	Psychoeducation and assessment
4	How can I relieve my pain?	Information: treatment alternatives including analgesics, relaxation, jaw exercises, different occlusal splints, acupuncture, and massage Assignment: identify pain modulators from pain diary, practice relaxation, and jaw exercises	Applied relaxation and skills training
5	The relation between stress and pain	Information: stress responses, the link between stress and pain Assignment: a questionnaire to assess the degree of stress, situational analyses, list desired ways to enhance recuperation, continue relaxation and jaw exercises	Functional analysis, problem-solving, relaxation, skills training
6	How can I reduce my stress and pain levels?	Information: stress and pain management; breathing exercises, diet, sleep, training, time management, setting boundaries and acceptance, continued relaxation and jaw exercises Assignment: mapping sleep, training and dietary habits in a weekly schedule and do an activity planning aiming to make desired changes regarding these, continued relaxation and jaw exercises	Committed action, relaxation, skills training
7	Summary and a plan for future action	Information: Setbacks and maintenance planning, a summary of the program Assignment: Develop a maintenance plan	Summary and maintenance plan

**Figure 1.** Screenshots from the guided internet-based multimodal pain program (top left: start page; top right: a page with a jaw exercises instructional video; lower left: a page with the recurrent rating of pain frequency, stress level, and life satisfaction to keep track of progression; lower right: an overview of treatment progression).



## Control Condition

Participants in the active control group received a hard Michigan-type stabilization splint placed in the upper jaw [27]. The occlusal splint was chosen as the control treatment because it is a conventional and reversible treatment for TMD pain, and it has known moderate efficacy [5]. The splints were made of acrylic, covered all the maxillary teeth, and had a smooth, flat surface. It was fitted to allow all supporting teeth to contact simultaneously on jaw closure, and canine guidance and disocclusion of the posterior teeth were provided in the lateral and protrusive excursion. Participants were instructed to wear it only at night. A clinical check-up of the participants in the occlusal splint group was performed within 2-6 weeks after treatment start. Scheduled time at the dentist was in total 60 minutes for this therapy.

## Follow-up

Follow-up was conducted with questionnaires comprising a short version of the DC/TMD instruments. The questionnaires were sent to participants by mail at 3 and 6 months after treatment start. If the questionnaire was not returned within a month, a reminder was sent by email or by phone.

## Outcome Measures

### Preliminary Evaluation of Participant Responses to Intervention

Measures specific to TMD included in the DC/TMD were used as outcome measures, guided by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations on core domains [28,29]. Primary outcomes included characteristic pain intensity, pain-related disability, and physical functioning (jaw functional limitation score). Secondary outcomes were depression and anxiety as measures of emotional functioning. The IMMPACT core domains also include ratings of global improvement and satisfaction with treatment, but this was not assessed in this study. In addition to the IMMPACT core domains, catastrophizing and stress were evaluated as secondary outcomes.

### Feasibility

Feasibility evaluation of the study included testing the study protocol and estimating recruitment and attrition rates in the current research setting. With preliminary results, post hoc power and sample size calculation were performed. Recruitment and attrition rates along with a new estimated sample size

allowed a decision on the feasibility of conducting a future larger randomized controlled trial to be made. Usage metrics of internet-based multimodal pain program were not assessed.

### **Adverse Effects**

Adverse events for the internet-based multimodal pain program treatment were collected through the follow-up phone calls. Also, all participants randomized to internet-based multimodal pain program were given written information at the 3-month follow-up about the opportunity to withdraw from the study and receive rescue treatment if symptoms worsened or if they were not satisfied with the treatment outcome. In a similar manner, participants randomized to occlusal splint therapy were free to withdraw from the study at any time, without penalty or loss of benefits, to seek additional treatment.

### **Statistical Analysis**

Nonparametric statistics were used due to the characteristics of the pain-related variables. For descriptive statistics, median, 25th percentile, and 75th percentile, as well as the number of observations or proportions, are reported. To explore the potential effects of the internet-based multimodal pain program in this early-stage trial, per protocol between-group comparisons of change in characteristic pain intensity level, pain-related disability, jaw functional limitation score, degree of depression, anxiety, catastrophizing, and stress at the 3- and 6-month follow-up were performed with the Wilcoxon rank-sum test. Friedman analysis of variance was used for within-group comparisons. Significant results from the Friedman analysis were followed by posthoc Wilcoxon matched-pair signed-rank testing. In addition, Spearman rank-order was used to test possible association between positive and negative treatment expectancy and changes in outcome measures at the 6-month follow-up within respective treatment groups. Dropout analysis

was performed to test the difference between completers and dropouts within allocated treatment groups. Furthermore, a comparison of baseline characteristics of participants between study groups was made. Chi-square test was used for categorical variables and the Wilcoxon rank-sum test was used for ordinal variables.

Estimation of required sample size was based on a previous neuroimaging study to detect statistically significant results in functional MRI [30]. According to this, 20 participants in each group should be recruited; however, the goal was to recruit 30 participants in each group to compensate for dropouts. It was also considered to be sufficient to provide useful information about the feasibility of a future larger randomized controlled trial. A posthoc power analysis with characteristic pain intensity as the outcome measure was performed. The sample size required for a parametric test on the between-group comparison was computed, and 15% was added, as nonparametric testing could be used in a future randomized controlled trial [31].

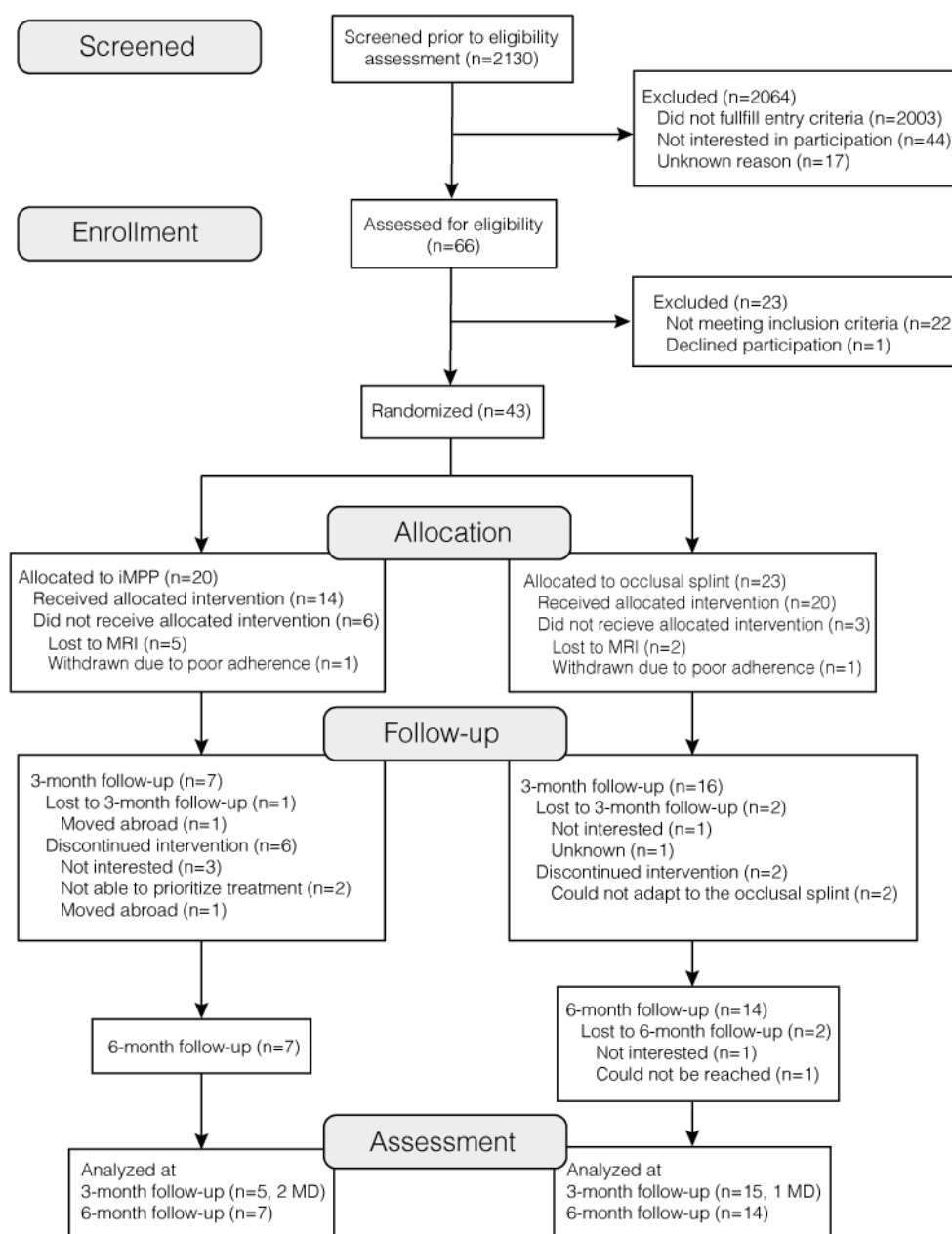
All data management and statistical analyses were performed unblinded using StataSE (version 15.1; StataCorp LLC). A probability level of  $P < .05$  was considered significant.

## **Results**

### **Participants**

Participant flow through each stage of the trial is presented in Figure 2. The baseline characteristics of the sample are described in Table 2. Comparisons of baseline demographic, clinical, and psychosocial characteristics showed no statistical difference between treatment groups except that the occlusal splint group had a significantly greater number of jaw muscles with pain on palpation ( $P = .049$ ; Table 2).

**Figure 2.** Flow diagram of the unblinded, parallel-arm, pilot randomized trial. iMPP: internet-based multimodal pain program; MRI: magnetic resonance imaging.





**Table 2.** Demographic, clinical, and psychosocial characteristics at baseline for patients with chronic temporomandibular disorder pain randomized to treatment with a guided internet-based multimodal pain program (iMPP, n=20) or occlusal splint (control, n=23).

Variable	iMPP <sup>a</sup>	Controls	P value
<b>Demographics</b>			
Age (years), median (IQR)	27 (23-37)	27 (23-38)	.71
<b>Sex, n (%)</b>			.54
Male	5 (25)	4 (17)	
Female	15 (75)	19 (83)	
<b>Birth country, n (%)</b>			.76
Sweden or other Nordic country	18 (90)	20 (87)	
Other	2 (10)	3 (13)	
<b>Civil status, n (%)</b>			.09
Single	8 (40)	4 (17)	
Married or de facto	6 (30)	14 (61)	
Divorced	2 (10)	0 (0)	
Other	4 (20)	5 (22)	
<b>Education level, n (%)</b>			.43
University	9 (45)	12 (52)	
Professional training	1 (5)	3 (13)	
High school	10 (50)	7 (30)	
Elementary school	0 (0)	1 (4)	
<b>Employment status, n (%)</b>			.29
Employed	17 (85)	19 (83)	
Unemployed	1 (5)	4 (17)	
Retired	0 (0)	0 (0)	
Registered disabled	1 (5)	0 (0)	
Missing data	1 (5)	0 (0)	
<b>Clinical characteristics</b>			
Characteristic pain intensity (0-10), median (IQR)	4.7 (3.7-5.3)	4 (3.7-5.7)	.61
Duration of temporomandibular disorder pain (months), median (IQR)	30 (12-120)	102 (12-300)	.36
Pain free mouth opening (mm), median (IQR)	41 (33-51)	36 (27-51)	.46
Maximum Unassisted mouth opening (mm), median (IQR)	52 (48-57)	49 (41-59)	.29
Number of muscles with pain on palpation (0-4), median (IQR)	4 (2-4)	4 (4-4)	.049
<b>Pain distribution, n (%)</b>			.74
Face	11 (55)	11 (48)	
Face and neck	9 (45)	12 (52)	
Other parts of the body	11 (55)	11 (48)	
<b>Self-reported comorbidities, n (%)</b>			.40
Gastrointestinal disorders	5 (25)	1 (4)	
Neurological disorder	1 (5)	0 (0)	
Psychiatric disorder	6 (30)	3 (13)	
Tinnitus	6 (30)	7 (30)	
Other pain states	6 (30)	7 (30)	
<b>Psychosocial characteristics, median (IQR)</b>			

Variable	iMPP <sup>a</sup>	Controls	<i>P</i> value
Pain-related disability (GCPS DS <sup>b</sup> , 0-10)	0.8 (0.0-2.0)	0.3 (0.0-2.3)	.98
Awake parafunctional behaviors (OBC <sup>c</sup> , 0-76)	26 (21-31)	24 (19-30)	.71
Jaw function limitation (JFL <sup>d</sup> , 0-10)	0.5 (0-1.4)	0.8 (0.1-1.8)	.55
Depression (PHQ-9 <sup>e</sup> , 0-27)	7 (3-11)	5 (2-10)	.52
Anxiety (GAD-7 <sup>f</sup> , 0-21)	5 (2-8)	5 (2-9)	.52
Unspecific physical symptoms (PHQ-15 <sup>g</sup> , 0-30)	8 (6-12)	8 (5-10)	.86
Stress (PSS-10 <sup>h</sup> , 0-40)	13 (10-21)	12 (9-20)	.68
Catastrophizing (PCS <sup>i</sup> , 0-52)	16 (7-25)	12 (5-18)	.20
<b>Treatment expectation (SETS<sup>j</sup>, 1-7), n (%)</b>			
Positive expectancy	5.3 (4.7-5.7)	5.0 (4.3-5.7)	.56
Negative expectancy	2.0 (2.0-3.3)	2.3 (2.0-3.0)	.81

<sup>a</sup>iMPP: Internet-based multimodal pain program.

<sup>b</sup>GCPS DS: Graded Chronic Pain Scale Disability Score.

<sup>c</sup>OBC: Oral Behaviors Checklist.

<sup>d</sup>JFL: Jaw Functional Limitation Scale-8.

<sup>e</sup>PHQ-9: Patient Health Questionnaire-9.

<sup>f</sup>GAD-7: Generalized Anxiety Disorders Assessment -7.

<sup>g</sup>PHQ-15: Patient Health Questionnaire-15.

<sup>h</sup>PSS-10: Perceived Stress Scale-10.

<sup>i</sup>PCS: Pain Catastrophizing Scale.

<sup>j</sup>SETS: Stanford Expectations of Treatment Scale.

## Preliminary Evaluation of Participant Responses to Intervention

The between-group analysis of change in outcome measures (Table 3) showed no significant difference between the treatment groups at any follow-up regarding characteristic pain intensity (3 months:  $P=.58$ ; 6 months:  $P=.41$ ), pain-related disability (3 months:  $P=.51$ ; 6 months:  $P=.12$ ), jaw functional limitation (3 months:  $P=.45$ ; 6 months:  $P=.90$ ), degree of depression (3 months:  $P=.64$ ; 6 months:  $P=.65$ ), anxiety (3 months:  $P=.93$ ; 6 months:  $P=.31$ ), stress (3 months:  $P=.66$ ; 6 months:  $P=.74$ ), or catastrophizing (3 months:  $P=.86$ ; 6 months:  $P=.85$ ).

Within-group analysis (Table 3) showed a significant reduction in jaw function limitation at the 6-month follow-up in the internet-based multimodal pain program group (Friedman:  $\chi^2_2=10.2$ ,  $P=.04$ ; Wilcoxon:  $z=-2.3$ ,  $P=.02$ ). In the occlusal splint group, characteristic pain intensity was significantly reduced at both 3- and 6-month follow-up (Friedman:  $\chi^2_2=25.1$ ,  $P=.01$ ; Wilcoxon 3 month:  $z=-3.0$ ,  $P=.003$ ; Wilcoxon 6 month:  $z=-3.3$ ,  $P=.001$ ) and jaw function limitation was reduced at the 6-month follow-up (Friedman:  $\chi^2_2=20.0$ ,  $P=.045$ ; Wilcoxon:  $z=-2.33$ ,  $P=.02$ ).

**Table 3.** Difference in outcome measure scores at 3- and 6-month follow-up, as well as results from within- and between-group comparisons.

Outcomes <sup>a</sup>			Between-group differences				Within-group differences			
			Wilcoxon rank-sum				Friedman		Wilcoxon signed-rank	
	3 months <sup>b</sup>	6 months <sup>c</sup>	3 months		6 months				3 months	6 months
			z score	P value	z score	P value	$\chi^2$	P value	z score	P value
<b>Primary outcomes</b>										
<b><math>\Delta</math>GCPS CPI<sup>d</sup></b>			0.56	.58	0.82	.41				
iMPP <sup>e</sup>	–1.3 (–2.0, –0.3)	–1.6 (–4.0, –0.3)					4.8	.31	N/A <sup>f</sup>	N/A
Occlusal splint	–2.0 (–3.3, –1.0)	–2.5 (–3.3, –1.3)					25.1	.01	–3.01	.003
<b><math>\Delta</math>GCPS DS<sup>g</sup></b>			0.67	.51	1.55	.12				
iMPP	0.0 (–1.0, 0.0)	0.0 (–0.3, 0.3)					5.5	.24	N/A	N/A
Occlusal splint	–0.3 (–1.3, 0.0)	–0.7 (–2.0, 0.0)					20.2	.06	N/A	N/A
<b><math>\Delta</math>JFL<sup>h</sup></b>			–0.75	.45	0.12	.90				
iMPP	–0.5 (–0.8, –0.3)	–0.5 (–0.9, –0.1)					10.2	.04	–1.91	.06
Occlusal splint	0.0 (–0.9, 0.0)	–0.8 (–1.1, 0.0)					20.0	.045	–1.69	.09
<b>Secondary outcomes</b>										
<b><math>\Delta</math>PHQ-9<sup>i</sup></b>			0.47	.64	–0.45	.65				
iMPP	1.0 (–1.0, 2.0)	0.0 (–6.0, 2.0)					9.9	.04	0.54	.59
Occlusal splint	0.0 (–3.0, 1.0)	–1.0 (–2.0, 2.0)					23.5	.02	–0.60	.55
<b><math>\Delta</math>GAD-7<sup>j</sup></b>			0.09	.93	–1.02	.31				
iMPP	0.0 (0.0, 0.0)	0.0 (–6.0, 1.0)					9.9	.04	–0.16	.88
Occlusal splint	–1.0 (–2.0, 3.0)	0.0 (–2.0, 1.0)					27.5	.007	0.11	.91
<b><math>\Delta</math>PSS-10<sup>k</sup></b>			0.44	.66	–0.34	.74				
iMPP	4.0 (–2.0, 6.0)	0.0 (–8.0, 2.0)					10.9	.03	0.41	.69
Occlusal splint	0.0 (–4.0, 5.0)	–0.5 (–5.0, 3.0)					27.9	.006	0.26	.80
<b><math>\Delta</math>PCS<sup>l</sup></b>			0.18	.86	0.19	.85				
iMPP	–1.0 (–6.0, 0.0)	–4.0 (–6.0, 1.0)					9.5	.049	–0.54	.59
Occlusal splint	–2.0 (–7.0, 0.0)	–3.0 (–11.0, 0.0)					20.8	.052	N/A	N/A

<sup>a</sup>Difference in outcome measure between follow-ups and baseline.<sup>b</sup>At 3-month follow-up—internet-based multimodal pain program group: n=5 and occlusal splint group: n=15.<sup>c</sup>At 6-month follow up—internet-based multimodal pain program group: n=7 and occlusal splint group: n=14.<sup>d</sup>GCPS CPI: Graded Chronic Pain Scale Characteristic Pain Intensity.<sup>e</sup>iMPP: internet-based multimodal pain program.<sup>f</sup>N/A: not applicable.<sup>g</sup>GCPS DS: Graded Chronic Pain Scale Disability Score.<sup>h</sup>JFL: Jaw Functional Limitation Scale-8.<sup>i</sup>PHQ-9: Patient Health Questionnaire-9 (for depression).<sup>j</sup>GAD-7: Generalized Anxiety Disorders-7.<sup>k</sup>PSS-10: Perceived Stress Scale-10.<sup>l</sup>PCS: Pain Catastrophizing Scale.

## Pretreatment Expectations

Comparison of pretreatment expectations on allocated treatment did not show any significant difference in positive ( $P=.56$ ) or

negative expectancy ( $P=.81$ ) between groups (Table 2). In the internet-based multimodal pain program group there was a significant correlation between positive expectancy at baseline and change in pain-related disability ( $\rho=0.78$ ,  $P=.04$ ).

There was also a significant negative correlation between positive expectancy and change in the degree of depression ( $\rho=-0.78$ ,  $P=.04$ ). In the occlusal splint group, positive expectancy was positively correlated to change in

catastrophizing ( $\rho=0.66$ ,  $P=.01$ ). No significant correlation was found between negative expectancy and change in treatment outcomes (Table 4).

**Table 4.** Spearman correlations for positive and negative treatment expectancy between baseline and difference in outcome measurements 6 months after treatment start.

Variable	Internet-based multimodal pain program (n=7)				Occlusal splint (n=14)			
	Positive expectancy		Negative expectancy		Positive expectancy		Negative expectancy	
	$\rho$	<i>P</i> value	$\rho$	<i>P</i> value	$\rho$	<i>P</i> value	$\rho$	<i>P</i> value
Characteristic pain intensity	0.26	.57	0.19	.69	0.25	.39	0.24	.40
Pain-related disability (GCPS DS <sup>a</sup> , 0-10)	0.78	.04	0.65	.11	0.49	.08	-0.21	.48
Jaw function limitation (JFL <sup>b</sup> , 0-10)	-0.36	.43	-0.11	.81	0.18	.56	0.14	.65
Depression (PHQ-9 <sup>c</sup> , 0-27)	-0.78	.04	-0.54	.21	-0.15	.61	-0.12	.69
Anxiety (GAD-7 <sup>d</sup> , 0-21)	-0.42	.34	-0.19	.68	-0.48	.08	-0.16	.58
Stress (PSS-10 <sup>e</sup> , 0-40)	-0.24	.61	0.04	.93	0.17	.55	0.24	.41
Catastrophizing (PCS <sup>f</sup> , 0-52)	-0.40	.38	0.32	.49	0.66	.01	0.25	.39
Positive expectancy (SETS <sup>g</sup> , 1-7)	N/A <sup>h</sup>	N/A	0.20	.41	N/A	N/A	0.07	.77
Negative expectancy (SETS, 1-7)	0.20	.41	N/A	N/A	0.07	.77	N/A	N/A

<sup>a</sup>GCPS DS: Graded Chronic Pain Scale Disability Score.

<sup>b</sup>JFL: Jaw Functional Limitation Scale-8.

<sup>c</sup>PHQ-9: Patient Health Questionnaire-9.

<sup>d</sup>GAD-7: Generalized Anxiety Disorders -7.

<sup>e</sup>PSS-10: Perceived Stress Scale-10.

<sup>f</sup>PCS: Pain Catastrophizing Scale.

<sup>g</sup>SETS: Stanford Expectations of Treatment Scale.

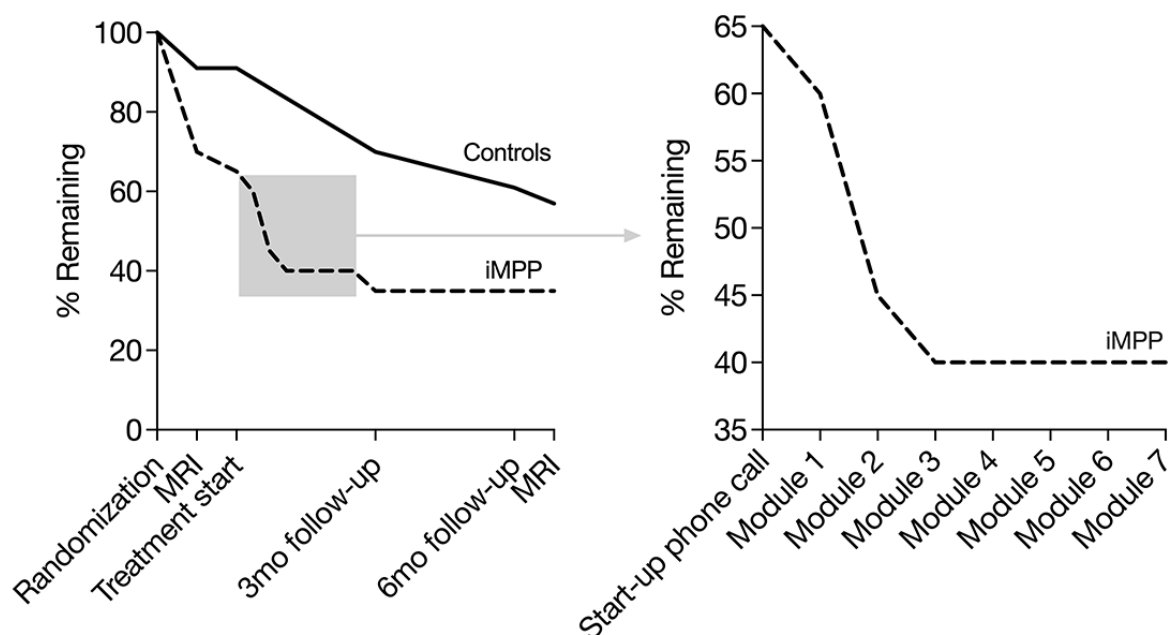
<sup>h</sup>N/A: not applicable.

## Feasibility

An average of 101 patients were screened per month, and 5 patients per month went through the inclusion examination. From this, 2 participants per month were enrolled, and 43 participants were recruited during the a priori determined recruitment period—fewer than had initially been planned. An attempt was made to extend the recruitment period, but it was considered futile under the current circumstances. The attrition rate showed an unbalanced dropout between groups (Figure 3). The proportion of dropout at the 6-month follow-up was 65% (13/20) for the internet-based multimodal pain program group and 39% (9/23) for the occlusal splint group. In total, 51%

(22/43) of the enrolled patients did not complete the 6-month follow-up. Figure 3 also shows how many participants completed each module in the internet-based multimodal pain program treatment. Of the 20 participants in the internet-based multimodal pain program, 14 started treatment, and 8 completed all 7 modules of the program. Of those who completed the program, the time spent between the start-up phone calls and the last follow-up phone call was a median of 14 (IQR 8-15) weeks. No adverse events were reported by participants randomized to internet-based multimodal pain program, and none choose to receive additional treatment at the 3-month follow-up.

**Figure 3.** Attrition (left: the proportion of patients remaining in each treatment arm by event; right: an enlarged view of the proportion of internet-based multimodal pain program patients remaining by treatment module). iMPP: internet-based multimodal pain program; MRI: magnetic resonance imaging of the brain.



### Dropout Analysis

The dropout analysis showed no difference between completers and dropouts in the internet-based multimodal pain program group regarding demographic factors, clinical or psychosocial characteristics (Multimedia Appendix 3). In the occlusal splint group, the dropouts were significantly younger ( $P=.02$ ), had a lower proportion of married or de facto ( $P=.01$ ), a lower proportion of participants with full time employment ( $P<.01$ ), and a higher number of unspecific physical symptoms compared to completers ( $P=.03$ ; Multimedia Appendix 3).

### Posthoc Power

Posthoc power calculations showed a power of 11% and a sample size of 292 required for detecting a significant difference in characteristic pain intensity between 2 groups with  $\alpha=.05$  and  $\beta=.80$  (characteristic pain intensity at the 6-month follow-up—internet-based multimodal pain program: mean 2.85; occlusal splint: mean 2.17; entire sample SD 1.92).

## Discussion

### General

This study set out to assess the effectiveness of an internet-based multimodal pain program for adults with chronic TMD pain in comparison to occlusal splint therapy in general dentistry. No differences in clinical outcomes were found between the 2 treatment modalities. However, within the internet-based multimodal pain program group, jaw functional limitation was reduced and within the occlusal splint group, characteristic pain intensity and jaw functional limitation was reduced. As the study proceeded, a substantially higher dropout rate than expected was observed. Consequently, this pilot study showed that a randomized controlled trial with this design is not feasible.

### No Difference in Treatment Effect

This study did not find any difference in any of the recommended IMMPACT core outcome measures between the internet-based multimodal pain program and occlusal splint therapy. One possible explanation for this is the small sample size due to fewer participants recruited than planned and the substantial dropout rate. The lack of sufficient power due to the small sample size is a major limitation of this study regarding interpretation of treatment effect. Another possible explanation is our choice to include an active control condition. In this study, the effectiveness of internet-based multimodal pain program was compared to occlusal splint therapy—a treatment with moderate efficacy on TMD pain [5]. Our choice of this control condition in an early phase trial can be debated. However, according to the Helsinki declaration, “The benefits, risks, burdens, and effectiveness of a new intervention must be tested against those of the best proven intervention(s), except in the following circumstances: where no proven intervention exists, the use of placebo, or no intervention, is acceptable [32].”

### The Internet-Based Multimodal Pain Program Improves Jaw Function

Physical function is an important factor to assess in chronic pain studies [28]. In this study, we used the Jaw Function Limitation scale to assess jaw function. Within-group analysis indicated that the internet-based multimodal pain program increases jaw function for those with chronic TMD pain. This can be expected because the internet-based multimodal pain program comprises jaw exercises as well as psychoeducation and situational analyses that aim to reduce fear of movement [33]. The multimodal internet-based pain program therapy, therefore, seems to improve physical function. However, the findings cannot be generalized due to the small sample size.



Potential treatment effects on the other outcome measures were not possible to determine in this study. The lack of change regarding pain intensity and pain-related disability was somewhat unexpected, as others have shown that multimodal treatment and cognitive behavior therapy are effective for patients with chronic orofacial pain [34-36], and internet interventions have been shown to improve pain interference and pain intensity [9,37].

The IMMPACT core outcome domain of participant ratings of global improvement was not assessed in this study since it is not a part of the DC/TMD instrument that was used in this study. Future studies should consider assessing global change because it could provide an estimate of overall change in symptoms in addition to those of the primary outcomes. This may be especially important when investigating a multicomponent treatment package where the relative contribution of each active component is unknown.

### **Occlusal Splint Treatment Reduces Pain and Improves Jaw Function**

Within the occlusal splint group, characteristic pain intensity and jaw functional limitation were significantly reduced at the 6-month follow-up compared to baseline. The treatment effect of occlusal splint treatment on pain intensity was expected and is consistent with the literature [38]. A reduction in jaw functional limitation after occlusal splint therapy has also been reported [39]. It can, therefore, be assumed that the delivered control condition worked as expected.

### **Pretreatment Expectation**

IMMPACT recommends that expectations be considered as a phenotypic measure in clinical trials, and in this study, it was assessed with the multidimensional Stanford Expectations of Treatment Scale [40]. Research suggests that expectations influence the treatment outcome, and patients with positive expectations seem to benefit from medical treatment to a greater extent for a variety of conditions, including chronic pain [41]. The negative relation between positive expectancy and degree of depression in the internet-based multimodal pain program group is in line with this. Interestingly, this study also found a positive relation between positive expectancy and pain-related disability in the internet-based multimodal pain program group and that positive expectancy was positively associated with catastrophizing in the occlusal splint group. These relationships are interesting because it might suggest that positive expectations that are too high could have a negative impact on treatment outcome and also highlights the complex relationship of expectations on treatment and treatment outcome. Further studies are recommended to better understand the relationship between pretreatment expectations and outcomes of orofacial pain management protocols.

### **Randomized Control Trial Not Feasible**

Individuals with chronic TMD pain were effectively recruited with this study protocol. All eligible patients except one chose to participate in the study, suggesting that participation was appealing. Retaining participants, on the other hand, turned out to be more of a challenge.

The high proportion of loss to follow-up confirms the difficulty of collecting data in longitudinal studies. However, the total proportion of dropouts in this trial (22/43, 51%) was relatively high compared to those for other clinical chronic pain trials, for which attrition ranged from 5% to 46% [42]. It was also surprising that some participants dropped out before the treatment start. A possible explanation for this might be the mandatory MRI examinations before and after the treatment period, particularly given that the MRI examinations were performed in a different city than that in which the patients were recruited. In addition, technical issues with the MRI scanner during a 6-month period in 2017 caused further inconvenience for the participants.

In addition to high attrition, unbalanced dropout for the 2 treatment modalities at the 6-month follow-up was observed. This imbalance could be explained neither by a difference between the treatment groups regarding baseline characteristics, treatment expectations, adverse effects, or the need of rescue treatment nor by differences in baseline characteristics between dropouts and completers. The dropout rate in the occlusal splint group was 39% (9/23). A systematic review of randomized controlled trials on occlusal treatments in TMD reported a dropout rate of less than 10% [43]. We believe that the MRI examinations were a major factor in the unexpectedly high dropout rate in this group. In the internet-based multimodal pain program group, the dropout rate of 65% is comparable to those of other internet-based trials; attrition has been recognized as one of the methodological challenges in the evaluation of eHealth apps [44], and future studies should consider an allocation process that compensates for this risk. It is, however, clear that a randomized controlled trial with the current design is not feasible from economic and ethical perspectives due to the high attrition rate, even if the study may be considered practically feasible.

### **Future Studies**

Further work is required to explore the potential of internet-based treatment for chronic orofacial pain. However, since the feasibility of this study design was suboptimal, different approaches have to be considered. For a future trial, adjustments in study design should be considered to reduce nonusage and dropout attrition. Inspired by a prior study [45] on an eHealth app for pain with low nonusage attrition, changes to the internet-based multimodal pain program could include optimization of scheme and triggers for reminders and offer different levels of clinical support according to the patient's preferences. Furthermore, the internet-based treatment used by that study had fewer modules that were less time consuming compared to those in this study's internet-based multimodal pain program [45]. Fewer and more focused modules should, therefore, be considered. Other possible changes to the internet-based multimodal pain program to make it more user-friendly include converting the offline workbook to a digital format and developing an app for mobile devices. In addition to improving the internet-based multimodal pain program itself, changes in the study protocol and design should be considered. We believe that the additional pre and posttreatment MRI examination was the primary reason for the high dropout rate in this study. Accordingly, a study of the treatment effect of

internet-based multimodal pain program should not include MRI examinations. Also, conversion to a hybrid trial model with follow-up data collection online could make it more convenient to participate in the study and increase retention. Finally, novel study designs such as *n*-of-1 or noninferiority trials should be considered to optimize the quality and quantity of data acquired [46,47].

## Conclusion

This study was not able to demonstrate a difference in treatment outcome between an internet-based multimodal pain program

and occlusal splint therapy in patients with chronic TMD pain. However, within the internet-based multimodal pain program group, the results suggested that internet-based multimodal pain program improves jaw function. The results also confirmed the treatment effect of occlusal splint therapy in chronic TMD pain. Furthermore, the outcome of this pilot study showed that a randomized controlled trial with this design is not feasible due to a too high attrition rate.

## Acknowledgments

This study was supported by funds from the Oral Health-Related Research by Region Skåne, Sweden; the Swedish Dental Society, Sweden; and National Dental Care Skåne AB, Sweden. The funders had no role in the study design, data collection and analysis, preparation of the manuscript, or decision to publish. The authors would like to warmly thank Professor Ewa Carin Ekberg and Psykologpartners AB for the valuable collaboration in developing the internet-based multimodal pain program; Dr David Brohede MSc in psychology, for valuable support and comments on Table 2; and National Dental Care Skåne AB for the collaboration in this study. Special thanks to the personnel at the Fäladstorget clinic who helped us recruit, examine, and treat the participants of this study. We would also like to thank Professor Per Erik Isberg for valuable input on the statistical methods.

## Authors' Contributions

PA conceived the original idea, designed and planned the research project, helped with the analysis and interpretation of data, and helped with the editing of the manuscript. PS helped with the interpretation of data, editing of the manuscript, and revised it critically for important intellectual content. JL coordinated the data collection; managed, analyzed, and interpreted data; and drafted and edited the manuscript with input from all the authors. All authors have given their approval of the final manuscript.

## Conflicts of Interest

The Faculty of Odontology, Malmö University, Sweden may have a financial interest in the internet-based multimodal pain program, as the internet-based multimodal pain program may be licensed from us to other Swedish Dental Service organizations and specialist clinics.

### Multimedia Appendix 1

Informed consent documentation.

[PDF File (Adobe PDF File), 176 KB - [jmir\\_v22i10e22326\\_app1.pdf](#)]

### Multimedia Appendix 2

Demo information.

[PDF File (Adobe PDF File), 48 KB - [jmir\\_v22i10e22326\\_app2.pdf](#)]

### Multimedia Appendix 3

Table of baseline characteristics of dropouts and completers of the sample (patients with chronic temporomandibular pain, *n*=43) within the allocated treatment groups.

[PDF File (Adobe PDF File), 146 KB - [jmir\\_v22i10e22326\\_app3.pdf](#)]

## References

1. Metoder för behandling av långvarig smärta : en systematisk litteraturöversikt.V. 2. Stockholm: Statens beredning för medicinsk utvärdering [Swedish Agency For Health Technology Assessment And Assessment Of Social Services]; 2006.
2. LeResche L. Epidemiology of temporomandibular disorders: implications for the investigation of etiologic factors. *Crit Rev Oral Biol Med* 1997;8(3):291-305. [doi: [10.1177/10454411970080030401](#)] [Medline: [9260045](#)]
3. Sessle B. Orofacial pain : from basic science to clinical management : the transfer of knowledge in pain research to education. Chicago: Quintessence Pub; 2008.
4. Dahlström L, Carlsson GE. Temporomandibular disorders and oral health-related quality of life. A systematic review. *Acta Odontol Scand* 2010 Mar;68(2):80-85. [doi: [10.3109/00016350903431118](#)] [Medline: [20141363](#)]
5. National Board of Health and Welfare. Nationella riktlinjer för vuxentandvård 2011 : stöd för styrning och ledning. Stockholm: Socialstyrelsen; 2011.

6. Ström L, Pettersson R, Andersson G. A controlled trial of self-help treatment of recurrent headache conducted via the Internet. *J Consult Clin Psychol* 2000 Aug;68(4):722-727. [Medline: [10965647](#)]
7. Williams DA, Kuper D, Segar M, Mohan N, Sheth M, Clauw DJ. Internet-enhanced management of fibromyalgia: a randomized controlled trial. *Pain* 2010 Dec;151(3):694-702 [FREE Full text] [doi: [10.1016/j.pain.2010.08.034](#)] [Medline: [20855168](#)]
8. Buhrman M, Fältenhag S, Ström L, Andersson G. Controlled trial of Internet-based treatment with telephone support for chronic back pain. *Pain* 2004 Oct;111(3):368-377. [doi: [10.1016/j.pain.2004.07.021](#)] [Medline: [15363881](#)]
9. Slattery BW, Haugh S, O'Connor L, Francis K, Dwyer CP, O'Higgins S, et al. An Evaluation of the Effectiveness of the Modalities Used to Deliver Electronic Health Interventions for Chronic Pain: Systematic Review With Network Meta-Analysis. *J Med Internet Res* 2019 Jul 17;21(7):e11086 [FREE Full text] [doi: [10.2196/11086](#)] [Medline: [31317869](#)]
10. Dear BF, Titov N, Perry KN, Johnston L, Wootton BM, Terides MD, et al. The Pain Course: a randomised controlled trial of a clinician-guided Internet-delivered cognitive behaviour therapy program for managing chronic pain and emotional well-being. *Pain* 2013 Jun;154(6):942-950. [doi: [10.1016/j.pain.2013.03.005](#)] [Medline: [23688830](#)]
11. De Boer MJ, Versteegen GJ, Vermeulen KM, Sanderman R, Struys MMRF. A randomized controlled trial of an Internet-based cognitive-behavioural intervention for non-specific chronic pain: an effectiveness and cost-effectiveness study. *Eur J Pain* 2014 Nov;18(10):1440-1451. [doi: [10.1002/ejp.509](#)] [Medline: [24777973](#)]
12. Eccleston C, Fisher E, Craig L, Duggan GB, Rosser BA, Keogh E. Psychological therapies (Internet-delivered) for the management of chronic pain in adults. *Cochrane Database Syst Rev* 2014;2:CD010152. [doi: [10.1002/14651858.CD010152.pub2](#)] [Medline: [24574082](#)]
13. Schiffman E, Ohrbach R, Truelove E, Look J, Anderson G, Goulet J, International RDC/TMD Consortium Network, International association for Dental Research, Orofacial Pain Special Interest Group, International Association for the Study of Pain. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for Clinical and Research Applications: recommendations of the International RDC/TMD Consortium Network\* and Orofacial Pain Special Interest Group†. *J Oral Facial Pain Headache* 2014;28(1):6-27 [FREE Full text] [doi: [10.11607/jop.1151](#)] [Medline: [24482784](#)]
14. Lövgren A, Häggman-Henrikson B, Visscher CM, Lobbezoo F, Marklund S, Wänman A. Temporomandibular pain and jaw dysfunction at different ages covering the lifespan--A population based study. *Eur J Pain* 2016 Apr;20(4):532-540. [doi: [10.1002/ejp.755](#)] [Medline: [26311138](#)]
15. Lövgren A, Visscher CM, Häggman-Henrikson B, Lobbezoo F, Marklund S, Wänman A. Validity of three screening questions (3Q/TMD) in relation to the DC/TMD. *J Oral Rehabil* 2016 Oct;43(10):729-736. [doi: [10.1111/joor.12428](#)] [Medline: [27573533](#)]
16. Vilanova LSR, Garcia RCMR, List T, Alstergren P. Diagnostic criteria for temporomandibular disorders: self-instruction or formal training and calibration? *J Headache Pain* 2015;16:505 [FREE Full text] [doi: [10.1186/s10194-015-0505-9](#)] [Medline: [25916330](#)]
17. Von Korff M. Assessment of Chronic Pain in Epidemiological and Health Services Research: Empirical Bases and New Direction. In: Turk DC, Melzack R, editors. *Handbook of Pain Assessment*. New York: Guilford Press; 2011:455-473.
18. Pain drawing. International Network for Orofacial Pain and Related Disorders Methodology (INFORM). URL: [https://ubwp.buffalo.edu/rdc-tmdinternational/wp-content/uploads/sites/58/2017/01/DC-TMD-pain-drawing\\_2013\\_05\\_12.pdf](https://ubwp.buffalo.edu/rdc-tmdinternational/wp-content/uploads/sites/58/2017/01/DC-TMD-pain-drawing_2013_05_12.pdf) [accessed 2020-09-28]
19. Ohrbach R, Larsson P, List T. The jaw functional limitation scale: development, reliability, and validity of 8-item and 20-item versions. *J Orofac Pain* 2008;22(3):219-230. [Medline: [18780535](#)]
20. Markiewicz MR, Ohrbach R, McCall WD. Oral behaviors checklist: reliability of performance in targeted waking-state behaviors. *J Orofac Pain* 2006;20(4):306-316. [Medline: [17190029](#)]
21. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001 Sep;16(9):606-613 [FREE Full text] [Medline: [11556941](#)]
22. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom Med* 2002;64(2):258-266. [doi: [10.1097/00006842-200203000-00008](#)] [Medline: [11914441](#)]
23. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006 May 22;166(10):1092-1097. [doi: [10.1001/archinte.166.10.1092](#)] [Medline: [16717171](#)]
24. Nordin M, Nordin S. Psychometric evaluation and normative data of the Swedish version of the 10-item perceived stress scale. *Scand J Psychol* 2013 Dec;54(6):502-507. [doi: [10.1111/sjop.12071](#)] [Medline: [24118069](#)]
25. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. *Psychological Assessment* 1995;7(4):524-532. [doi: [10.1037/1040-3590.7.4.524](#)]
26. Younger J, Gandhi V, Hubbard E, Mackey S. Development of the Stanford Expectations of Treatment Scale (SETS): a tool for measuring patient outcome expectancy in clinical trials. *Clin Trials* 2012 Dec;9(6):767-776. [doi: [10.1177/1740774512465064](#)] [Medline: [23169874](#)]
27. Ramfjord SP, Ash MM. Reflections on the Michigan occlusal splint. *J Oral Rehabil* 1994 Sep;21(5):491-500. [doi: [10.1111/j.1365-2842.1994.tb01164.x](#)] [Medline: [7996334](#)]
28. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2003 Dec;106(3):337-345. [Medline: [14659516](#)]

29. Haythornthwaite JA. IMMPACT recommendations for clinical trials: opportunities for the RDC/TMD. *J Oral Rehabil* 2010 Oct;37(10):799-806. [doi: [10.1111/j.1365-2842.2010.02152.x](https://doi.org/10.1111/j.1365-2842.2010.02152.x)] [Medline: [20887278](#)]
30. Gracely RH, Petzke F, Wolf JM, Clauw DJ. Functional magnetic resonance imaging evidence of augmented pain processing in fibromyalgia. *Arthritis Rheum* 2002 May;46(5):1333-1343 [FREE Full text] [doi: [10.1002/art.10225](https://doi.org/10.1002/art.10225)] [Medline: [12115241](#)]
31. GraphPad Prism 7 Statistics Guide - Sample size for nonparametric tests. GraphPad Software. URL: [https://www.graphpad.com/guides/prism/7/statistics/stat\\_sample\\_size\\_for\\_nonparametric.htm](https://www.graphpad.com/guides/prism/7/statistics/stat_sample_size_for_nonparametric.htm) [accessed 2020-09-28]
32. General Assembly of the World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *J Am Coll Dent* 2014;81(3):14-18. [Medline: [25951678](#)]
33. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain* 2000 Apr;85(3):317-332. [Medline: [10781906](#)]
34. Aggarwal VR, Tickle M, Javidi H, Peters S. Reviewing the evidence: can cognitive behavioral therapy improve outcomes for patients with chronic orofacial pain? *J Orofac Pain* 2010;24(2):163-171. [Medline: [20401354](#)]
35. Aggarwal VR, Fu Y, Main CJ, Wu J. The effectiveness of self-management interventions in adults with chronic orofacial pain: A systematic review, meta-analysis and meta-regression. *Eur J Pain* 2019 May;23(5):849-865. [doi: [10.1002/ejp.1358](https://doi.org/10.1002/ejp.1358)] [Medline: [30620145](#)]
36. Türp JC, Jokstad A, Motschall E, Schindler HJ, Windecker-Gétaz I, Ettlin DA. Is there a superiority of multimodal as opposed to simple therapy in patients with temporomandibular disorders? A qualitative systematic review of the literature. *Clin Oral Implants Res* 2007 Jun;18 Suppl 3:138-150. [doi: [10.1111/j.1600-0501.2007.01480.x](https://doi.org/10.1111/j.1600-0501.2007.01480.x)] [Medline: [17594378](#)]
37. Buhrman M, Gordh T, Andersson G. Internet interventions for chronic pain including headache: A systematic review. *Internet Interv* 2016 May;4:17-34 [FREE Full text] [doi: [10.1016/j.invent.2015.12.001](https://doi.org/10.1016/j.invent.2015.12.001)] [Medline: [30135787](#)]
38. Al-Moraissi EA, Farea R, Qasem KA, Al-Wadeai MS, Al-Sabahi ME, Al-Iryani GM. Effectiveness of occlusal splint therapy in the management of temporomandibular disorders: network meta-analysis of randomized controlled trials. *Int J Oral Maxillofac Surg* 2020 Jan 22. [doi: [10.1016/j.ijom.2020.01.004](https://doi.org/10.1016/j.ijom.2020.01.004)] [Medline: [31982236](#)]
39. Doepel M, Nilner M, Ekberg E, LE Bell Y. Long-term effectiveness of a prefabricated oral appliance for myofascial pain. *J Oral Rehabil* 2012 Apr;39(4):252-260. [doi: [10.1111/j.1365-2842.2011.02261.x](https://doi.org/10.1111/j.1365-2842.2011.02261.x)] [Medline: [21985440](#)]
40. Laferton JAC, Kube T, Salzmann S, Auer CJ, Shedden-Mora MC. Patients' Expectations Regarding Medical Treatment: A Critical Review of Concepts and Their Assessment. *Front Psychol* 2017;8:233 [FREE Full text] [doi: [10.3389/fpsyg.2017.00233](https://doi.org/10.3389/fpsyg.2017.00233)] [Medline: [28270786](#)]
41. Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, et al. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. *Pain* 2016 Sep;157(9):1851-1871 [FREE Full text] [doi: [10.1097/j.pain.0000000000000602](https://doi.org/10.1097/j.pain.0000000000000602)] [Medline: [27152687](#)]
42. Turk DC, Rudy TE. Neglected factors in chronic pain treatment outcome studies--referral patterns, failure to enter treatment, and attrition. *Pain* 1990 Oct;43(1):7-25. [Medline: [2277718](#)]
43. Forssell H, Kalso E, Koskela P, Vehmanen R, Puukka P, Alanen P. Occlusal treatments in temporomandibular disorders: a qualitative systematic review of randomized controlled trials. *Pain* 1999 Dec;83(3):549-560. [doi: [10.1016/s0304-3959\(99\)00160-8](https://doi.org/10.1016/s0304-3959(99)00160-8)] [Medline: [10568864](#)]
44. Eysenbach G. The law of attrition. *J Med Internet Res* 2005;7(1):e11 [FREE Full text] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](#)]
45. Dear BF, Gandy M, Karin E, Fogliati R, Fogliati VJ, Staples LG, et al. The Pain Course: 12- and 24-Month Outcomes From a Randomized Controlled Trial of an Internet-Delivered Pain Management Program Provided With Different Levels of Clinician Support. *J Pain* 2018 Dec;19(12):1491-1503. [doi: [10.1016/j.jpain.2018.07.005](https://doi.org/10.1016/j.jpain.2018.07.005)] [Medline: [30099209](#)]
46. Porcino AJ, Shamseer L, Chan AW, Kravitz RL, Orkin A, Punja S, SPENT group. SPIRIT extension and elaboration for n-of-1 trials: SPENT 2019 checklist. *BMJ* 2020 Mar 27;368:m122. [doi: [10.1136/bmj.m122](https://doi.org/10.1136/bmj.m122)] [Medline: [32107202](#)]
47. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006 Mar 08;295(10):1152-1160. [doi: [10.1001/jama.295.10.1152](https://doi.org/10.1001/jama.295.10.1152)] [Medline: [16522836](#)]

## Abbreviations

**DC/TMD:** Diagnostic Criteria for Temporomandibular Disorders

**IMMPACT:** Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials

**MRI:** magnetic resonance imaging

**TMD:** temporomandibular disorders

*Edited by G Eysenbach; submitted 09.07.20; peer-reviewed by P Gazerani, J Egan; comments to author 30.07.20; revised version received 10.08.20; accepted 18.08.20; published 13.10.20.*

*Please cite as:*

*Lam J, Svensson P, Alstergren P*

*Internet-Based Multimodal Pain Program With Telephone Support for Adults With Chronic Temporomandibular Disorder Pain: Randomized Controlled Pilot Trial*

*J Med Internet Res 2020;22(10):e22326*

URL: <http://www.jmir.org/2020/10/e22326/>

doi: [10.2196/22326](https://doi.org/10.2196/22326)

PMID: [33048053](https://pubmed.ncbi.nlm.nih.gov/33048053/)

©Julia Lam, Peter Svensson, Per Alstergren. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 13.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Identifying the Value of an eHealth Intervention Aimed at Cognitive Impairments: Observational Study in Different Contexts and Service Models

Monika Jurkeviciute<sup>1</sup>, MSc; Lex van Velsen<sup>2</sup>, PhD; Henrik Eriksson<sup>1</sup>, PhD; Svante Lifvergren<sup>1,3</sup>, MD, PhD; Pietro Davide Trimarchi<sup>4</sup>, PhD; Ulla Andin<sup>3</sup>, MD, PhD; Johan Svensson<sup>3,5</sup>, MD, PhD

<sup>1</sup>Centre for Healthcare Improvement, Chalmers University of Technology, Gothenburg, Sweden

<sup>2</sup>eHealth Group, Roessingh Research and Development, Enschede, Netherlands

<sup>3</sup>Skaraborg Hospital Group, Lidköping, Sweden

<sup>4</sup>IRCCS Fondazione Don Carlo Gnocchi, Milan, Italy

<sup>5</sup>Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

**Corresponding Author:**

Monika Jurkeviciute, MSc

Centre for Healthcare Improvement

Chalmers University of Technology

Vera Sandbergs Allé 8

Gothenburg, 41296

Sweden

Phone: 46 766061558

Email: [monika.jurkeviciute@chalmers.se](mailto:monika.jurkeviciute@chalmers.se)

## Abstract

**Background:** Value is one of the central concepts in health care, but it is vague within the field of summative eHealth evaluations. Moreover, the role of context in explaining the value is underexplored, and there is no explicit framework guiding the evaluation of the value of eHealth interventions. Hence, different studies conceptualize and operationalize value in different ways, ranging from measuring outcomes such as clinical efficacy or behavior change of patients or professionals to measuring the perceptions of various stakeholders or in economic terms.

**Objective:** The objective of our study is to identify contextual factors that determine similarities and differences in the value of an eHealth intervention between two contexts. We also aim to reflect on and contribute to the discussion about the specification, assessment, and relativity of the “value” concept in the evaluation of eHealth interventions.

**Methods:** The study concerned a 6-month eHealth intervention targeted at elderly patients (n=107) diagnosed with cognitive impairment in Italy and Sweden. The intervention introduced a case manager role and an eHealth platform to provide remote monitoring and coaching services to the patients. A model for evaluating the value of eHealth interventions was designed as monetary and nonmonetary benefits and sacrifices, based on the value conceptualizations in eHealth and marketing literature. The data was collected using the Mini-Mental State Examination (MMSE), the clock drawing test, and the 5-level EQ-5D (EQ-5D-5L). Semistructured interviews were conducted with patients and health care professionals. Monetary data was collected from the health care and technology providers.

**Results:** The value of an eHealth intervention applied to similar types of populations but differed in different contexts. In Sweden, patients improved cognitive performance (MMSE mean 0.85, SD 1.62,  $P<.001$ ), reduced anxiety (EQ-5D-5L mean 0.16, SD 0.54,  $P=.046$ ), perceived their health better (EQ-5D-5L VAS scale mean 2.6, SD 9.7,  $P=.035$ ), and both patients and health care professionals were satisfied with the care. However, the Swedish service model demonstrated an increased cost, higher workload for health care professionals, and the intervention was not cost-efficient. In Italy, the patients were satisfied with the care received, and the health care professionals felt empowered and had an acceptable workload. Moreover, the intervention was cost-effective. However, clinical efficacy and quality of life improvements have not been observed. We identified 6 factors that influence the value of eHealth intervention in a particular context: (1) service delivery design of the intervention (process of delivery), (2) organizational setup of the intervention (ie, organizational structure and professionals involved), (3) cost of different treatments, (4) hourly rates of staff for delivering the intervention, (5) lifestyle habits of the population (eg, how physically active they were in their daily life and if they were living alone or with family), and (6) local preferences on the quality of patient care.

**Conclusions:** Value in the assessments of eHealth interventions need to be considered beyond economic terms, perceptions, or behavior changes. To obtain a holistic view of the value created, it needs to be operationalized into monetary and nonmonetary outcomes, categorizing these into benefits and sacrifices.

(*J Med Internet Res* 2020;22(10):e17720) doi:[10.2196/17720](https://doi.org/10.2196/17720)

## KEYWORDS

eHealth value; evaluation of value; eHealth intervention; cognitive impairment; role of context; cost benefit

## Introduction

### The Concept of Value in eHealth

The concept of value has taken a central role in health care, including eHealth development and evaluation. This trend has been ongoing in health care since 2006, when Porter and Teisberg introduced the value-based health care concept to improve and innovate healthcare [1], resulting in increased interest in the *value* concept. For example, some argue that there has been a shift in the health care discourse, from improvement of quality to improvement of value [2]. Moreover, many improvements using the value concept have been reported in recent years [3]. However, the concept of value is somewhat unclear within the context of summative eHealth evaluation. Subsequently, it has been approached in different ways in various studies. Sometimes, value has been investigated as positive outcomes such as clinical efficacy or behavior change of patients or professionals [4,5]. Others have aimed to identify value through the perceptions of various stakeholders [6,7]. In order to arrange reimbursement for an eHealth service and to support decisions about investments in technology development or implementation, the value of eHealth has also been interpreted in economic terms (ie, whether health outcomes justify the costs) [8]. While different evaluation frameworks have addressed some aspects of value [9], there is no framework that explicitly guides the evaluation of value in eHealth interventions. The immature conceptual and methodological base of value in eHealth can create confusion due to the large number of studies that are hard to compare, learn from, and transfer from one context to another.

To explore the conceptualizations of value beyond the area of eHealth, inspiration could be taken from other disciplines, such as the marketing of products and services. One way to approach value can be as benefits and sacrifices, which measure service quality in relation to cost [10] and can reflect both monetary and nonmonetary outcomes [11]. In addition, value occurs from

the interaction between a subject (eg, a patient) and an object (an eHealth intervention) and is relative. Relativity means that value is comparative (signifying that the value of one eHealth intervention can be compared to the other), personal (what is valuable for one end-user or stakeholder is not necessarily valuable for the other), and situational (ie, value depends on the context of use) [10].

Assuming that value is relative, one can question the usefulness of evaluating a certain eHealth intervention in a specific context. How do the results translate from one context to the next when the concept of value, costs, and benefits are potentially different? Systematic reviews that have analyzed eHealth intervention studies in dementia care [12] and eHealth evaluation frameworks [13] revealed that the role of context has been neglected. The current discourse treats the context as the circumstances under which an intervention is effective or not [14]. However, there are no studies that have explicitly investigated how the context influences the value of the eHealth intervention. Knowledge of these contextual factors can help to translate the interventions into practice or new settings [15-17].

To sum up, one way of conceptualizing the assessment of value in eHealth interventions could be to combine the benefit-and-sacrifice approach from marketing [8] with the value-based health care logic that assesses patient outcomes against cost [1]. In this view, benefits could refer to financial earnings or savings as monetary benefits, and eHealth service quality or utility as nonmonetary benefits [18]. Sacrifices relate to the financial investment and expenditure as monetary sacrifices, and social disadvantages (ie, what it takes to provide the service physically or emotionally) as nonmonetary sacrifices. In addition, it might also be fruitful to add an emphasis on the context, which has been lacking in eHealth studies [12,13,17]. The proposed structure of assessing the value of an eHealth intervention is depicted in [Textbox 1](#).

**Textbox 1.** Structure of the value assessment of an eHealth intervention.

<b>Benefits</b> <ul style="list-style-type: none"><li>• Monetary</li><li>• Nonmonetary</li></ul>
<b>Sacrifices</b> <ul style="list-style-type: none"><li>• Monetary</li><li>• Nonmonetary</li></ul>

In this study, we identified the value of a nonpharmacological eHealth intervention, which combined an integrated care model

with eHealth and targeted to treat cognitive impairment in elderly populations in Italy and Sweden.

The 2 objectives that guided this study were (1) to identify the contextual factors that determined the similarities and differences in the value of an eHealth intervention between the 2 contexts, and (2) to reflect on and contribute to the discussion about the specification, assessment, and relativity of the “value” concept in the evaluation of eHealth interventions.

## Context

This study was a part of the European Union–funded project “Digital Environment for Cognitive Inclusion” (DECI). The entire DECI project was performed over 3 years and focused on the development of digital solutions to improve the care of elderly individuals with mild cognitive impairment (MCI) and mild dementia (MD).

In our study, we examined the implementation of a 6-month eHealth intervention among patients recruited in Sweden and Italy. The inclusion criteria for participants were  $\geq 60$  years of age, a diagnosis of MCI or a diagnosis of dementia according to DSM-5 criteria, a clinical dementia rating (CDR) of  $\leq 1$ , living at home, and the ability to provide informed consent or the availability of a proxy for informed consent. The exclusion criteria were living in a care institution, previous or present major psychiatric illness (eg, schizophrenia, bipolar disorder,

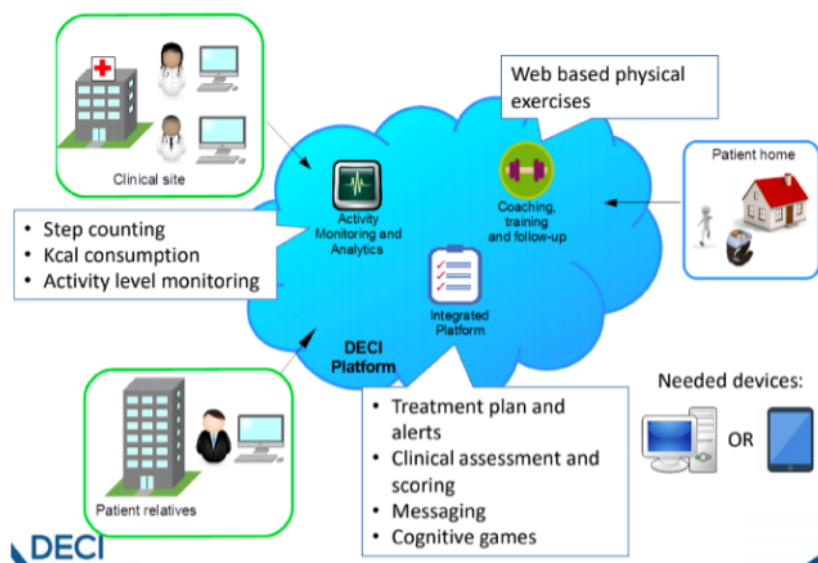
or recurrent major depression), previous or present major neurological illness other than MCI or MD (eg, stroke, multiple sclerosis, brain tumor, traumatic brain injury), the presence of other serious comorbidities (eg, severe chronic obstructive pulmonary disease, severe heart disease, or severe chronic kidney failure), a history of drug or alcohol abuse, severe sensory impairments (mainly visual and auditory), a history of intellectual disability or other developmental diseases, and a life expectancy of less than 1 year (as judged by a clinician). The specific procedures for the inclusion of patients in Italy and Sweden differed. After inclusion, the 6-month eHealth intervention was initiated.

Clinical efficacy variables and quality of life were assessed at baseline and at the end of the 6-month intervention. Further data in terms of monetary and nonmonetary outcomes were obtained by performing semistructured interviews 6 months after the eHealth intervention and by collecting information from the health care providers after the intervention had been completed.

## The DECI Intervention

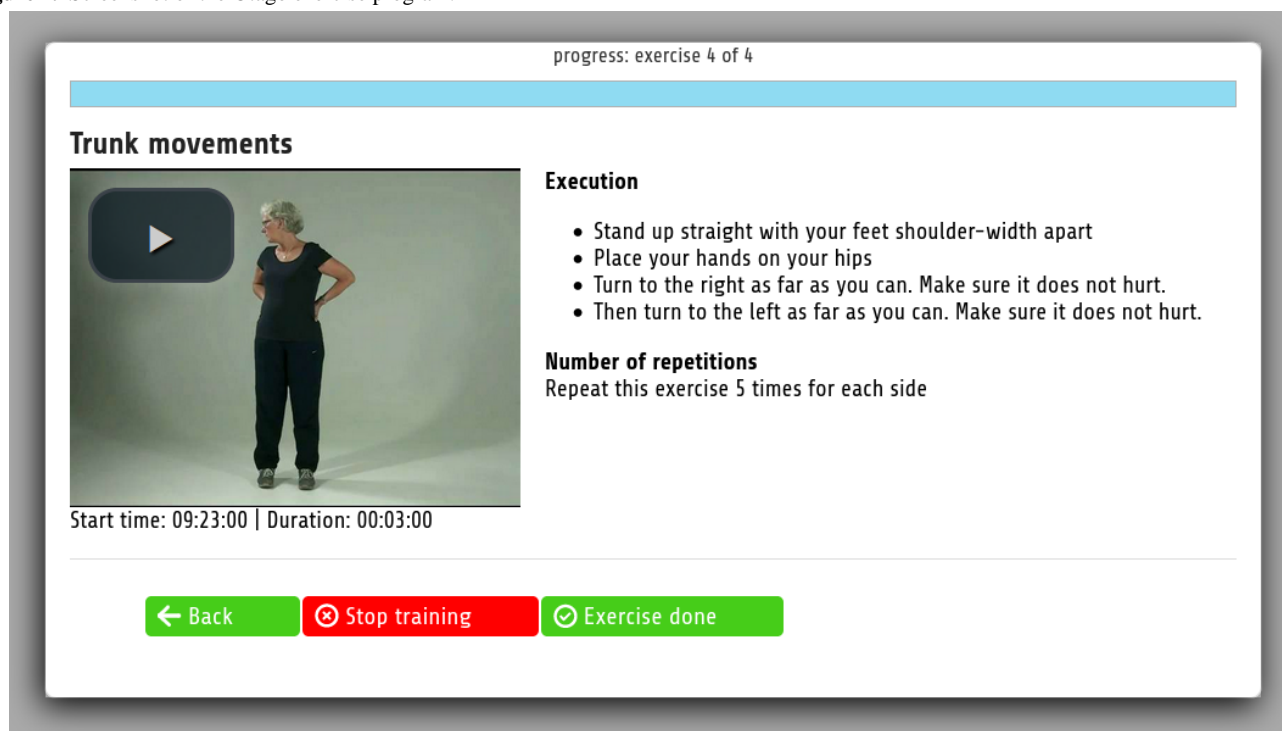
The DECI intervention consisted of a set of integrated eHealth applications and a service model, as depicted in Figure 1.

**Figure 1.** Components of the Digital Environment for Cognitive Inclusion (DECI) intervention.



The digital component of the intervention was centered around a web-based portal that disclosed a set of several services via single sign-on. A digital platform that served as the data repository for a patient allowed for the following main functionalities: messaging among the key actors for a patient’s care (the patient, professional caregivers, family members, and informal caregivers), sending data to and retrieving data from a patient’s electronic medical record (EMR) and other digital applications, and the provision and processing of patient-centered surveys. A web-based service provided the

Otago fall prevention program [19] (used for improving physical health [20]) through video instruction, accompanied by written instructions and a voice-over that pronounced these written instructions. Patients could indicate the difficulty they had with exercises; based upon this feedback, the service decided whether (1) they could continue to the next level of the Otago program after completing a week, (2) they would remain at the same level, or (3) they would go down a level. Figure 2 depicts a screenshot of an exercise in the Otago program.

**Figure 2.** Screenshot of the Otago exercise program.

Smartbrain [21], a web-based collection of exercises aimed at cognitive stimulation normally provided to patients with MCI, Alzheimer disease, Parkinson disease, and similar conditions, offered patients different difficulty levels. In addition, a physical activity dashboard supported by the ADAMO care watch (Caretek s.r.l) and that included pedometer functionality [22] allowed care professionals to inspect a patient's physical activity (in terms of daily steps and active hours) and provided the patient with daily feedback about his or her physical activity levels.

The patients received a tablet for accessing digital services. Next, the case manager role was introduced as the organizational element of the intervention. The case manager was responsible for coordinating care and introducing the technology to the patients.

### *The Italian Context of the DECI Intervention*

The Italian National Health System provides free health care for patients throughout the country. It is funded by taxes and charges for some services. The DECI study was carried out in the Milan area (Lombardy region) at the Istituto Palazzolo (IP), Fondazione Don Gnocchi (FDG). All services are integrated to offer a care program that is shared between the elderly patients and the caregivers, and to provide follow-up along all the clinical pathways. The DECI project mainly targeted patients enrolled at the Memory Clinic of the Istituto Palazzolo.

Patients eligible for inclusion were identified using the IP-FDG Memory Clinic database and were then contacted to receive information on the study and a proposal of participation. Then, an inclusion visit was arranged for written informed consent, clinical anamnesis, and a neuropsychological and functional examination of patients that accepted participation and met the inclusion criteria. Afterward, patients received training and delivery of the study materials. If needed, a follow-up visit was

arranged. Lastly, clinical and satisfaction assessments were performed at the exit visit.

The Italian DECI team consisted of 1 senior physician specialized in geriatrics and dementia, 1 social worker (case manager), 2 neuropsychologists, and 1 engineer. A physiotherapist from the rehabilitation services gave advice but did not participate actively in visits and follow-up. The patients received a 10.1-inch Samsung Android tablet with a 4G SIM card to guarantee connectivity. The case manager and one of the neuropsychologists provided instructions and training for the patient regarding the DECI platform, usually during an individualized 90-minute meeting at the memory clinic. Related information was also available through the case manager by phone or through the message system of the DECI platform. The patients used the tablet for physical and cognitive training during the 6-month intervention. An online help desk by phone was established during working hours (Monday to Friday). Inclusion visits and follow-up were carried out at the memory clinic.

### *The Swedish Context of the DECI Intervention*

Health care in Sweden is tax-funded and provided by the communities (elderly care) and the counties (specialized and primary care). The DECI study was carried out in the Skaraborg region, which has a permanent interorganizational network for integrated care that delivers mobile, coordinated person-centered care for patients with various chronic diseases. DECI patients receive mobile, networked care managed by the Swedish DECI team.

The study purpose was described in advertisements in local media and leaflets that were distributed in local care units. Then, patients who contacted the Swedish DECI team were invited for screening. Patients who met the inclusion criteria and agreed to participate were visited at home for written informed consent,

clinical anamnesis, and neuropsychological and functional examination; inclusion and exclusion criteria were checked further. During subsequent visits, patients received training and delivery of the material. Lastly, clinical and satisfaction assessments were performed at the exit visit at the patient's home.

The DECI team consisted of 1 senior physician specialized in geriatrics from Skaraborg Hospital Group and 1 experienced nurse from one of the communities (case manager). An occupational therapist and a physiotherapist from 2 other communities in Skaraborg gave advice but did not participate in the home visits. All patients received a 10.1-inch Samsung Android tablet with a 4G SIM card to guarantee connectivity. The patients were assessed using memory tests. The case manager and the geriatrician provided instructions and training for the patient regarding the DECI platform, usually during a

1-hour long meeting at the patient's home. Information was also given during subsequent visits at the patients' home by the case manager, in some cases with the geriatrician also present. The patient used the tablet for physical and cognitive training during the 6-month intervention. An online help desk by phone was established during working hours (Monday to Friday). Follow-up was carried out at the patients' homes.

## Methods

### Value Specification

The evaluation conducted in this study was based on [Textbox 1](#). In order to meet the requirements and aims of the DECI project, the model presented in [Textbox 1](#) was populated with the variables collected during the project. The specification of value for the DECI intervention is depicted in [Textbox 2](#).

**Textbox 2.** Value specification in the Digital Environment for Cognitive Inclusion (DECI) intervention.

<b>Benefits</b>
<ul style="list-style-type: none"> <li>• <i>Monetary</i> <ul style="list-style-type: none"> <li>• Income</li> <li>• Prevented cost of treatment</li> </ul> </li> <li>• <i>Nonmonetary</i> <ul style="list-style-type: none"> <li>• Clinical efficacy</li> <li>• Quality of life</li> <li>• Patient satisfaction</li> <li>• Job satisfaction</li> </ul> </li> </ul>
<b>Sacrifices</b>
<ul style="list-style-type: none"> <li>• <i>Monetary</i> <ul style="list-style-type: none"> <li>• Investment</li> <li>• Operating expenses</li> <li>• Cost of spent time</li> </ul> </li> <li>• <i>Nonmonetary</i> <ul style="list-style-type: none"> <li>• Patient safety</li> <li>• Workload</li> </ul> </li> </ul>

The monetary benefits were operationalized as income and the prevented cost of treatment. The nonmonetary benefits were expressed as clinical efficacy, quality of life, patient satisfaction, and job satisfaction. The monetary sacrifices were operationalized as investment, operating expenses, and the cost of spent time. The nonmonetary sacrifices were expressed as patient safety and workload.

### Data Collection and Analysis

Data to assess the monetary and nonmonetary outcomes of DECI ([Textbox 2](#)) were collected from the Italian and Swedish health care providers involved in the project. Details of the collection and analysis of data are described below.

### Monetary Benefits

#### Income

Income data was related to the annual state reimbursement for the treatment of MCI for a single patient. The total yearly income was calculated by multiplying the projected yearly population in the treatment by the yearly state reimbursement per patient.

#### Prevented Cost of Treatment

Prevented cost included the postponement of care and the prevented cost of treatment of falls. Postponement of care was based on the delayed conversion from MCI to MD. Data was collected using the Clinical Dementia Rating Scale (CDR) [23]



at baseline and at follow-up after 6 months. The annual cost of treatment of MD was collected from the Italian and Swedish health care providers involved in the project. In the analysis of data, the conversion rate (changes in CDR after 6 months) was multiplied by the annual targeted population and the cost of MD treatment.

Data regarding falls were theorized based on the assumptions extracted from the relevant literature, since such data was not collected in the DECI study. Elderly people (from 65 years of age and older) fall 0.33 times a year [24], which can create care needs such as visits to a general practitioner (GP; 9%), emergency visits (5%) [25], or treatments for fracture (3%, based on the professional judgment of health care professionals in the DECI study). Previous studies involving the Otago physical activity coaching program have demonstrated that it helps to prevent falls by 68% [26]. The cost information for GP visits, emergency visits, and treatments for fracture was collected from the Italian and Swedish health care providers involved in the project. During the data analysis, a preventable number of falls was calculated by multiplying 0.33 by 68% and by a targeted population size. Then, a number of prevented GP visits, emergency visits, and fractures were calculated by multiplying the number of falls by 9%, 5%, and 3%, respectively. To calculate the cost, the results were multiplied by the costs for a single GP visit, emergency visit, and fracture treatment. The results were summed up to calculate the total preventable costs due to falls.

### ***Nonmonetary Benefits***

#### **Clinical Efficacy**

The Mini-Mental State Examination (MMSE) [27] and the clock drawing test (CDT) [28] were used to assess the cognitive performance of the patients. The instruments were administered at baseline (T0) and at follow-up after 6 months of intervention (T1). Within-group differences were calculated by comparing values at T1 with those at T0 using the Wilcoxon test. Between-group differences were analyzed by comparing the changes from baseline (T1–T0) using the Mann-Whitney U test. The statistical analysis was performed using SPSS for Windows (version 24; IBM Corp). A *P* value of <.05 was considered statistically significant.

#### **Quality of Life**

The EQ-5D-5L [29] questionnaire was used to estimate quality of life at baseline and at follow-up after 6 months. Within-group differences were calculated using the Wilcoxon test, and between-group differences were analyzed by comparing the changes from baseline using the Mann-Whitney U test.

#### **Patient Satisfaction**

Data were collected using semistructured interviews with patients in Sweden and Italy after 6 months of having used the DECI services (the interview protocol can be found in [Multimedia Appendix 1](#)). The sampling was purposeful [30] in order to be careful not to create too big of a cognitive burden during the assessment activities. Thematic analysis [31] of the data helped to identify the perceived benefits by the patients and the necessary sacrifices in order to deliver it.

#### **Job Satisfaction**

Data were collected using in-depth semistructured interviews with the health care professionals providing the DECI services in Italy and Sweden services (the interview protocol can be found in [Multimedia Appendix 2](#)). Thematic analysis [31] of the data identified the perceived benefits by the health care professionals and the necessary sacrifices in order to deliver it.

### ***Monetary Sacrifices***

#### **Cost of Spent Time**

The spent time included the direct provision of the DECI services and participation in the multi-disciplinary meetings (those hours did not overlap). The hours spent by healthcare professionals were collected from the Swedish and Italian healthcare providers. The cost was then calculated by multiplying the number of hours by the hourly tariff. The total cost of spent time was a sum of costs of the different professional categories.

#### **Investment**

Investment related to the one-time cost of starting to use the DECI technologies. Investment data were collected via email from the healthcare providers.

#### **Operating Expenses**

Operating expenses included the annual cost of hardware (tablets), servers, the usage fees of DECI technologies, maintenance, and a help-desk function. Data were collected from the healthcare and technology providers using an Excel file containing various categories of the operating expenses. The usage fees of the DECI technology were stable, except for the fee for the ADAMO activity monitoring device (a wristwatch) that depended on the number of users.

### ***Nonmonetary Sacrifices***

#### **Patient Safety**

Data were collected using semistructured interviews with patients and health care professionals. The data were thematically analyzed in order to identify safety-related issues.

#### **Workload**

Data were collected through semistructured interviews with health care professionals. The data were thematically analyzed in order to identify the workload-related issues.

#### **Monetary Benefit/Sacrifice Ratio**

The ratio was calculated by dividing the sum of monetary benefits by the sum of monetary sacrifices (Ratio=Benefits/Sacrifices). The ratio was calculated for the scenarios at year 1, year 2, and year 3 of using the DECI services.

#### **Ethics**

The DECI study was approved by the Ethical Committee of the Fondazione Don Carlo Gnocchi and the Regional Ethical Committee of Gothenburg.

## Results

### Demographic Characteristics

Table 1 depicts the demographic characteristics of the patients involved in the study.

**Table 1.** Demographic characteristics of patients in Italy (n=53) and Sweden (n=54) at baseline.

Characteristics	Patients in Italy (n=53)	Patients in Sweden (n=54)	P value <sup>a</sup>
Age in years, mean (SD)	77.6 (5.3)	74.8 (5.9)	<.001
<b>Gender, n (%)</b>			.66
Female	27 (51)	30 (56)	
Male	26 (49)	24 (44)	
<b>Diagnosis, n (%)</b>			.007
MCI <sup>b</sup>	39 (74)	49 (91)	
MD <sup>c</sup>	14 (26)	5 (9)	
Education years, mean (SD)	9.2 (4.3)	11.6 (2.9)	<.001
MMSE <sup>d</sup> (range 0-30), mean (SD)	26.6 (2.9)	28.2 (1.4)	<.001
CDT <sup>e</sup> (range 0-5), mean (SD)	3.30 (1.38)	4.81 (0.52)	<.001

<sup>a</sup>Differences between groups were examined using the Mann-Whitney U test for continuous variables and using chi-square tests for categorical variables.

<sup>b</sup>MCI: mild cognitive impairment.

<sup>c</sup>MD: mild dementia.

<sup>d</sup>MMSE: Mini-Mental State Examination.

<sup>e</sup>CDT: clock drawing test.

### Monetary Benefits

#### Prevented Cost of Treatment

The results in Italy showed a 10% lower conversion rate from MCI to MD compared to the conversion rates of patients in regular care. In the Swedish patients, who were younger than the Italian patients and had a less marked decrease in cognitive function at baseline (higher MMSE and CDT scores; Table 1), the results showed a 2% higher conversion rate from MCI to MD compared to the patients in regular care, thus bringing no preventable costs from the postponed care. Since the postponement of care (which is the element of the preventable

costs) was zero, the Swedish preventable costs consisted of the prevention of falls only.

#### Income

The Italian and Swedish health care systems reimburse the cost of time when providing the eHealth-supported care. Therefore, the cost of time spent on providing the DECI treatment (a category in monetary sacrifices) was considered as income for year 1, year 2, and year 3.

#### Summary of Monetary Benefits

Table 2 shows a summary of the monetary benefits per patient in Italy and Sweden for 3 years.

**Table 2.** Monetary benefits, in euros (a currency exchange rate of EUR €1=US \$1.18 is applicable).

Monetary benefits in Italy and Sweden	Year 1 (€)	Year 2 (€)	Year 3 (€)
<b>Preventable costs - Italy</b>			
Total	191.696	287.544	335.468
Per patient	1916	1916	1916
<b>Preventable costs - Sweden</b>			
Total	6155	9232	10.770
Per patient	12	12	12
<b>Income - Italy</b>			
Total	67.938	101.906	118.891
Per patient	679	679	679
<b>Income - Sweden</b>			
Total	826.882	1.240.324	1.447.044
Per patient	1653	1653	1653

The large differences in monetary values between Italy and Sweden occurred due to the different targeted population sizes. In Italy, it was 100 patients in year 1, 150 patients in year 2, and 175 patients in year 3. In Sweden, it was 500 patients in year 1, 750 patients in year 2, and 875 patients in year 3.

## Nonmonetary Benefits

### *Clinical Efficacy*

The mean (SD) changes in MMSE and CDT scores in Italy were -0.14 (2.86) and 0.23 (1.52), respectively; in Sweden, these changes were 0.85 (1.62) and -0.11 (0.57), respectively. In Italy, both MMSE and CDT scores at the 6-month follow-up were similar to those at baseline ( $P=.35$  and  $P=.34$ , respectively). In Sweden, MMSE scores were higher (better) at the 6-month follow-up compared to those at baseline ( $P<.001$ ), whereas CDT scores were unchanged ( $P=.18$ ). When comparing the results in Italy versus those in Sweden, the changes in MMSE scores were significantly greater in the Swedish cohort ( $P=.004$ ), whereas there were no differences in CDT scores ( $P=.15$ ). Thus, as determined by MMSE scores, cognitive performance was improved in the Swedish study population but not in the Italian study population.

### *Quality of Life*

In Italy, there was no difference between the 6-month follow-up and baseline in any of the EQ-5D-5L subscales (mobility,  $P=.41$ ; self-care,  $P=.41$ ; activity,  $P=.58$ ; pain,  $P=.16$ ; anxiety,  $P=.59$ ) and the EQ-5D-5L visual analog scale (VAS),  $P=.53$ . In Sweden, there were improvements in the mean (SD) changes in the EQ-5D-5L subscale anxiety and the EQ-5D-5L VAS scale, which were 0.16 (0.54) and 2.6 (9.7), respectively. The 6-month follow-up values in these variables were also significantly different from those at baseline ( $P=.046$  and  $P=.035$ , respectively), confirming beneficial effects in these variables in the Swedish cohort. Other EQ-5D-5L subscales were unchanged in Sweden: mobility,  $P=.20$ ; self-care,  $P=.16$ ; activity,  $P=.20$ ; and pain,  $P=.26$ . However, when comparing the results in Italy versus those in Sweden, there were no between-group differences. Therefore, although some improvement was observed in Sweden, there was no difference in the effect on quality of life between the countries.

### *Patient Satisfaction*

In Italy, 10 patients were interviewed. The patients perceived the DECI service as simple and as containing engaging exercises, which helped the patients become more physically and cognitively active. The patients were willing to try more advanced exercises matching their physical condition, and it was deemed that a customized exercise program could increase the motivation of the patients. Some patients struggled to navigate the technologies and sought help from health care professionals or family members. Like in Sweden, the help-desk function could help to reduce the load on clinicians for solving technical questions or problems. The activity monitoring watch was appreciated by the patients, but its design could be improved in order to meet the aesthetic standards of the elderly.

In Sweden, 10 patients were interviewed. The DECI service was appreciated by the patients due to the clinician-monitored

exercise possibilities at home and multiple home visits by health care professionals. The patients expressed a willingness to continue using the physical and cognitive activity programs. Previous information technology (IT) experience was determined to be helpful in navigating the tablet. Less experienced patients relied on family members for help, while others called health care professionals. Therefore, an IT help desk that helps to solve tablet-related issues is a necessary element of the service. The activity monitoring watch did not meet patients' expectations without a pedometer and a display.

### *Job Satisfaction*

In Italy, 4 health care professionals were interviewed. Their occupations were a geriatrician, a social worker, a neuropsychologist, and a clinical neuropsychologist. The health care professionals felt enabled to build relationships with patients through a dedicated case manager. The physical and cognitive activity coaching programs gave the professionals tools that could not be found in usual care practices. The exercises motivated and engaged the patients, which positively contributed to the job satisfaction of the professionals. The ability to remotely monitor patients' performances and adherence to the tools was seen as an additional value. The professionals also utilized the messaging and data sharing functions in the DECI platform, which facilitated the interdisciplinary work.

In Sweden, 2 health care professionals were interviewed. Their occupations were a geriatrician and a nurse. The professionals were satisfied that the DECI services helped to form positive relationships with the patients while visiting them in their home environments. The case manager's role was perceived as rewarding. Professionals could deepen their knowledge of the patients when observing physical status in their usual environment. The professionals felt empowered since they could offer digital tools to the patients with beneficial, customized exercises for different muscle groups and cognitive conditions. However, the messaging function in the DECI platform was perceived as having low value, since the mobile team was mostly on the road and used mobile phones for communication.

## Monetary Sacrifices

### *Investment*

In Italy, the one-time investment concerned the cost of staff for server preparation in the hospital. In Sweden, the investment concerned the cost of staff for server preparation and installation of the DECI technologies. Additionally, investments in Sweden also entailed one-time costs for purchasing tablets, including 4G-sim cards for training for patients ( $n=52$ ) and members of the DECI-team ( $n=2$ ).

### *Operating Expenses*

In both Italy and Sweden, the highest usage fee concerned the ADAMO activity monitoring device. The annual operating costs in Italy include purchasing the tablets (a stable number of tablets purchased for patients every year, depreciation in 12 months), server hosting, the configuration of the tablets, the personnel cost of maintenance and the help desk, licenses, and 4G connectivity. In Sweden, it consists of the server hosting,

first-line help desk, and management costs. Operating expenses are depicted in [Table 3](#).

**Table 3.** Monetary sacrifices, in euros (a currency exchange rate of EUR €1=US \$1.18 is applicable).

Monetary sacrifices in Italy and Sweden	Year 1 (€)	Year 2 (€)	Year 3 (€)
<b>Cost of spent time - Italy</b>			
Total	71.161	106.743	124.533
Per patient	711	711	711
<b>Cost of spent time - Sweden</b>			
Total	693.773	969.160	1.130.687
Per patient	1387	1387	1387
<b>Investment - Italy</b>			
Total	140	0	0
Per patient	1.4	0	0
<b>Investment - Sweden</b>			
Total	18.636	18.636	18.636
Per patient	37	24	21
<b>Operating expenses - Italy</b>			
Total	126.746	171.896	194.471
Per patient	1267	1145	1111
<b>Operating expenses - Sweden</b>			
Total	348.015	499.965	575.940
Per patient	696	666	658

### Cost of Spent Time

The cost was calculated based on the hours spent on the DECI service provision and the multi-disciplinary meetings ([Table 3](#)). The following professional categories were involved in Italy: physician, nurse practitioner, physiotherapist, technician, case manager (social worker), and psychologist. In Sweden, the cost was calculated for a geriatrician, an occupational therapist, a physiotherapist, and a nurse practitioner.

[Table 3](#) shows a summary of the monetary sacrifices per patient in Italy and Sweden for 3 years. The large differences in monetary values between Italy and Sweden occurred due to the different targeted population sizes and care models.

### Nonmonetary Sacrifices

#### Patient Safety

In Italy, the ADAMO activity monitoring device caused an allergic reaction in 1 patient, due to sensitivity to nickel and plastic. The health care professionals in Italy (n=4, 1 geriatrician, 1 social worker, 1 neuropsychologist, and 1 clinical neuropsychologist) and Sweden (n=2, 1 geriatrician and 1 nurse) noted that the physical activity training program Otago could

be less safe for older people if used without supervision. Hence, the level of exercise difficulty is of high importance.

#### Workload

In Italy, the 4 health care professionals dedicated, on average, 29.3 hours a week to engage with the existing patients at the memory clinic that took part in the DECI study (the result is based on their self-reported data). A substantial amount of this time was dedicated to digital data entry for the DECI study. This time was used for patient inclusion, training, phone calls, solving technical issues, and digital data entry for the study.

In Sweden, the 2 health care professionals spent, on average, 52.5 hours per week, thus reporting overtime. This time was used for the full-scale dementia examination, training, phone calls, home visits to the patients, solving technical issues, and digital data entry for the study. The result is based on their self-reported data.

#### Monetary Benefit/Sacrifice Ratio

[Table 4](#) depicts a summary of monetary benefits and sacrifices, and provides a calculation of the benefit/sacrifice ratio in Italy and Sweden for 3 years.

**Table 4.** Summary of the DECI scenario in monetary value, in euros (a currency exchange rate of EUR €=US \$1.18 is applicable).

Benefits and sacrifices in Italy and Sweden	Year 1	Year 2	Year 3
<b>Total monetary benefits – Italy (€)</b>			
Total	259.633	389.450	454.359
Per patient	2596	2596	2596
<b>Total monetary benefits – Sweden (€)</b>			
Total	833.037	1.249.555	1.457.815
Per patient	1666	1666	1666
<b>Total monetary sacrifices – Italy (€)</b>			
Total	198.047	278.639	319.004
Per patient	1980	1857	1822
<b>Total monetary sacrifices – Sweden (€)</b>			
Total	1.060.424	1.487.761	1.725.263
Per patient	2120	1983	1971
Benefit/sacrifice ratio - Italy	1.31	1.39	1.42
Benefit/ sacrifice ratio - Sweden	0.78	0.84	0.84

The benefit/sacrifice ratio showed that the Italian intervention could bring positive monetary value from the first year onward. In Sweden, the intervention did not bring monetary value during the first 3 years. However, the gap between monetary benefits and sacrifices reduces with the growing number of patients.

## Discussion

### Principal Findings

This study was guided by 2 objectives: (1) to identify the contextual factors that determine the similarities and differences in the value of an eHealth intervention between 2 contexts, and (2) to reflect on and contribute to the discussion about the specification, assessment, and the relativity of the “value” concept in evaluating eHealth interventions. This study was based on the implementation of an eHealth platform for remote home monitoring of physical and cognitive activity for people suffering from cognitive impairment in Italy and Sweden.

The findings of this study show that there is a differing value derived from the implementation of the same eHealth technology to similar types of populations in different contexts.

We translated value into benefits and sacrifices and assessed these for the intervention in 2 countries. In Sweden, the identified benefits of the eHealth intervention included improved cognitive performance assessed by the MMSE, reduced patient anxiety assessed by the EQ-5D-5L, better perceived health estimated using the EQ-5D-5L VAS scale, and satisfaction with the care received. However, these benefits can require sacrifices, such as an increased cost and higher workload for health care professionals. With the service model of Sweden (home visits), a lower number of patients could be visited per day. Additionally, the relatively high hourly rates of the staff increased the cost of the intervention. In Italy, the identified benefits included patient satisfaction with the care received, empowered health care professionals, and acceptable workloads. Moreover, for the Italian patients, who were older and had a

more marked decrease in cognitive function (lower MMSE and CDT scores) at baseline than the Swedish patients, the intervention could bring positive monetary value from the first year onward. This was the result of a higher preventable cost of treatment and state reimbursement, in comparison to the cost that was based on relatively lower hourly staff rates and the service model that reduced the time spent per patient (the intervention was implemented on a sample of existing patients of the clinic and the visits were performed at the clinic). However, the clinical efficacy and quality-of-life improvements have not been observed over the course of the 6-month intervention in Italy.

In this study, the monetary side of value was influenced by factors such as (1) the service delivery design of the intervention (process of delivery), (2) the organizational setup of the intervention (ie, organizational structure and professionals involved), (3) the cost of different treatments, and (4) the hourly rates of staff for delivering the intervention. These factors affected the cost-effectiveness through the expenses incurred (including potentially preventable costs due to the intervention) and necessary investments. The nonmonetary side of the value of the intervention was also influenced by the service delivery design and organizational setup of the intervention, as well as by a fifth factor: (5) the lifestyle habits of the population (eg, how physically active they were in their daily lives and if they were living alone or with family). Finally, the value of eHealth should be seen against the sixth factor: (6) local preferences on the quality of patient care. This study showed that even the non-cost-efficient intervention can be viewed as valuable locally and deemed worthy of implementation. In such a case, local preferences on the quality of patient care can be a decisive factor. Particular positive nonmonetary outcomes might be valued highly enough to proceed with adopting the eHealth-supported service model. Moreover, it should not be neglected that the service delivery design and organizational



set-up of the intervention can be adjusted to make it more cost-efficient.

A summary of the contextual factors affecting the value of eHealth intervention is provided in [Textbox 3](#).

**Textbox 3.** Conceptual model for eHealth value specification and the influencing contextual factors.

<p><b>Contextual factors (benefits)</b></p> <ul style="list-style-type: none"> <li>• <i>Monetary</i> <ul style="list-style-type: none"> <li>• Service delivery design of an intervention (process)</li> <li>• Organizational setup of an intervention (structure)</li> <li>• Cost of different treatments</li> <li>• Hourly rates of staff for delivering an intervention</li> </ul> </li> <li>• <i>Nonmonetary</i> <ul style="list-style-type: none"> <li>• Service delivery design of an intervention (process)</li> <li>• Organizational setup of an intervention (structure)</li> <li>• Lifestyle habits of the population</li> <li>• Local preferences on the quality of patient care</li> </ul> </li> </ul> <p><b>Contextual factors (sacrifices)</b></p> <ul style="list-style-type: none"> <li>• <i>Monetary</i> <ul style="list-style-type: none"> <li>• Service delivery design of an intervention (process)</li> <li>• Organizational setup of an intervention (structure)</li> <li>• Cost of different treatments</li> <li>• Hourly rates of staff for delivering an intervention</li> </ul> </li> <li>• <i>Nonmonetary</i> <ul style="list-style-type: none"> <li>• Service delivery design of an intervention (process)</li> <li>• Organizational setup of an intervention (structure)</li> </ul> </li> </ul>
---

## Limitations

The identified contextual factors affecting the value of eHealth interventions could be limited because the study was based on a summative eHealth evaluation conducted in 2 countries. To study 2 countries in-depth demanded time and effort, and including more countries was not feasible. Other contextual setups could enrich the list of the factors and need to be investigated further. Also, the study was constrained by a 6-month follow-up time for the patients. A longer follow-up time could enrich the contextual factors and the proposed eHealth value conceptualization by revealing long-term effects.

## Comparison with Prior Work

The conceptual model proposed for assessing the value of eHealth interventions ([Textbox 3](#)) was built on prominent eHealth evaluation frameworks [32,33], in addition to previous conceptualizations of value in the eHealth [4-8] and marketing literature [10]. We argue that value needs to be operationalized in both monetary and nonmonetary outcomes, and our model suggests categorizing them into benefits and sacrifices. In practice, an evaluation study on value needs to adapt the model to its needs by populating the model with the themes of evaluation. It is important not to overlook the nonmonetary

aspects that can reveal a broader and more accurate view of the value created (in contrast to the cost-versus-outcomes view [1,8,34,35]). We propose that all the parts of the model (monetary and nonmonetary benefits/sacrifices) need to be assessed in order to obtain a holistic view of the value created.

Regarding the conceptualization of value for eHealth interventions, our study showed that an overly limited view on value is obtained if assessing it as only positive outcomes, such as behavior change or clinical efficacy [4,5], perceptions regarding the added value by various stakeholders [6,7], or economic outcomes [8]. The view of value as clinical efficacy [4] is sometimes not possible when studies have a shorter follow-up time. Furthermore, this study showed that economic outcomes might be an overly limited measure of value, when the non-cost-efficient intervention can be viewed as valuable locally for other positive outcomes and deemed worthy for implementation in practice despite the increased cost. We suggest the following conceptualization of value for eHealth interventions:

*Value is a holistic view of the created monetary and nonmonetary benefits of eHealth that require monetary and nonmonetary sacrifices in a particular context.*

Finally, we argue that more studies need to apply a holistic view to the summative assessments of value, rather than focusing on defining value for the stakeholders during the design process [36]. This holistic value assessment could help to clarify which adjustments could be made to reduce sacrifices and maximize benefits. Moreover, such assessments (especially across multiple contexts) could improve the transferability of eHealth interventions.

## Conclusions

This study offers a step toward better conceptualization of the value concept within the eHealth context. We argue that it should be interpreted as both monetary and nonmonetary benefits and sacrifices achieved in a context. Next, we argue that without considering value beyond economic terms or assessing it as merely perceptions, a full picture of the value created might be missed. We offer a model that provides some conceptual considerations for the assessment of value of eHealth. Applied within a summative evaluation in this cross-context study, it can be a useful starting point for future research.

## Acknowledgments

The collection of data used in this study was financed by the European Union's Horizon 2020 research and innovation program under grant agreement no. 643588. The funding organization had no role in designing the study, collecting and analyzing the data, or in preparing this manuscript.

## Conflicts of Interest

LvV works at one of the enterprises that supplied an eHealth solution to the DECI project and the current study.

### Multimedia Appendix 1

Patient interview protocol.

[DOCX File, 17 KB - [jmir\\_v22i10e17720\\_app1.docx](#)]

### Multimedia Appendix 2

Healthcare professional interview protocol.

[DOCX File, 17 KB - [jmir\\_v22i10e17720\\_app2.docx](#)]

## References

1. Porter M, Teisberg E. Redefining Competition in Health Care. Boston: Harvard Business School Press; 2006.
2. Gray M. Value: Operations Research and the new health care paradigm. *Oper Res Health Care* 2012;1:20-21 [FREE Full text] [doi: [10.1016/j.orhc.2012.01.002](#)]
3. Colldén C, Gremyr I, Hellström A, Sporraeus D. A value-based taxonomy of improvement approaches in healthcare. *J Health Organ Manag* 2017 Jun 19;31(4):445-458. [doi: [10.1108/jhom-08-2016-0162](#)]
4. Ahern DK, Kreslake JM, Phalen JM. What is eHealth (6): perspectives on the evolution of eHealth research. *J Med Internet Res* 2006 Mar;8(1):e4 [FREE Full text] [doi: [10.2196/jmir.8.1.e4](#)] [Medline: [16585029](#)]
5. Talboom-Kamp EP, Verdijk NA, Harmans LM, Numans ME, Chavannes NH. An eHealth Platform to Manage Chronic Disease in Primary Care: An Innovative Approach. *Interact J Med Res* 2016 Feb 09;5(1):e5. [doi: [10.2196/ijmr.4217](#)]
6. Runz-Jørgensen SM, Schiøtz ML, Christensen U. Perceived value of eHealth among people living with multimorbidity: a qualitative study. *J Comorb* 2017;7(1):96-111 [FREE Full text] [doi: [10.15256/joc.2017.7.98](#)] [Medline: [29359124](#)]
7. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and Facilitators to eHealth Use in Daily Practice: Perspectives of Patients and Professionals in Dermatology. *J Med Internet Res* 2017 Sep 05;19(9):e300 [FREE Full text] [doi: [10.2196/jmir.7512](#)] [Medline: [28874336](#)]
8. Bergmo TS. How to Measure Costs and Benefits of eHealth Interventions: An Overview of Methods and Frameworks. *J Med Internet Res* 2015 Nov;17(11):e254 [FREE Full text] [doi: [10.2196/jmir.4521](#)] [Medline: [26552360](#)]
9. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *J Med Internet Res* 2017 Nov 01;19(11):e367 [FREE Full text] [doi: [10.2196/jmir.8775](#)] [Medline: [29092808](#)]
10. Gummerus J. Value creation processes and value outcomes in marketing theory. *Mark Theory* 2013 Jan 24;13(1):19-46. [doi: [10.1177/1470593112467267](#)]
11. Ulaga W, Chacour S. Measuring Customer-Perceived Value in Business Markets. *Ind Mark Manag* 2001 Aug;30(6):525-540. [doi: [10.1016/s0019-8501\(99\)00122-4](#)]
12. Christie HL, Bartels SL, Boots LM, Tange HJ, Verhey FR, de Vugt ME. A systematic review on the implementation of eHealth interventions for informal caregivers of people with dementia. *Internet Interv* 2018 Sep;13:51-59. [doi: [10.1016/j.invent.2018.07.002](#)]

13. Eslami Andargoli A, Scheepers H, Rajendran D, Sohal A. Health information systems evaluation frameworks: A systematic review. *Int J Med Inform* 2017 Jan;97:195-209. [doi: [10.1016/j.ijmedinf.2016.10.008](https://doi.org/10.1016/j.ijmedinf.2016.10.008)]
14. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. *Int J Med Inform* 2014 Jul;83(7):e1-e11. [doi: [10.1016/j.ijmedinf.2013.07.002](https://doi.org/10.1016/j.ijmedinf.2013.07.002)]
15. Edwards N, Barker PM. The Importance of Context in Implementation Research. *J Acquir Immune Defic Syndr* 2014;67:S157-S162. [doi: [10.1097/qai.0000000000000322](https://doi.org/10.1097/qai.0000000000000322)]
16. Vernooij-Dassen M, Moniz-Cook E. Raising the standard of applied dementia care research: addressing the implementation error. *Aging Ment Health* 2014 May 12;18(7):809-814. [doi: [10.1080/13607863.2014.899977](https://doi.org/10.1080/13607863.2014.899977)]
17. Ammenwerth E, Brender J, Nykänen P, Prokosch H, Rigby M, Talmon J. Visions and strategies to improve evaluation of health information systems. In: *International Journal of Medical Informatics*. 2004 Jun Presented at: Visions and strategies to improve evaluation of health information systems: Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. *Int J Med Inform* ;73(6); 2004; Innsbruck p. 479-491. [doi: [10.1016/j.ijmedinf.2004.04.004](https://doi.org/10.1016/j.ijmedinf.2004.04.004)]
18. Brady MK, Robertson CJ. An exploratory study of service value in the USA and Ecuador. *Int J Serv Ind Manag* 1999 Dec;10(5):469-486. [doi: [10.1108/09564239910289003](https://doi.org/10.1108/09564239910289003)]
19. Campbell AJ, Robertson MC, Gardner MM, Norton RN, Tilyard MW, Buchner DM. Randomised controlled trial of a general practice programme of home based exercise to prevent falls in elderly women. *BMJ* 1997 Oct 25;315(7115):1065-1069. [doi: [10.1136/bmj.315.7115.1065](https://doi.org/10.1136/bmj.315.7115.1065)]
20. Dekker-van Weering M, Jansen-Kosterink S, Frazer S, Vollenbroek-Hutten M. User Experience, Actual Use, and Effectiveness of an Information Communication Technology-Supported Home Exercise Program for Pre-Frail Older Adults. *Front. Med* 2017 Nov 27;4:208. [doi: [10.3389/fmed.2017.00208](https://doi.org/10.3389/fmed.2017.00208)]
21. Tárraga L, Boada M, Modinos G, Espinosa A, Diego S, Morera A, et al. A randomised pilot study to assess the efficacy of an interactive, multimedia tool of cognitive stimulation in Alzheimer's disease. *J Neurol Neurosurg Psychiatry* 2006 Oct;77(10):1116-1121 [FREE Full text] [doi: [10.1136/jnnp.2005.086074](https://doi.org/10.1136/jnnp.2005.086074)] [Medline: [16820420](https://pubmed.ncbi.nlm.nih.gov/16820420/)]
22. Magistro D, Brustio PR, Ivaldi M, Eslinger DW, Zecca M, Rainoldi A, et al. Validation of the ADAMO Care Watch for step counting in older adults. *PLoS ONE* 2018 Feb 9;13(2):e0190753. [doi: [10.1371/journal.pone.0190753](https://doi.org/10.1371/journal.pone.0190753)]
23. Morris JC. Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. *Int. Psychogeriatr* 2005 Jan 10;9(S1):173-176. [doi: [10.1017/s1041610297004870](https://doi.org/10.1017/s1041610297004870)]
24. Gillespie L, Robertson M, Gillespie W, Sherrington C, Gates S, Clemson L, et al. Interventions for preventing falls in older people living in the community. *Cochrane database syst rev* 2012;9:CD007146. [doi: [10.1002/14651858.cd007146.pub3](https://doi.org/10.1002/14651858.cd007146.pub3)]
25. Berg WP, Alessio HM, Mills EM, Tong C. Circumstances and consequences of falls in independent community-dwelling older adults. *Age Ageing* 1997;26(4):261-268. [doi: [10.1093/ageing/26.4.261](https://doi.org/10.1093/ageing/26.4.261)]
26. Thomas S, Mackintosh S, Halbert J. Does the 'Otago exercise programme' reduce mortality and falls in older adults?: a systematic review and meta-analysis. *Age Ageing* 2010 Sep 04;39(6):681-687. [doi: [10.1093/ageing/afq102](https://doi.org/10.1093/ageing/afq102)]
27. Folstein MF, Folstein SE, McHugh PR. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *Int. J. Geriatr. Psychiatry* 1975;12(3):189-198. [doi: [10.1002/\(sici\)1099-1166\(199805\)13:5<285::aid-gps753>3.0.co;2-v](https://doi.org/10.1002/(sici)1099-1166(199805)13:5<285::aid-gps753>3.0.co;2-v)]
28. Shulman KI. Clock-drawing: is it the ideal cognitive screening test? *Int J Geriatr Psychiatry* 2000 Jun;15(6):548-561. [doi: [10.1002/1099-1166\(200006\)15:6<548::aid-gps242>3.0.co;2-u](https://doi.org/10.1002/1099-1166(200006)15:6<548::aid-gps242>3.0.co;2-u)]
29. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 1990 Dec;16(3):199-208. [doi: [10.1016/0168-8510\(90\)90421-9](https://doi.org/10.1016/0168-8510(90)90421-9)]
30. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Adm Policy Ment Health* 2015 Sep;42(5):533-544 [FREE Full text] [doi: [10.1007/s10488-013-0528-y](https://doi.org/10.1007/s10488-013-0528-y)] [Medline: [24193818](https://pubmed.ncbi.nlm.nih.gov/24193818/)]
31. Aronson J. A pragmatic view of thematic analysis. *Qual Rep*(1). 1995. URL: <https://nsuworks.nova.edu/tqr/vol2/iss1/3/> [accessed 2019-12-28]
32. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. *Int J Technol Assess Health Care* 2012 Jan;28(1):44-51. [doi: [10.1017/S0266462311000638](https://doi.org/10.1017/S0266462311000638)] [Medline: [22617736](https://pubmed.ncbi.nlm.nih.gov/22617736/)]
33. Lampe K, Mäkelä M, Garrido MV, Anttila H, Autti-Rämö I, Hicks NJ, et al. The HTA Core Model: A novel method for producing and reporting health technology assessments. *Int J Technol Assess Health Care* 2009 Dec 23;25(S2):9-20. [doi: [10.1017/s0266462309990638](https://doi.org/10.1017/s0266462309990638)]
34. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. *Health Aff* 2014 Jul;33(7):1115-1122. [doi: [10.1377/hlthaff.2014.0147](https://doi.org/10.1377/hlthaff.2014.0147)] [Medline: [25006136](https://pubmed.ncbi.nlm.nih.gov/25006136/)]
35. Porter ME. What Is Value in Health Care? *N Engl J Med* 2010 Dec 23;363(26):2477-2481. [doi: [10.1056/NEJMp1011024](https://doi.org/10.1056/NEJMp1011024)]
36. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. *J Med Internet Res* 2011 Dec;13(4):e111 [FREE Full text] [doi: [10.2196/jmir.1672](https://doi.org/10.2196/jmir.1672)] [Medline: [22155738](https://pubmed.ncbi.nlm.nih.gov/22155738/)]

## Abbreviations

**CDR:** clinical dementia rating  
**CDT:** clock drawing test  
**DECI:** Digital Environment for Cognitive Inclusion  
**FDG:** Fondazione Don Gnocchi  
**GP:** general practitioner  
**IP:** Istituto Palazzolo  
**IT:** information technology  
**MCI:** mild cognitive impairment  
**MD:** mild dementia  
**MMSE:** Mini-Mental State Examination

*Edited by G Eysenbach; submitted 07.01.20; peer-reviewed by E Aramaki, F Adebesein; comments to author 10.03.20; revised version received 17.05.20; accepted 14.06.20; published 08.10.20.*

*Please cite as:*

*Jurkeviciute M, van Velsen L, Eriksson H, Lifvergren S, Trimarchi PD, Andin U, Svensson J*

*Identifying the Value of an eHealth Intervention Aimed at Cognitive Impairments: Observational Study in Different Contexts and Service Models*

*J Med Internet Res* 2020;22(10):e17720

URL: <http://www.jmir.org/2020/10/e17720/>

doi: [10.2196/17720](https://doi.org/10.2196/17720)

PMID: [33064089](https://pubmed.ncbi.nlm.nih.gov/33064089/)

©Monika Jurkeviciute, Lex van Velsen, Henrik Eriksson, Svante Lifvergren, Pietro Davide Trimarchi, Ulla Andin, Johan Svensson. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 11.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Efficacy of Electronic Acupuncture Shoes for Chronic Low Back Pain: Double-Blinded Randomized Controlled Trial

Bo-Yan Yeh<sup>1</sup>, MD; Geng-Hao Liu<sup>1,2,3</sup>, MD, PhD; Tzung-Yan Lee<sup>4,5</sup>, PhD; Alice May-Kuen Wong<sup>6</sup>, MD; Hen-Hong Chang<sup>7,8\*</sup>, MD, PhD; Yu-Sheng Chen<sup>1,9\*</sup>, MD, PhD

<sup>1</sup>Division of Acupuncture and Moxibustion, Department of Traditional Chinese Medicine, Chang Gung Memorial Hospital, Linkou, Taoyuan, Taiwan

<sup>2</sup>School of Traditional Chinese Medicine, Chang Gung University, Taoyuan, Taiwan

<sup>3</sup>Sleep Center, Chang Gung Memorial Hospital, Taoyuan, Taiwan

<sup>4</sup>Graduate Institute of Traditional Chinese Medicine, School of Chinese Medicine, College of Medicine, Chang Gung University, Taoyuan, Taiwan

<sup>5</sup>Department of Traditional Chinese Medicine, Chang Gung Memorial Hospital, Keelung, Taiwan

<sup>6</sup>Department of Physical Medicine and Rehabilitation, Chang Gung Memorial Hospital, Taoyuan, Taiwan

<sup>7</sup>Graduate Institute of Integrated Medicine, College of Chinese Medicine, and Chinese Medicine Research Center, China Medical University, Taichung, Taiwan

<sup>8</sup>Department of Chinese Medicine, China Medical University Hospital, Taichung, Taiwan

<sup>9</sup>Graduate Institute of Clinical Medical Sciences, Chang Gung University, Taoyuan, Taiwan

\*these authors contributed equally

**Corresponding Author:**

Yu-Sheng Chen, MD, PhD

Division of Acupuncture and Moxibustion, Department of Traditional Chinese Medicine

Chang Gung Memorial Hospital, Linkou

No.123, Dinghu Rd., Guishan Township

Taoyuan, 33378

Taiwan

Phone: 886 33196200 ext 2613

Email: [cusmp01@cgmh.org.tw](mailto:cusmp01@cgmh.org.tw)

## Abstract

**Background:** Chronic low back pain is a common problem and is associated with high costs, including those related to health care and indirect costs due to absence at work or reduced productivity. Previous studies have demonstrated that acupuncture or electroacupuncture can relieve low back pain. Electronic acupuncture shoes (EAS) are a novel device designed in this study. This device combines the properties of acupuncture and transcutaneous electrical nerve stimulation for clinical use.

**Objective:** The aim of this study was to evaluate the efficacy of EAS in patients with chronic low back pain.

**Methods:** In this prospective double-blinded randomized controlled study, the data of 83 patients who experienced chronic low back pain were analyzed. Patients came to our clinic for 20 visits and underwent assessment and treatment. Patients were randomly allocated to receive either EAS plus placebo nonsteroidal anti-inflammatory drugs (NSAIDs) (EAS group, n=42) or sham EAS plus NSAIDs (NSAID group, n=41). The visual analog scale (VAS) score and range of motion were assessed at baseline, before and after each EAS treatment, and 2 weeks after the last treatment. The time for achieving pain remission was recorded. Quality of life was assessed at the 2nd, 14th, and 20th visits.

**Results:** After 6 weeks of treatment, the treatment success rate in each visit in the EAS group was higher than that in the NSAID group, as revealed by the intention-to-treat (ITT) and per-protocol (PP) analyses, but significant differences were observed only during the 16th visit in the ITT analysis (EAS group: 31/37, 84% and NSAID group: 21/34, 62%;  $P=.04$ ). The change in the VAS score from baseline in each visit in the EAS group was greater than that in the NSAID group, as revealed by the ITT and PP analyses, and significant differences were observed in the 5th visit and 9th visit in the ITT analysis ( $P=.048$  and  $P=.048$ , respectively). Significant differences were observed in the left rotation in the 2nd visit and 4th visit ( $P=.049$  and  $P=.03$ , respectively). No significant differences were observed in the VAS score before and after treatment in each visit and in the quality of life in both groups.

**Conclusions:** EAS might serve as a reliable alternative therapeutic tool for patients with chronic low back pain who are contraindicated for oral NSAIDs.



**Trial Registration:** ClinicalTrials.gov NCT02468297 <https://clinicaltrials.gov/ct2/show/NCT02468297>

(*J Med Internet Res* 2020;22(10):e22324) doi:[10.2196/22324](https://doi.org/10.2196/22324)

## KEYWORDS

acupuncture; electronic acupuncture shoes; low back pain; medical device; self-treatment; mHealth

## Introduction

Low back pain (LBP) is a common problem worldwide, and its prevalence ranges from 22% to 48% [1]. The lifetime prevalence of LBP is 84% [2]. LBP is associated with high costs, including those related to health care and indirect costs from missed work or reduced productivity [1]. Acupuncture is a cost-effective treatment strategy for chronic LBP (CLBP) [3]. CLBP is commonly defined as back pain that persists for at least 12 weeks. In clinics, physicians usually prescribe nonsteroidal anti-inflammatory drugs (NSAIDs) to relieve pain. However, these NSAIDs cause side effects such as nausea, peptic ulcer, gastrointestinal bleeding, and elevated blood pressure.

Lee et al [4] proposed that acupuncture is a simple and effective strategy for relieving pain, but it cannot improve the loss of function and disability resulting from LBP. The guidelines of the American College of Physicians strongly recommend acupuncture for the selective nonpharmacologic treatment of CLBP [1]. The mechanism of acupuncture analgesia is associated with central neurotransmitters, immune cytokines, and cytokines from the spinal glial cells [5]. These substances can produce various effects such as analgesic, muscle relaxation, anti-inflammatory, mild anxiolytic, and antidepressant effects [6].

Electroacupuncture is defined as applying electrical stimulation to acupuncture needles [6]. This strategy may result in a faster analgesic and anesthetic effect, and high-frequency electroacupuncture has been reported to control pain more effectively than low-frequency electroacupuncture [7]. Electronic acupuncture shoes (EAS) are a newly designed device; these shoes show a combination of the properties of electroacupuncture and transcutaneous electrical nerve stimulation (TENS) and employ a pain-controlling mechanism different from that of acupoint TENS [8-10].

Electronic systems have been incorporated as a component in many medical devices. For example, the electronic system in the smart assistive knee brace serves as a driving force through heating and cooling on the shape memory alloy to induce a cycle of contraction and elongation [11]. However, in the EAS, the electronic system plays the key role in the therapeutic effect by applying the current loop between KII and *shimian* to induce reactions in the human body through the central [12,13] and peripheral pathways [14-16], respectively. The electronic system is incorporated into the shoes with an appropriate parameter setting such that the location of the electrical stimulation is accurate. If a person contracts a new type of infectious disease such as COVID-19, EAS can serve as a very good modality for home health care and can provide therapeutic effects immediately with few contraindications. This device can be used for patients who fear acupuncture needles, and it can be

manipulated by patients themselves. Our study compared the noninferiority effects of EAS with NSAIDs for the treatment of patients with CLBP.

## Methods

### Patient Selection

This study was approved by the Institutional Review Board of the Chang Gung Medical Foundation, and this trial was registered at the ClinicalTrials.gov website (NCT02468297). We registered this trial retrospectively because of the following reasons. We began to enroll patients in April 2004 but stopped in November 2008. We needed to ensure that the amplitude of the electric current was around an appropriate range. Thus, the trial protocol was revised for performing a test to determine the resistance of the sole of every participant. After the revision of the protocol, we re-enrolled patients in April 2009. The manufacturer of EAS did not want to disclose the trade secret before the trial had been completed.

Informed consent was obtained from each participant. A total of 90 patients of both sexes were prospectively screened for study participation. All patients were outpatients of the Department of Rehabilitation, Orthopedics, and Chinese Medicine. We selected patients with the following inclusion criteria: diagnosis of CLBP, that is, the location of pain was below the 12th rib and above the horizontal gluteal crease and lasted for more than 12 weeks; age range of 20 to 60 years; and provision of signed informed consent. Patients were excluded if he or she met the following medical conditions, including cancer, rheumatoid arthritis, renal stone, diabetes mellitus, pacemaker implantation, under steroid treatment, fracture or surgical history of the back, spinal cord compression syndrome (eg, herniated intervertebral disc or spinal cord disorder), visceral organ infectious disease (eg, pancreatitis and pyelonephritis), or visual coordination disorder. Women who experienced menopause before 50 years of age, underwent ovary excision, or were pregnant were excluded. Further, women considering to be pregnant were asked to not enroll, and enrolled women were asked to agree to contraception or abstinence. Moreover, before joining the trial, their pregnancy test must be negative. Patients were also excluded if they were not free from previous participation in other trials within 30 days before joining this trial, showed contraindications to ibuprofen, or if they had poor heart, liver, gastrointestinal tract, and renal function. Further, if the physician suspected that EAS treatment might have adverse effects on the patient based on the physical examination and laboratory data at the first visit, the patient was not enrolled.

### Study Setting

All patients were randomly assigned to either EAS group (EAS plus placebo NSAIDs) or NSAID group (sham EAS plus

NSAIDs) based on a concealed allocation approach. A computerized random number table was used to determine group allocation. Numbered opaque sealed envelopes contained ibuprofen or placebo analgesics and numbered EAS, and no restrictions were placed on randomization. The envelopes for the analgesics and EAS were kept by a pharmacist who was not an assessor of the study. All assessors and patients were blinded to group allocation.

### EAS and NSAID settings

In the first visit (first week), patients were assessed by physical examination, vital sign evaluation, blood test (complete blood count, clinical chemistry, and pregnancy test), radiography of the lumbar spine in the anterior-posterior and lateral views, deep tendon reflex test, and sensation, range of motion, and pain scales. Subsequently, patients received EAS treatment (EAS group) or placebo EAS (NSAID group) for 1 hour 3 times a week for 6 weeks. Patients were assessed at baseline (first visit), before and after each EAS treatment (2nd-19th visit), and 2 weeks after the last treatment (20th visit).

### Experimental Group

In the first week of treatment, patients in the EAS group received placebo analgesics. Before EAS treatment, patients were asked to remove any conducting metal accessories, watch, mobile, and socks. Each patient had their own EAS.

During treatment, if the patient experienced any discomfort, indicated by symptoms such as extreme LBP, spasm in both lower limbs, paralysis, tachycardia, or dizziness, the assessors recorded the symptoms and their duration. If the symptoms were slight, EAS treatment was ceased for 5-10 minutes and then continued after the symptoms were relieved. If the symptoms were severe, EAS treatment was stopped at once, and the assessors recorded the reasons for stopping the management and the treatment duration. After each treatment, the patient rested for 3-5 minutes, relaxed the joints of the lower limbs, and then underwent the efficacy assessment.

### Control Group

In the first week of treatment, patients in the NSAID group received 400 mg ibuprofen 3 times a day. Subsequently, they received EAS treatment without electric current 3 times a week for 6 weeks, as in the EAS group.

### Outcome Measurements and Follow-Up

#### *Primary Outcome Measurement: Pain Intensity*

Pain intensity was assessed using the visual analog scale (VAS). The VAS is an 11-point scale ranging from 0 to 10. A VAS score of 0 implies absence of pain, and a score of 10 implies unbearable pain. Patients rated their pain levels before and after each EAS treatment (2nd-19th visit); otherwise, pain levels were assessed at the first visit and at 2 weeks after the last treatment (20th visit) [17].

#### *Secondary Outcome Measurements*

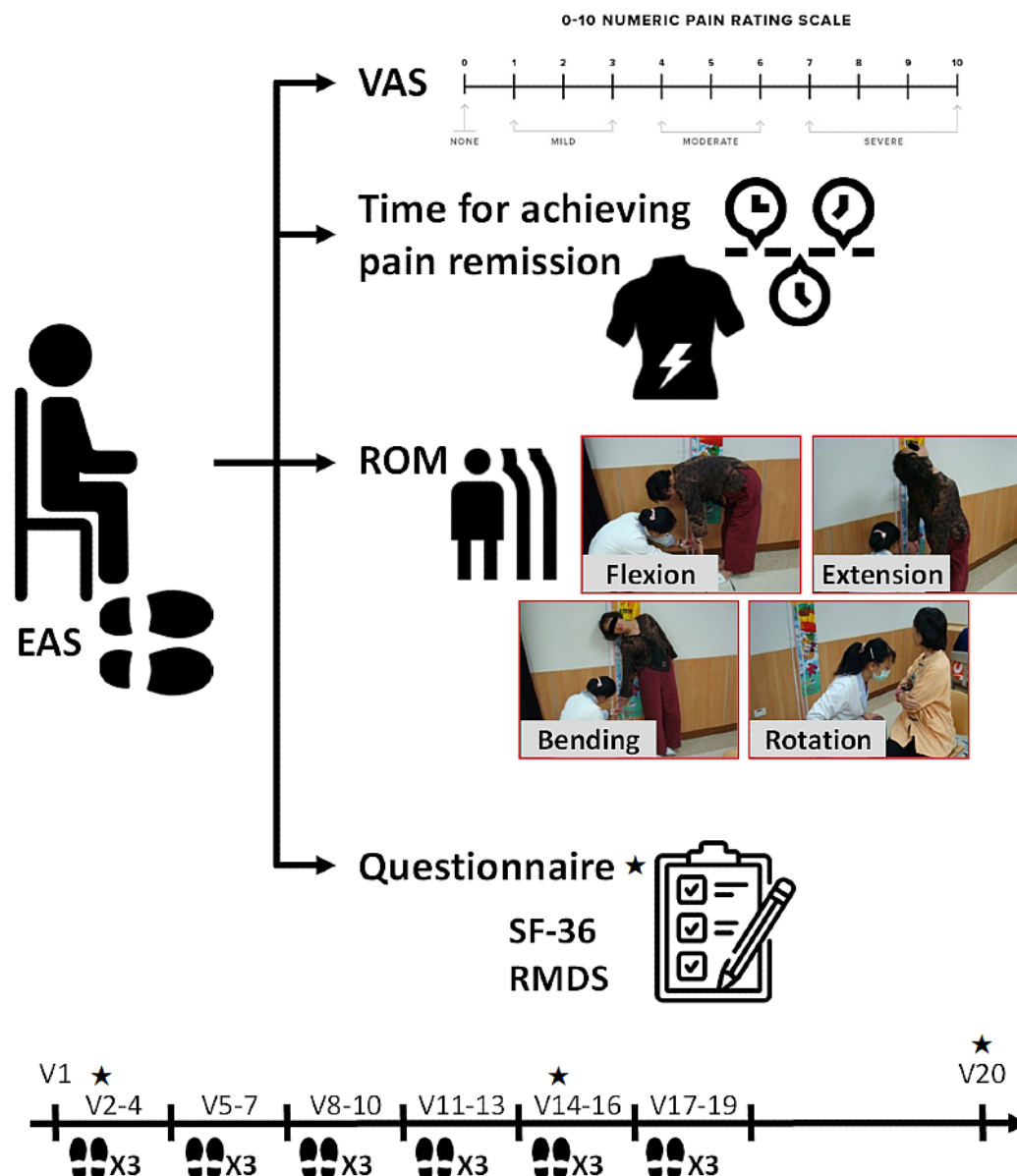
##### **Time for Achieving Pain Remission**

This measurement represented the course of pain remission and was defined as the time point when the VAS score began to decrease.

##### **Range of Motion**

The patient stood upright without shoes and with both heels close together and toes slightly apart by 15°. During measurement, the knee was extended and hands were relaxed and dropped naturally. The distance from the middle finger of both hands to the floor was measured under lumbar flexion, extension, and lateral bending. Patients sat on chairs with their feet at shoulder width. Their elbows were in flexion position, and their hands were placed in front of the chest. The angle of the protractor on the wrist was measured under lumbar rotation. The range of motion of the lumbar spine was assessed before and after each EAS treatment (2nd-19th visit); otherwise, the range of motion was assessed at the first visit and at 2 weeks after the last treatment (20th visit). The study protocol, the circumstance of the experimental investigation, and the study assessments are shown in [Figure 1](#).

**Figure 1.** The study protocol, the circumstance of the experimental investigation, and the study assessments. Stars in the figure indicate that questionnaires were administered on the 2nd, 14th, and 20th visits. EAS: electronic acupuncture shoes; VAS: visual analog scale; ROM: range of motion; SF-36: 36-item short form; RMDS: Roland Morris Disability Scale; V: visiting day.



### Quality of Life

The 36-item short form (SF-36) health survey was used to assess the health-related quality of life. It consists of 36 questions grouped into 8 domains: general health (6 items), vitality (4 items), physical function (10 items), bodily pain (2 items), physical role limitation (4 items), emotional role limitation (3 items), social function (2 items), and mental health (5 items). For each domain, scores range from 0 to 100, and higher scores reflect better quality of life [18].

### Maintaining Treatment Effect

The Roland Morris Disability Scale (RMDS) questionnaire was used to assess the functional disability due to LBP. This questionnaire consists of 24 questions that focus on the regular activities of daily living. Each affirmative answer corresponds

to 1 point, and the total number of points determines the final score. The total score ranges from 0 to 24, and higher scores reflect increased disability. Scores higher than 14 indicate severe impairment [19]. The SF-36 and the RMDS questionnaire were administered to patients at the 2nd, 14th, and 20th visits.

### Statistical Analysis

The objective of the statistical analysis was to effectively determine whether the difference between the treatment success rate of the experimental group was at least 30% more than that of the control group, where treatment success denoted VAS score after treatment being lower than that before treatment.  $\alpha$  and  $\beta$  were set as .10 and .20, respectively. At a test power of 80%, the estimated effective sample size was 66. Thus, according to an experimental group-control group ratio of 1:1, the number of patients assigned to each group was 33. To

reasonably assess the treatment outcome, the patients were required to undergo more than 12 treatments before their treatment outcome was included in the assessment. The SAS program (version 9.3) was used to analyze the data (SAS Institute Inc). Chi-squared test and Fisher exact test were used for the comparison of the categorical variables between the groups. For continuous variables, Student *t* test (two-tailed) was applied for 2 independent samples between the groups, while paired *t* test was chosen for within-group evaluation. The results are reported as mean (SD). A *P* value of  $<.05$  was considered statistically significant.

## Results

### Analytic Paradigms

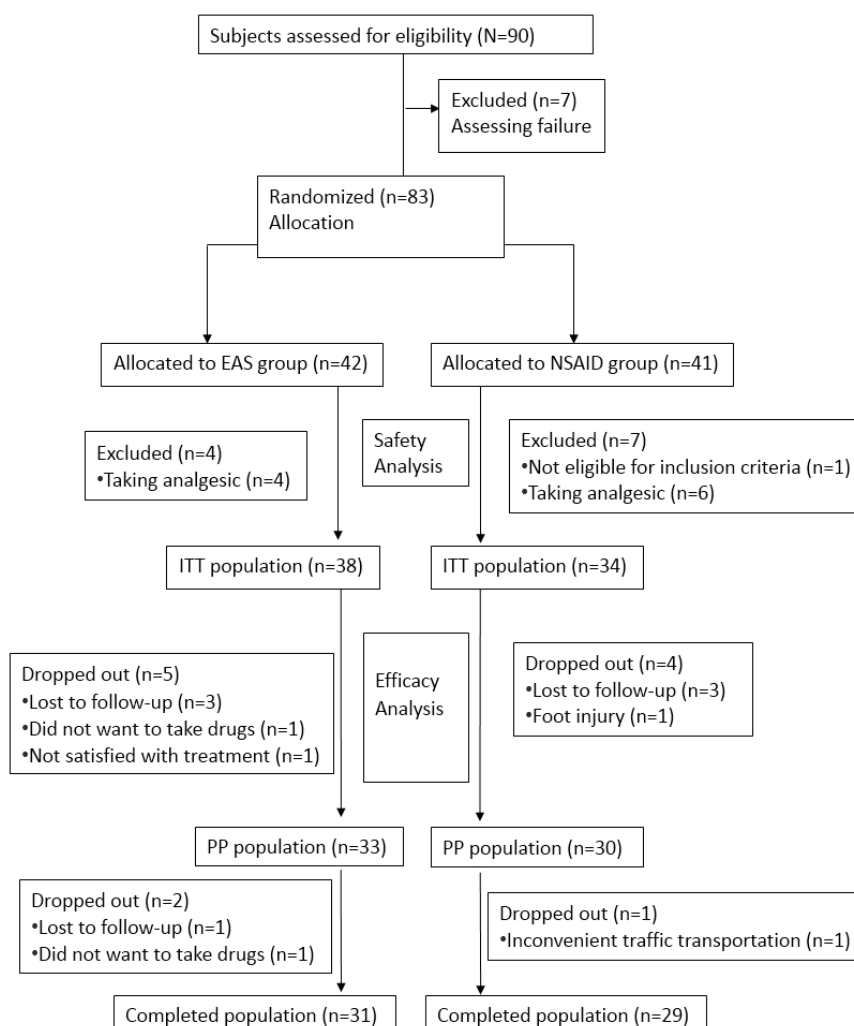
The population evaluated in the intention-to-treat (ITT) analysis included those who met the inclusion and exclusion criteria and

who received treatment at least once without violating the protocol. The population evaluated in the per-protocol (PP) analysis was included as those receiving treatment at least 12 times without violating the protocol.

### Subject Characteristics

Eighty-three patients were enrolled in this study between April 2009 and January 2012. One patient was excluded because of not meeting the inclusion criteria, and 10 patients were excluded because they received analgesics. The CONSORT (Consolidated Standards of Reporting Trials) ([Multimedia Appendix 1](#)) flow diagram for the study is shown in [Figure 2](#). Patient characteristics, including age, sex, weight, height, and duration of pain, were similar between the groups, except for age in the ITT analysis, which differed significantly (EAS group, 45.7 years and NSAID group, 41.1 years,  $P=.04$ ) ([Table 1](#)).

**Figure 2.** CONSORT flow diagram. CONSORT: Consolidated Standards of Reporting Trials; EAS: electronic acupuncture shoes; NSAID: nonsteroidal anti-inflammatory drug; ITT: intention-to-treat; PP: per-protocol.



**Table 1.** Characteristics of the patients and duration of pain.

Characteristics	Intention-to-treat population			Per-protocol population		
	Group EAS <sup>a</sup>	Group NSAID <sup>b</sup>	<i>P</i> value	Group EAS	Group NSAID	<i>P</i> value
Patients, n	38	34		33	30	
Age (years), mean (SD)	45.7 (10.19)	41.1 (8.10)	.04 <sup>c</sup>	45.3 (10.58)	42.0 (8.15)	.17
<b>Gender</b>			.11			.18
Females, n	28	19		25	18	
Males, n	10	15		8	12	
Weight (kg), mean (SD)	58.5 (11.25)	63.5 (12.52)	.08	57.4 (9.02)	62.5 (12.49)	.07
Height (cm), mean (SD)	161.8 (8.18)	164.1 (9.32)	.28	162.0 (8.25)	163.2 (9.55)	.59
Duration of pain (years), mean (SD)	8.6 (7.79)	8.5 (6.73)	.94	8.7 (8.09)	8.9 (6.75)	.94

<sup>a</sup>EAS: electronic acupuncture shoes.<sup>b</sup>NSAID: nonsteroidal anti-inflammatory drug.<sup>c</sup>Only this value was significant at  $P < .05$ .

## Adverse Effects

No severe adverse effects were reported in this study. Moderate and mild adverse effects such as pain in the extremities, back pain, hypoesthesia, and arthralgia were recorded. As shown in

**Table 2**, no significant differences were observed between both the groups. All adverse effects occurred during the period of EAS (or sham) treatment and were relieved after the completion of EAS (or sham) treatment without sequelae.

**Table 2.** Data of the adverse effects in the patients.

Adverse effects	Group EAS <sup>a</sup> , n=42, n (%)	Group NSAID <sup>b</sup> , n=41, n (%)
Patients with at least one adverse effect	22 (52)	19 (46)
Feeling hot	5 (12)	2 (5)
Pain in the sole	6 (14)	6 (15)
Arthralgia	3 (7)	2 (5)
Back pain	5 (12)	1 (2)
Limb discomfort	4 (10)	0 (0)
Muscle tightness	4 (10)	2 (5)
Musculoskeletal pain	4 (10)	1 (2)
Pain in extremity	13 (31.0)	9 (22)
Sensation of heaviness	3 (7)	1 (2)
Hypoesthesia	13 (31)	7 (17)

<sup>a</sup>EAS: electronic acupuncture shoes.<sup>b</sup>NSAID: nonsteroidal anti-inflammatory drug.

## Primary Outcome Measurements

### Treatment Success Rate

Treatment success was defined as the VAS score after the intervention being lower than that before the intervention. The treatment success rate in each visit is shown in **Table 3**. During the period from the 2nd visit to 5th visit, the NSAID and EAS groups were prescribed ibuprofen and placebo, respectively, for 7 days, and the treatment success rate in each visit was higher in the EAS group in both the ITT and PP analyses, but without

significant differences. During the period from the 6th visit to 19th visit, the treatment success rate in each visit in the EAS group was higher than that in the NSAID group in the ITT and PP analyses, and significant differences were observed only during the 16th visit in the ITT analysis (EAS group: 31/37, 84% and NSAID group: 21/34, 62%;  $P = .04$ ). The treatment success rate in the last visit (20th visit) in the EAS group was lower than that in 19th visit in both the ITT and PP analyses. The treatment success rate in the EAS group was higher than that in the NSAID group, but there were no significant differences.



**Table 3.** Data of the success rate of the treatment at each visit.

Visit	Success rate in the intention-to-treat population			Success rate in the per-protocol population		
	Group EAS <sup>a</sup> , n=38 <sup>b</sup> , n (%)	Group NSAID <sup>c</sup> , n=34, n (%)	P value	Group EAS, n=33, n (%)	Group NSAID, n=30, n (%)	P value
2	19 (50)	16 (47)	.72	17 (52)	15 (50)	.90
3	24 (65)	21 (62)	.79	21 (64)	19 (63)	.98
4	24 (65)	21 (62)	.79	21 (64)	19 (63)	.98
5	26 (70)	19 (56)	.21	23 (70)	18 (60)	.42
6	27 (73)	22 (65)	.45	24 (73)	21 (70)	.81
7	25 (67)	22 (65)	.80	22 (67)	20 (67)	>.99
8	28 (76)	22 (65)	.31	25 (76)	20 (67)	.43
9	28 (76)	21 (62)	.21	25 (76)	19 (63)	.28
10	28 (76)	22 (65)	.31	25 (76)	20 (67)	.43
11	30 (81)	23 (68)	.19	27 (82)	21 (70)	.27
12	29 (78)	23 (68)	.31	26 (79)	21 (70)	.42
13	31 (84)	24 (71)	.18	28 (85)	22 (73)	.26
14	31 (84)	22 (65)	.07	28 (85)	20 (67)	.09
15	29 (78)	22 (65)	.20	26 (79)	20 (67)	.28
16	31 (84)	21 (62)	.04 <sup>d</sup>	28 (85)	19 (63)	.05
17	26 (70)	20 (59)	.31	23 (70)	18 (60)	.42
18	27 (73)	21 (62)	.31	24 (73)	19 (63)	.42
19	29 (78)	20 (59)	.08	26 (79)	18 (60)	.11
20	25 (68)	20 (59)	.45	22 (67)	18 (60)	.58

<sup>a</sup>EAS: electronic acupuncture shoes.

<sup>b</sup>n=37 after the 2nd visit because the patient number 55 was excluded from the intention-to-treat analysis (the patient was not satisfied with treatment and she quitted the study).

<sup>c</sup>NSAID: nonsteroidal anti-inflammatory drug.

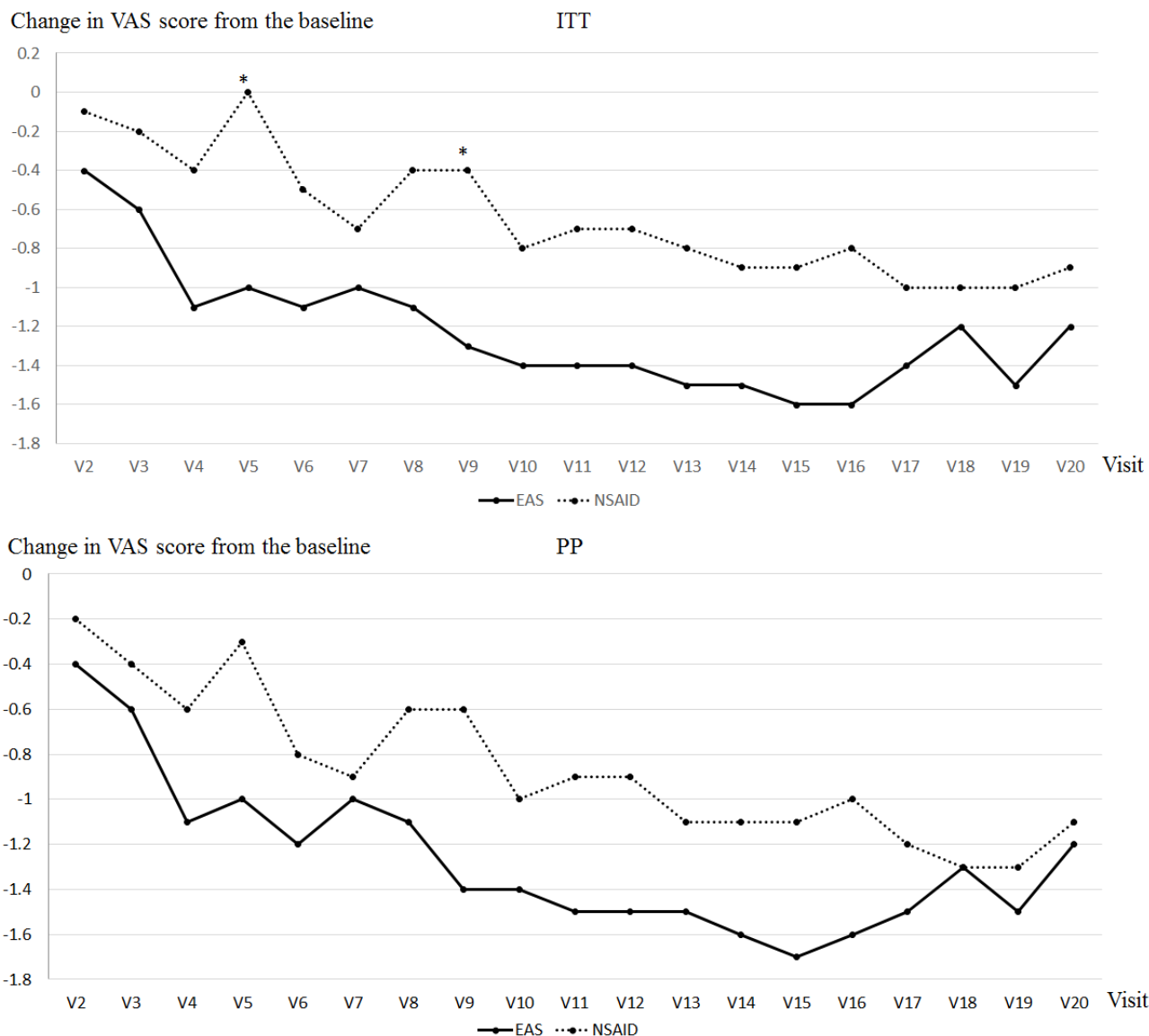
<sup>d</sup>Only this value was significant at  $P<.05$ .

### Changes in the VAS Score From the Baseline

We calculated the change in the VAS score from the baseline after each treatment, and the results are shown in [Figure 3](#). The VAS score at baseline was the VAS score measured during the first visit. At the first visit, mean (SD) VAS scores in the EAS and NSAID groups were 4.4 (2.46) and 4.2 (2.35), respectively, in the ITT analysis. The mean (SD) VAS scores in the EAS and NSAID groups were 4.4 (2.58) and 4.2 (2.46), respectively, in the PP analysis. No significant differences were observed between both groups. The change in the VAS score from baseline at each visit in the EAS group was greater than that in the NSAID group in the ITT and PP analyses, and significant

differences were observed at the 5th visit and 9th visit in the ITT analysis ( $P=.048$  and  $P=.048$ , respectively). In both the ITT and PP analyses, VAS scores in the EAS and NSAID groups gradually decreased during the period from the 2nd visit to the 4th visit but increased in the 5th visit. During the period from the 6th visit to 19th visit, the VAS score in the NSAID group fluctuated up and down. However, the VAS score in the EAS group continued to decrease from the 6th visit to the 17th visit and fluctuated from the 17th visit to the 19th visit in the ITT analysis. In the ITT analysis, VAS scores in the last visit in both groups were less than that at baseline. A similar trend was observed in the PP analysis.

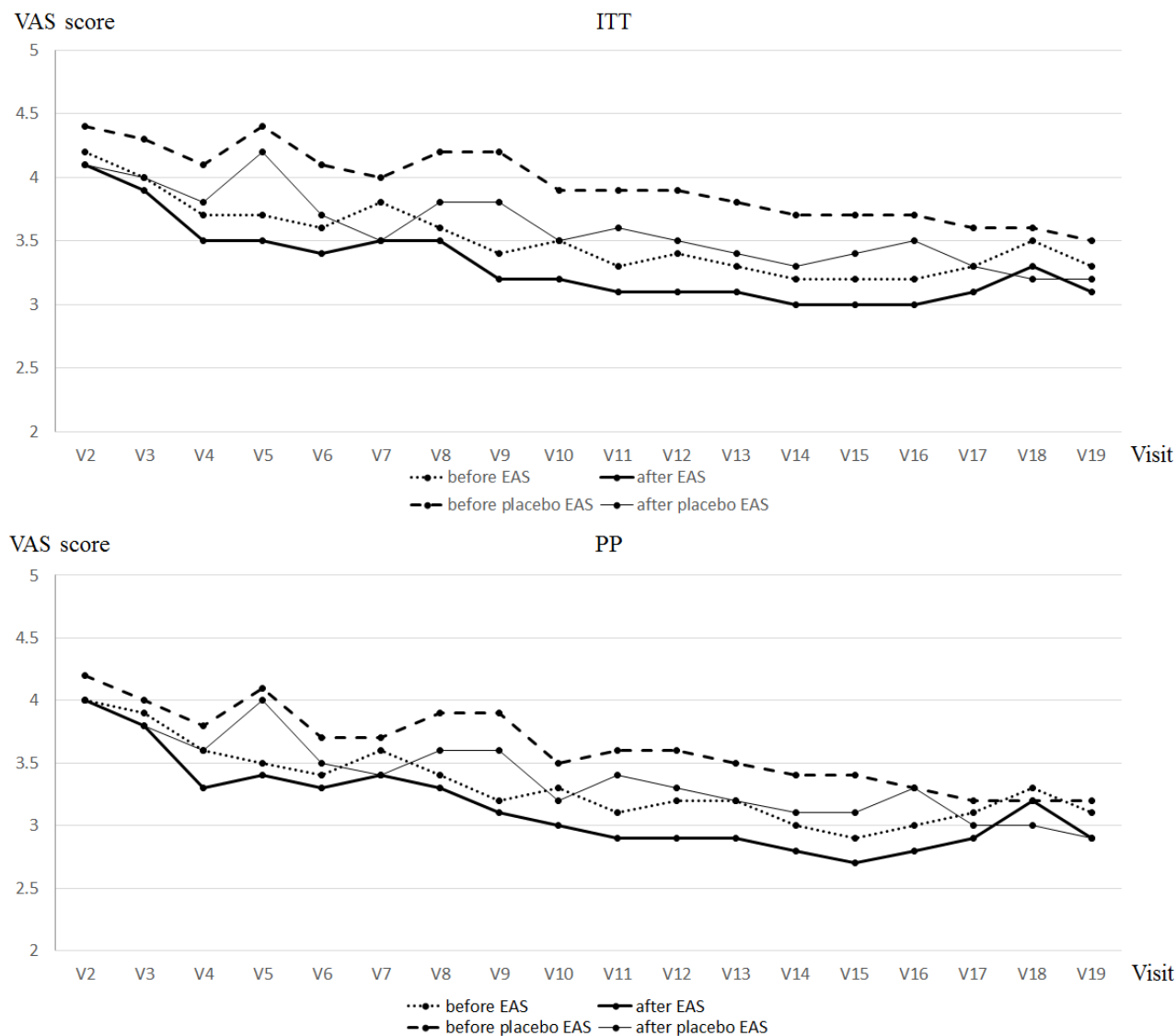
**Figure 3.** Change in visual analog scale (VAS) score from baseline in each visit in intention-to-treat and per-protocol analyses. Data are represented as means. The groups were compared at a significance level of .05. The x axis represents the day of visit for treatment, while the y axis indicates the change in VAS score from the baseline. EAS: electronic acupuncture shoes; NSAID: nonsteroidal anti-inflammatory drug; ITT: intention-to-treat; PP: per-protocol.



### VAS Scores Before and After Treatment in Each Visit

As presented in Figure 4, no significant differences were observed between the groups in both the ITT and PP analyses.

**Figure 4.** Visual analog scale (VAS) score before and after treatment in each visit in intention-to-treat and per-protocol analyses. Data are shown as means. The x axis represents the day of visit for treatment, while the y axis indicates the VAS scores. EAS: electronic acupuncture shoes; ITT: intention-to-treat; PP: per-protocol.



## Secondary Outcome Measurements

### Time for Achieving Pain Remission

This measurement represented the course of pain remission and was defined as the time point when the VAS score began to decrease. As presented in Table 4, the time for achieving pain

remission in the EAS group was 14.5 days, which was shorter than that in the NSAID group (17.4 days), but these differences were not significant in the ITT analysis. The time for achieving pain remission in the EAS group was 15.7 days, which was shorter than that in the NSAID group (16.9 days), but these differences were not significant in the PP analysis.

**Table 4.** Time for achieving pain remission.

Time for achieving pain remission	Intention-to-treat population			Per-protocol population		
	Group EAS <sup>a</sup> , n=38	Group NSAID <sup>b</sup> , n=34	P value	Group EAS, n=33	Group NSAID, n=30	P value
Mean (SD), days	14.5 (15.73)	17.4 (18.84)	.49	15.7 (16.56)	16.9 (19.07)	.79
Median days	8.0	8.0		8.0	8.0	
Range of days (min, max)	3.0, 63.0	2.0, 63.0		2.0, 63.0	2.0, 63.0	
95% CI	9.33-19.67	10.78-23.93		9.79-21.54	9.78-24.02	

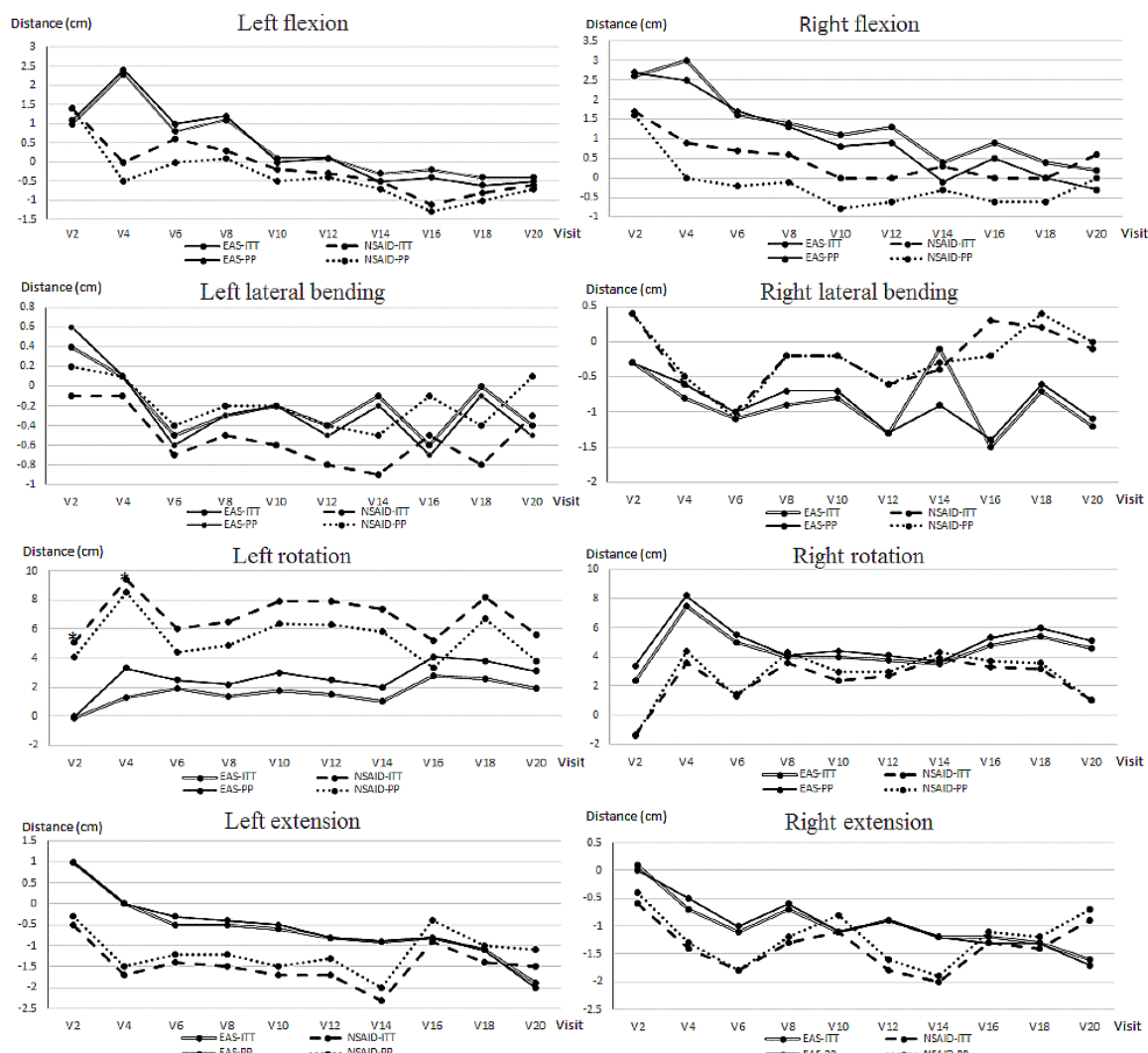
<sup>a</sup>EAS: electronic acupuncture shoes.<sup>b</sup>NSAID: nonsteroidal anti-inflammatory drug.

### Range of Motion

The range of motion included right and left flexion, lateral bending, rotation, and extension. The results are shown in Figure

5. No significant differences were observed between the groups in both the ITT and PP analyses. However, significant differences were observed in the left rotation at the 2nd visit and 4th visit ( $P=.049$  and  $P=.03$ , respectively).

**Figure 5.** Change in range of motion from baseline in intention-to-treat and per-protocol analyses. The x axis represents the day of visit for treatment, while the y axis indicates the distance (cm). EAS: electronic acupuncture shoes; NSAID: nonsteroidal anti-inflammatory drug; ITT: intention-to-treat; PP: per-protocol.



### Quality of Life and Maintaining Treatment Effect

The quality of life and maintaining treatment effect were assessed using SF-36 (Table 5) and RMDS (Table 6),

respectively. All the results showed no significant differences between the groups in both the ITT and PP analyses.

**Table 5.** Quality of life assessed using the 36-item short form health survey.

Visit, SF-36 <sup>a</sup> scale	Intention-to-treat population			Per-protocol population		
	Group EAS <sup>b</sup> , n=38 <sup>c</sup> , mean (SD)	Group NSAID <sup>d</sup> , n=34, mean (SD)	P value	Group EAS, n=33, mean (SD)	Group NSAID, n=30, mean (SD)	P value
<b>Visit 2 (baseline)</b>						
Physical function	73.6 (19.96)	75.1 (16.72)	.72	75.0 (19.16)	74.8 (17.49)	.97
Role limitation (Physical)	50.0 (43.11)	44.1 (39.91)	.55	54.5 (41.67)	44.2 (41.36)	.33
Role limitation (Emotional)	61.4 (42.82)	56.9 (39.81)	.64	66.7 (40.82)	60.0 (39.54)	.51
Vitality	49.1 (21.82)	50.9 (12.40)	.66	49.7 (20.32)	53.2 (10.04)	.39
Mental health	61.4 (15.06)	63.4 (15.39)	.58	61.3 (14.34)	66.0 (14.45)	.21
Social function	73.0 (18.50)	65.4 (21.77)	.12	72.7 (18.87)	66.7 (22.34)	.25
Bodily pain	55.6 (19.20)	53.9 (18.29)	.70	56.5 (19.18)	55.3 (18.20)	.80
General health	51.9 (20.55)	48.8 (18.89)	.50	53.1 (21.06)	49.6 (19.54)	.49
<b>Visit 14 (change from baseline)</b>						
Physical function	-1.6 (11.16)	-0.3 (19.26)	.73	-1.8 (11.98)	-0.3 (20.55)	.73
Role limitation (Physical)	1.3 (31.27)	8.1 (40.23)	.43	1.5 (33.62)	9.2 (42.79)	.43
Role limitation (Emotional)	3.5 (36.18)	-2.0 (36.64)	.53	4.0 (38.87)	-2.2 (39.08)	.53
Vitality	3.2 (11.07)	1.3 (11.83)	.48	3.8 (11.85)	1.5 (12.60)	.47
Mental health	2.0 (9.05)	-0.8 (12.33)	.27	2.3 (9.70)	-0.9 (13.15)	.27
Social function	2.3 (12.28)	4.4 (15.35)	.52	2.7 (13.17)	5.0 (16.28)	.53
Bodily pain	3.1 (12.8)	6.8 (15.61)	.27	3.6 (13.75)	7.8 (16.43)	.28
General health	2.4 (12.39)	3.2 (15.19)	.82	2.8 (13.28)	3.6 (16.16)	.83
<b>Visit 20 (change from baseline)</b>						
Physical function	0.0 (12.63)	-3.2 (13.25)	.29	0 (13.58)	3.7 (14.08)	.30
Role limitation (Physical)	3.3 (30.85)	11.0 (37.03)	.34	3.8 (33.14)	12.5 (39.25)	.34
Role limitation (Emotional)	0 (39.52)	4.9 (37.72)	.59	0 (42.49)	5.6 (40.19)	.60
Vitality	3.2 (15.33)	5.0 (14.09)	.62	3.8 (16.49)	5.7 (14.90)	.63
Mental health	1.5 (8.74)	1.0 (10.15)	.82	1.7 (9.38)	1.1 (10.82)	.82
Social function	4.3 (11.36)	0.7 (19.69)	.36	4.9 (12.08)	0.8 (21.00)	.36
Bodily pain	8.4 (15.13)	10.1 (14.71)	.63	9.6 (15.87)	11.4 (15.18)	.65
General health	2.4 (11.84)	3.5 (14.80)	.74	2.8 (12.69)	3.9 (15.73)	.75

<sup>a</sup>SF-36: 36-item of a short form health survey.<sup>b</sup>EAS: electronic acupuncture shoes.<sup>c</sup>n=37 after the 2nd visit because the patient number 55 was excluded from the intention-to-treat analysis (the patient was not satisfied with treatment and she quitted the study).<sup>d</sup>NSAID: nonsteroidal anti-inflammatory drug.



**Table 6.** Quality of life assessed using the Roland Morris Disability Scale.

Visit, Roland Morris Disability Scale	Intention-to-treat population			Per-protocol population		
	Group EAS <sup>a</sup> , n=38 <sup>b</sup>	Group NSAID <sup>c</sup> , n=34	<i>P</i> value	Group EAS, n=33	Group NSAID, n=30	<i>P</i> value
<b>Visit 2 (baseline)</b>			.77			.57
Mean (SD)	7.3 (4.64)	7.6 (4.62)		7.1 (4.91)	7.8 (4.88)	
Median	6.5	7.0		6.0	7.5	
Range (min, max)	0.0, 17.0	1.0, 18.0		0.0, 17.0	1.0, 18.0	
<b>Visit 14 (baseline)</b>			.77			.57
Mean (SD)	7.3 (4.64)	7.6 (4.62)		7.1 (4.91)	7.8 (4.88)	
Median	6.5	7.0		6.0	7.5	
Range (min, max)	0.0, 17.0	1.0, 18.0		0.0, 17.0	1.0, 18.0	
<b>Visit 20 (baseline)</b>			.78			.49
Mean (SD)	-2.1 (4.79)	-2.4 (4.62)		-1.1 (4.34)	-1.9 (4.50)	
Median	-2.0	-1.0		0.0	-1.0	
Range (min, max)	-11.0, 9.0	-16.0, 3.0		-11.0, 9.0	-16.0, 3.0	

<sup>a</sup>EAS: electronic acupuncture shoes.

<sup>b</sup>n=37 after the 2nd visit because the patient number 55 was excluded from the intention-to-treat analysis (the patient was not satisfied with treatment and she quitted the study).

<sup>c</sup>NSAID: nonsteroidal anti-inflammatory drug.

## Discussion

### Principal Results

We conducted a trial to compare the effects of EAS and conventional treatment (NSAIDs) for CLBP. In the ITT and PP analyses, the change in the VAS score from baseline in the EAS group in each visit was greater than that in the NSAID group, and significant differences were observed in the 5th visit and 9th visit in the ITT analysis ( $P=.048$  and  $P=.048$ , respectively). In the ITT and PP analyses, the treatment success rate in each visit in the EAS group was higher than that in the NSAID group, and significant differences were observed only in the 16th visit ( $P=.04$ ). In both the ITT and PP analyses, there were no significant differences in the VAS score before and after each treatment between the groups. No significant differences were observed in the time required for achieving pain remission, range of motion, SF-36 scores, and RMDS scores between the groups, but significant differences were observed in the left rotation at the 2nd and 4th visits.

### Comparison With Prior Work

EAS are a novel medical device, and it combine the properties of electroacupuncture and TENS and are different from acupoint TENS [8-10]. Electrodes of TENS are placed on the painful area of the body, that is, close to the lesion or near the nerve bundles proximal to the painful area [8]. Instead of local treatment, EAS are worn on the feet with electrical stimulation over 2 acupoints (KI1 and *shimian*) on the soles. Moreover, the administration of both conventional TENS and acupuncture-like TENS is required to achieve physiological intentions (paresthesia and muscle twitch, respectively) to confirm its effectiveness, thereby making patient blinding impossible [8-10].

By contrast, EAS outputs complex waveforms composed of low-frequency and middle-frequency waves. Furthermore, the amplitude of the current is so low that patients wearing EAS barely notice they are under treatment.

In clinical practice, needling on *yongquan* (KI1) is usually employed for treating conditions of disturbance of consciousness [20]. Few studies have evaluated the effect of electroacupuncture on KI1 [21]. *Shimian* is an extra acupoint, not located on the meridian, and it is employed to treat insomnia. According to the theory of traditional Chinese medicine, the flow of defensive *qi* (*wei qi*) travels around the outer side of the body (*yang*) through a special pattern and streams into the inner side of the body (*yin*) through the kidney meridian [22-24]. Moreover, in traditional Chinese medicine, one mechanism underlying insomnia is that the defensive *qi* cannot stream from the *yang* to *yin* [22,23]. Thus, many acupoints of the kidney meridian are believed to regulate the flow of defensive *qi* and are used to treat insomnia [25,26]. According to the theory of traditional Chinese medicine, chronic pain is a result of disharmony or depletion in the supply of *qi* [27,28]. For these reasons, we suspected that the regulation of the circulating defensive *qi* might relieve chronic pain. Therefore, we chose the *yongquan* and *shimian* as our treatment acupoints and applied electric current to these areas.

Yang et al [12] conducted an animal study, wherein the results revealed that the application of TENS on KI1 could produce analgesic effects; this was demonstrated through the increased response time of hind paw licking to thermal stimuli induced by complete Freund's adjuvant. The results of immunostaining and western blotting of the brain and spinal tissue of rats revealed that the application of TENS on KI1 inhibited ERK2 activation and c-Fos expression, which are associated with pain

perception. In that study, TENS was not applied as a conventional method to the painful lesion or nearby nerve bundles but on an acupoint distal to the painful site. Stimuli were transferred through the peripheral afferent fiber to the central nervous system and they caused changes in the brain network, thereby affecting pain perception and modulation [13]. These findings can explain that EAS may alleviate CLBP through a central pathway.

Furthermore, we think that EAS also work possibly through a peripheral pathway. Application of acupuncture on acupoints remote from the lesion or site of pain to alleviate disease or uncomfotableness is a fundamental concept based on the meridian theory in traditional Chinese medicine. A study demonstrated that acupuncture on distal acupoints could increase the blood flow of the meridian distribution area that the acupoints belong to [14]. A few studies found that CLBP is correlated with insufficient blood flow in the lumbopelvic region [15,16]. To summarize these theories and findings, we speculate EAS may alleviate CLBP by improving the blood flow.

The strengths of our study are as follows. First, we designed a randomized, double-blinded, placebo-controlled, dual-intervention (real EAS with placebo analgesic and sham EAS with ibuprofen) crossover trial to investigate the effect of EAS compared with that of conventional NSAIDs. Achieving participant blinding is very difficult in the design of an electroacupuncture or TENS study unless patients have little knowledge of electroacupuncture or TENS [8]. In this study, all patients were unaware of EAS because EAS are innovative medical devices. Even if patients were aware of EAS, they still could not differentiate between the real and sham treatment because the applied current was too small to be felt. Second, we employed rigorous methodology in sample size calculation and analytical paradigms (ITT and PP) to evaluate the effects of the intervention [29]. Third, we obtained various outcome measurements to compare the effects of EAS and NSAIDs; we focused not only on the pain itself but also on the mental status and the range of motion affecting the quality of life of the patients.

### Limitations and Future Directions

Our study has several limitations. First, we did not test the success of blinding by using an assessment tool such as that developed by Deyo et al [30], although patients were not knowledgeable about EAS and the applied current was extremely low. Second, the duration of ibuprofen intake was only 1 week in order to prevent the side effects such as

gastrointestinal, cardiovascular, renal, hepatic, and other systemic adverse effects [31–35]; therefore, we could not determine whether a longer intake of ibuprofen would achieve a higher efficacy. In clinical practice, the duration of NSAID use in patients with CLBP should be examined and evaluated regularly [36]. Treatments for CLBP include pharmacological, interventional, and surgical strategies [36]. Oral NSAIDs are widely used as first-line therapy for CLBP [36]. However, for specific populations with CLBP, the risk of adverse effects and drug-drug interactions [37–41] should be considered before prescribing oral NSAIDs. In this situation, developing an alternative therapy with similar efficacy to NSAIDs is crucial. The effect of TENS on CLBP is still debatable [42]; although the effectiveness of acupuncture is highly recommended [43], it depends on the performer.

EAS might serve as a reliable alternative therapeutic tool. Additional studies for evaluating the effect of EAS on other sites or categories of pain such as neck pain or pain caused by cancer should be conducted. EAS are more than just a tool for treating LBP. If sensors can be integrated to monitor physiological parameters and then shared on the internet, physicians will be able to personalize the amount of electrical stimulation through remote control [44]. This system will make a great contribution to mobile health, and in this era of new infectious diseases such as COVID-19, this feature will enhance the applicability of EAS [45].

### Conclusion

In this prospective double-blinded randomized controlled study, we demonstrated that EAS had a better treatment success rate and analgesic effect compared to NSAIDs in patients with CLBP during some of their visits for treatment, with partial improvement in the range of motion. EAS could be considered as a potential noninferior and reliable alternative therapeutic tool for patients with CLBP who are contraindicated for oral NSAIDs.

### Data Availability

There are restrictions on the availability of data for this study owing to the signed consent agreements regarding data security, which only allow access to external researchers for research monitoring purposes. Requestors wishing to access the trial data to replicate or check our analyses can apply to Protech Pharmaservices Corporation (contact@ppccro.com) after receiving permission from the East Bamboo Company Limited (ingo@eastbamboo.com.tw) and principal investigator Doctor Yu-Sheng Chen (cusp01@cgmh.org.tw).

### Acknowledgments

This study was funded by the East Bamboo Company Limited (<http://eastbamboo.com.tw>) for the Chang Gung Medical Foundation Linkou Chang Gung Memorial Hospital (<https://www1.cgmh.org.tw/intr/intr2/c3s400/e-mainpage.html>). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors are grateful for the invaluable help provided by the East Bamboo Company Limited and Protech Pharmaservices Corporation Contract Research Organization for allowing the completion of the study. This study was supported in part by Chang Gung Memorial Hospital, Linkou, Taiwan, under Contract No. CMRPG3D0891. This study received valuable suggestions from the “Chinese Medicine Research Center, China Medical University” from The Featured Areas Research Center Program within the framework of the

Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan (CMRC-CMA-6). Shih-An Chang and Yi-Xiang Weng suggested some revisions, and this manuscript was edited by Wallace Academic Editing.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT checklist.

[[PDF File \(Adobe PDF File\), 135 KB - jmir\\_v22i10e22324\\_app1.pdf](#)]

## References

1. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Clinical Guidelines Committee of the American College of Physicians. Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain: A Clinical Practice Guideline From the American College of Physicians. *Ann Intern Med* 2017 Apr 04;166(7):514-530. [doi: [10.7326/M16-2367](#)] [Medline: [28192789](#)]
2. Cassidy JD, Carroll LJ, Côté P. The Saskatchewan health and back pain survey. The prevalence of low back pain and related disability in Saskatchewan adults. *Spine (Phila Pa 1976)* 1998 Sep 01;23(17):1860-6; discussion 1867. [doi: [10.1097/00007632-199809010-00012](#)] [Medline: [9762743](#)]
3. Taylor P, Pezzullo L, Grant SJ, Bensoussan A. Cost-effectiveness of Acupuncture for Chronic Nonspecific Low Back Pain. *Pain Pract* 2014 Sep;14(7):599-606. [doi: [10.1111/papr.12116](#)] [Medline: [24138020](#)]
4. Lee J, Choi T, Lee MS, Lee H, Shin B, Lee H. Acupuncture for acute low back pain: a systematic review. *Clin J Pain* 2013 Feb;29(2):172-185. [doi: [10.1097/AJP.0b013e31824909f9](#)] [Medline: [23269281](#)]
5. Chen W, Fu G, Shen W. [Progresses of studies on acupuncture analgesia for postoperative reaction]. *Zhen Ci Yan Jiu* 2013 Feb;38(1):83-87. [Medline: [23650807](#)]
6. Comachio J, Oliveira Magalhães M, Nogueira Burke T, Vidal Ramos LA, Peixoto Leão Almeida G, Silva APMCC, et al. Efficacy of acupuncture and electroacupuncture in patients with nonspecific low back pain: study protocol for a randomized controlled trial. *Trials* 2015 Oct 15;16:469 [FREE Full text] [doi: [10.1186/s13063-015-0850-7](#)] [Medline: [26472590](#)]
7. Kwon YB, Kang MS, Son SS, Kim JT, Lee YH, Han HJ, et al. Different frequencies of electroacupuncture modified the cellular activity of serotonergic neurons in brainstem. *Am J Chin Med* 2000;28(3-4):435-441. [doi: [10.1142/S0192415X00000519](#)] [Medline: [11154058](#)]
8. Johnson MI, Paley CA, Howe TE, Sluka KA. Transcutaneous electrical nerve stimulation for acute pain. *Cochrane Database Syst Rev* 2015 Jun 15(6):CD006142. [doi: [10.1002/14651858.CD006142.pub3](#)] [Medline: [26075732](#)]
9. Francis RP, Johnson MI. The characteristics of acupuncture-like transcutaneous electrical nerve stimulation (acupuncture-like TENS): a literature review. *Acupunct Electrother Res* 2011;36(3-4):231-258. [doi: [10.3727/036012911803634139](#)] [Medline: [22443026](#)]
10. DeSantana JM, Walsh DM, Vance C, Rakel BA, Sluka KA. Effectiveness of transcutaneous electrical nerve stimulation for treatment of hyperalgesia and pain. *Curr Rheumatol Rep* 2008 Dec;10(6):492-499 [FREE Full text] [doi: [10.1007/s11926-008-0080-z](#)] [Medline: [19007541](#)]
11. Moslemi N, Gohari S, Mozafari F, Zardian MG, Burvill C, Yahya MY, et al. A novel smart assistive knee brace incorporated with shape memory alloy wire actuator. *Journal of Intelligent Material Systems and Structures* 2020 May 26;31(13):1543-1556. [doi: [10.1177/1045389x20922911](#)]
12. Yang L, Yang L, Gao X. Transcutaneous electrical nerve stimulation on Yongquan acupoint reduces CFA-induced thermal hyperalgesia of rats via down-regulation of ERK2 phosphorylation and c-Fos expression. *Anat Rec (Hoboken)* 2010 Jul;293(7):1207-1213 [FREE Full text] [doi: [10.1002/ar.21157](#)] [Medline: [20583264](#)]
13. Qin W, Tian J, Bai L, Pan X, Yang L, Chen P, et al. fMRI connectivity analysis of acupuncture effects on an amygdala-associated brain network. *Mol Pain* 2008 Nov 13;4:55. [doi: [10.1186/1744-8069-4-55](#)] [Medline: [19014532](#)]
14. Wang L, Zhang W, Xie H, Tian Y, Xu Y. [Experimental verification of "treating lumbar-back problems by puncturing Weizhong (BL 40)" with blood perfusion imaging technique]. *Zhen Ci Yan Jiu* 2007 Aug;32(4):247-251. [Medline: [17907387](#)]
15. Paungmali A, Henry LJ, Silitertpisan P, Pirunsan U, Uthaihpun S. Improvements in tissue blood flow and lumbopelvic stability after lumbopelvic core stabilization training in patients with chronic non-specific low back pain. *J Phys Ther Sci* 2016 Jan;28(2):635-640 [FREE Full text] [doi: [10.1589/jpts.28.635](#)] [Medline: [27064327](#)]
16. Espahbodi S, Doré CJ, Humphries KN, Hughes SPF. Color Doppler ultrasonography of lumbar artery blood flow in patients with low back pain. *Spine (Phila Pa 1976)* 2013 Feb 15;38(4):E230-E236. [doi: [10.1097/BRS.0b013e31827ecd6e](#)] [Medline: [23197013](#)]
17. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain* 2004 Aug;8(4):283-291. [doi: [10.1016/j.ejpain.2003.09.004](#)] [Medline: [15207508](#)]

18. Burholt V, Nash P. Short Form 36 (SF-36) Health Survey Questionnaire: normative data for Wales. *J Public Health (Oxf)* 2011 Dec;33(4):587-603. [doi: [10.1093/pubmed/fdr006](https://doi.org/10.1093/pubmed/fdr006)] [Medline: [21307049](https://pubmed.ncbi.nlm.nih.gov/21307049/)]
19. Heapy A, LaChappelle K, Higgins D, Buta E, Driscoll M, Kerns R. (115) Demographic, psychological, and clinical predictors of Roland Morris Disability Questionnaire score among patients with chronic low back pain. *The Journal of Pain* 2016 Apr;17(4):S4-S5. [doi: [10.1016/j.jpain.2016.01.018](https://doi.org/10.1016/j.jpain.2016.01.018)]
20. Zhang H, Sun X, Liu S, Chen Y, Ling F. Neuronal activation by acupuncture at Yongquan (KI1) and sham acupoints in patients with disorder of consciousness: a positron emission tomography study. *Neural Regen Res* 2014 Mar 01;9(5):500-501 [FREE Full text] [doi: [10.4103/1673-5374.130070](https://doi.org/10.4103/1673-5374.130070)] [Medline: [25206845](https://pubmed.ncbi.nlm.nih.gov/25206845/)]
21. Li F, Li L, Wang X, Bai Y, Jiawula A, Bu Q, et al. [Effect of electroacupuncture stimulation of "Baihui" (GV 20) and "Yongquan" (KI 1) on expression of hippocampal amyloid- $\beta$  and low density lipoprotein receptor-related protein-1 in APP/PS 1 transgenic mice]. *Zhen Ci Yan Jiu* 2015 Feb;40(1):30-4, 55. [Medline: [25845217](https://pubmed.ncbi.nlm.nih.gov/25845217/)]
22. Unschuld PU. *Huang Di Nei Jing Ling Shu: The ancient classic on needle therapy*. Oakland, CA: University of California Press; 2016.
23. Unschuld PU, Tessenow H, Zheng J. *Huang di nei jing su wen: An annotated translation of Huang Di's Inner Classic -Basic Questions*. Berkeley: University of California Press; 2011.
24. Ye X, Dong M. A review on different English versions of an ancient classic of Chinese medicine: Huang Di Nei Jing. *J Integr Med* 2017 Jan;15(1):11-18. [doi: [10.1016/S2095-4964\(17\)60310-8](https://doi.org/10.1016/S2095-4964(17)60310-8)] [Medline: [28088255](https://pubmed.ncbi.nlm.nih.gov/28088255/)]
25. Huang Y, Chen Y. [Xingjian (LR 2) penetrated to Yongquan (KI 1) for insomnia]. *Zhongguo Zhen Jiu* 2013 Apr;33(4):313. [Medline: [23819230](https://pubmed.ncbi.nlm.nih.gov/23819230/)]
26. Wang S. [Clinical observation on acupuncture at Zhaohai (KI 6) and Shenmai (BL 62) for treatment of insomnia]. *Zhongguo Zhen Jiu* 2005 Nov;25(11):771-772. [Medline: [16335201](https://pubmed.ncbi.nlm.nih.gov/16335201/)]
27. Yang KH, Kim YH, Lee MS. Efficacy of Qi-therapy (external Qigong) for elderly people with chronic pain. *Int J Neurosci* 2005 Jul;115(7):949-963. [doi: [10.1080/00207450590901378](https://doi.org/10.1080/00207450590901378)] [Medline: [16051542](https://pubmed.ncbi.nlm.nih.gov/16051542/)]
28. Lee MS, Yang KH, Huh HJ, Kim HW, Ryu H, Lee HS, et al. Qi therapy as an intervention to reduce chronic pain and to enhance mood in elderly subjects: a pilot study. *Am J Chin Med* 2001;29(2):237-245. [doi: [10.1142/S0192415X01000277](https://doi.org/10.1142/S0192415X01000277)] [Medline: [11527067](https://pubmed.ncbi.nlm.nih.gov/11527067/)]
29. Laureano MR, Onishi ET, Bressan RA, Neto PB, Castiglioni MLV, Batista IR, et al. The effectiveness of acupuncture as a treatment for tinnitus: a randomized controlled trial using (99m)Tc-ECD SPECT. *Eur Radiol* 2016 Sep;26(9):3234-3242. [doi: [10.1007/s00330-015-4164-7](https://doi.org/10.1007/s00330-015-4164-7)] [Medline: [26747258](https://pubmed.ncbi.nlm.nih.gov/26747258/)]
30. Deyo RA, Walsh NE, Schoenfeld LS, Ramamurthy S. Can trials of physical treatments be blinded? The example of transcutaneous electrical nerve stimulation for chronic pain. *Am J Phys Med Rehabil* 1990 Feb;69(1):6-10. [doi: [10.1097/00002060-199002000-00003](https://doi.org/10.1097/00002060-199002000-00003)] [Medline: [2137345](https://pubmed.ncbi.nlm.nih.gov/2137345/)]
31. Henry D, Lim LL, Garcia Rodriguez LA, Perez Gutthann S, Carson JL, Griffin M, et al. Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta-analysis. *BMJ* 1996 Jun 22;312(7046):1563-1566 [FREE Full text] [doi: [10.1136/bmj.312.7046.1563](https://doi.org/10.1136/bmj.312.7046.1563)] [Medline: [8664664](https://pubmed.ncbi.nlm.nih.gov/8664664/)]
32. McGettigan P, Henry D. Cardiovascular risk and inhibition of cyclooxygenase: a systematic review of the observational studies of selective and nonselective inhibitors of cyclooxygenase 2. *JAMA* 2006 Oct 04;296(13):1633-1644. [doi: [10.1001/jama.296.13.jrv60011](https://doi.org/10.1001/jama.296.13.jrv60011)] [Medline: [16968831](https://pubmed.ncbi.nlm.nih.gov/16968831/)]
33. McGettigan P, Henry D. Cardiovascular risk with non-steroidal anti-inflammatory drugs: systematic review of population-based controlled observational studies. *PLoS Med* 2011 Sep;8(9):e1001098 [FREE Full text] [doi: [10.1371/journal.pmed.1001098](https://doi.org/10.1371/journal.pmed.1001098)] [Medline: [21980265](https://pubmed.ncbi.nlm.nih.gov/21980265/)]
34. Whelton A. Renal and related cardiovascular effects of conventional and COX-2-specific NSAIDs and non-NSAID analgesics. *Am J Ther* 2000 Mar;7(2):63-74. [doi: [10.1097/00045391-200007020-00004](https://doi.org/10.1097/00045391-200007020-00004)] [Medline: [11319575](https://pubmed.ncbi.nlm.nih.gov/11319575/)]
35. Bessone F. Non-steroidal anti-inflammatory drugs: What is the actual risk of liver damage? *World J Gastroenterol* 2010 Dec 07;16(45):5651-5661 [FREE Full text] [doi: [10.3748/wjg.v16.i45.5651](https://doi.org/10.3748/wjg.v16.i45.5651)] [Medline: [21128314](https://pubmed.ncbi.nlm.nih.gov/21128314/)]
36. Morlion B. Chronic low back pain: pharmacological, interventional and surgical strategies. *Nat Rev Neurol* 2013 Aug;9(8):462-473. [doi: [10.1038/nrneurol.2013.130](https://doi.org/10.1038/nrneurol.2013.130)] [Medline: [23817349](https://pubmed.ncbi.nlm.nih.gov/23817349/)]
37. Roubille C, Martel-Pelletier J, Davy J, Haraoui B, Pelletier J. Cardiovascular adverse effects of anti-inflammatory drugs. *Antiinflamm Antiallergy Agents Med Chem* 2013;12(1):55-67. [doi: [10.2174/1871523011312010008](https://doi.org/10.2174/1871523011312010008)] [Medline: [23286294](https://pubmed.ncbi.nlm.nih.gov/23286294/)]
38. Pinto A, Farrar JT, Hersh EV. Prescribing NSAIDs to patients on SSRIs: possible adverse drug interaction of importance to dental practitioners. *Compend Contin Educ Dent* 2009 Apr;30(3):142-51; quiz 152, 154. [Medline: [19397069](https://pubmed.ncbi.nlm.nih.gov/19397069/)]
39. Farkouh ME, Greenberg JD, Jeger RV, Ramanathan K, Verheugt FWA, Chesebro JH, et al. Cardiovascular outcomes in high risk patients with osteoarthritis treated with ibuprofen, naproxen or lumiracoxib. *Ann Rheum Dis* 2007 Jun;66(6):764-770 [FREE Full text] [doi: [10.1136/ard.2006.066001](https://doi.org/10.1136/ard.2006.066001)] [Medline: [17412741](https://pubmed.ncbi.nlm.nih.gov/17412741/)]
40. Cheetham TC, Levy G, Niu F, Bixler F. Gastrointestinal safety of nonsteroidal antiinflammatory drugs and selective cyclooxygenase-2 inhibitors in patients on warfarin. *Ann Pharmacother* 2009 Nov;43(11):1765-1773. [doi: [10.1345/aph.1M284](https://doi.org/10.1345/aph.1M284)] [Medline: [19809010](https://pubmed.ncbi.nlm.nih.gov/19809010/)]
41. Mackenzie IS, Coughtrie MWH, MacDonald TM, Wei L. Antiplatelet drug interactions. *J Intern Med* 2010 Dec;268(6):516-529 [FREE Full text] [doi: [10.1111/j.1365-2796.2010.02299.x](https://doi.org/10.1111/j.1365-2796.2010.02299.x)] [Medline: [21073556](https://pubmed.ncbi.nlm.nih.gov/21073556/)]



42. Khadilkar A, Odebiyi DO, Brosseau L, Wells GA. Transcutaneous electrical nerve stimulation (TENS) versus placebo for chronic low-back pain. *Cochrane Database Syst Rev* 2008 Oct 08(4):CD003008 [[FREE Full text](#)] [doi: [10.1002/14651858.CD003008.pub3](#)] [Medline: [18843638](#)]
43. Hopton A, MacPherson H. Acupuncture for chronic pain: is acupuncture more than an effective placebo? A systematic review of pooled data from meta-analyses. *Pain Pract* 2010;10(2):94-102. [doi: [10.1111/j.1533-2500.2009.00337.x](#)] [Medline: [20070551](#)]
44. Chen Y, Lin C, Tsai M, Chuang T, Lee OK. Wearable Motion Sensor Device to Facilitate Rehabilitation in Patients With Shoulder Adhesive Capsulitis: Pilot Study to Assess Feasibility. *J Med Internet Res* 2020 Jul 23;22(7):e17032 [[FREE Full text](#)] [doi: [10.2196/17032](#)] [Medline: [32457026](#)]
45. Wijesooriya NR, Mishra V, Brand PLP, Rubin BK. COVID-19 and telehealth, education, and research adaptations. *Paediatr Respir Rev* 2020 Sep;35:38-42 [[FREE Full text](#)] [doi: [10.1016/j.prrv.2020.06.009](#)] [Medline: [32653468](#)]

## Abbreviations

**CLBP:** chronic low back pain  
**CONSORT:** Consolidated Standards of Reporting Trials  
**EAS:** electronic acupuncture shoes  
**ITT:** intention-to-treat  
**LBP:** low back pain  
**NSAIDs:** nonsteroidal anti-inflammatory drugs  
**PP:** per-protocol  
**RMDS:** Roland Morris Disability Scale  
**SF-36:** 36-item short form  
**TENS:** transcutaneous electrical nerve stimulation  
**VAS:** visual analog scale

*Edited by G Eysenbach; submitted 09.07.20; peer-reviewed by M S, S Gohari; comments to author 04.09.20; revised version received 15.09.20; accepted 22.09.20; published 26.10.20.*

### *Please cite as:*

Yeh BY, Liu GH, Lee TY, Wong AMK, Chang HH, Chen YS

*Efficacy of Electronic Acupuncture Shoes for Chronic Low Back Pain: Double-Blinded Randomized Controlled Trial*

*J Med Internet Res* 2020;22(10):e22324

URL: <http://www.jmir.org/2020/10/e22324/>

doi: [10.2196/22324](#)

PMID: [33104004](#)

©Bo-Yan Yeh, Geng-Hao Liu, Tzung-Yan Lee, Alice May-Kuen Wong, Hen-Hong Chang, Yu-Sheng Chen. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# 24-Month Outcomes of Primary Care Web-Based Depression Prevention Intervention in Adolescents: Randomized Clinical Trial

Benjamin Van Voorhees<sup>1\*</sup>, MD, MPH; Tracy R G Gladstone<sup>2\*</sup>, PhD; Kunmi Sobowale<sup>3</sup>, MD; C Hendricks Brown<sup>4</sup>, PhD; David A Aaby<sup>5</sup>, MS; Daniela A Terrizzi<sup>1</sup>, PhD, RN; Jason Canel<sup>6</sup>, MD; Eumene Ching<sup>7</sup>, MD; Anita D Berry<sup>8</sup>, MSN, CNP, APN, PMHS; James Cantorna<sup>9</sup>, MD; Milton Eder<sup>10</sup>, MD; William Beardslee<sup>11</sup>, MD; Marian Fitzgibbon<sup>1,12,13</sup>, PhD; Monika Marko-Holguin<sup>1</sup>, MSS; Linda Schiffer<sup>12</sup>, MS; Miae Lee<sup>1</sup>, MD, MS; Sarah A de Forest<sup>1</sup>, MD; Emily E Sykes<sup>1</sup>, MA; Jennifer H Suor<sup>14</sup>, PhD; Theodore J Crawford<sup>15</sup>, MSW; Katie L Burkhouse<sup>14</sup>, PhD; Brady C Goodwin<sup>1</sup>, BS; Carl Bell<sup>14,16</sup>, MD

<sup>1</sup>Department of General Pediatrics, University of Illinois at Chicago, College of Medicine, Chicago, IL, United States

<sup>2</sup>The Robert S and Grace W Stone Primary Prevention Initiatives, Wellesley Centers for Women, Wellesley College, Boston, MA, United States

<sup>3</sup>Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States

<sup>4</sup>Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

<sup>5</sup>Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

<sup>6</sup>NorthShore University Health System, Evanston, IL, United States

<sup>7</sup>Harvard Vanguard, Cambridge, MA, United States

<sup>8</sup>Almost Home Kids, Ann & Robert H Lurie Children's Hospital, Chicago, IL, United States

<sup>9</sup>Franciscan Medical Specialists, Munster, IN, United States

<sup>10</sup>Department of Family Medicine and Community Health, University of Minnesota, Minneapolis, MN, United States

<sup>11</sup>Judge Baker Center, Harvard Medical School, Roxbury Crossing, MA, United States

<sup>12</sup>Institute for Health Research and Policy, School of Public Health, University of Illinois at Chicago, Chicago, IL, United States

<sup>13</sup>University of Illinois Cancer Center, University of Illinois at Chicago, Chicago, IL, United States

<sup>14</sup>Department of Psychiatry, University of Illinois at Chicago, College of Medicine, Chicago, IL, United States

<sup>15</sup>Dr Karen Taylor-Crawford & Associates, Palos Park, IL, United States

<sup>16</sup>Department of Psychiatry, Windsor University, School of Medicine, Cayon St Kitts, Saint Kitts and Nevis

\*these authors contributed equally

## Corresponding Author:

Benjamin Van Voorhees, MD, MPH  
Department of General Pediatrics  
University of Illinois at Chicago  
College of Medicine  
840 South Wood Street (MC 856)  
Chicago, IL  
United States  
Phone: 1 312 996 8352  
Email: [bvanvoor@uic.edu](mailto:bvanvoor@uic.edu)

## Abstract

**Background:** Adolescent depression carries a high burden of disease worldwide, but access to care for this population is limited. Prevention is one solution to curtail the negative consequences of adolescent depression. Internet interventions to prevent adolescent depression can overcome barriers to access, but few studies examine long-term outcomes.

**Objective:** This study compares CATCH-IT (Competent Adulthood Transition with Cognitive Behavioral Humanistic and Interpersonal Training), an internet-based intervention, to a general health education active control for depression onset at 12 and 24 months in adolescents presenting to primary care settings.

**Methods:** A 2-site randomized trial, blinded to the principal investigators and assessors, was conducted comparing Competent Adulthood Transition with Cognitive Behavioral Humanistic and Interpersonal Training to health education to prevent depressive episodes in 369 adolescents (193 youths were randomly assigned to Competent Adulthood Transition with Cognitive Behavioral

Humanistic and Interpersonal Training and 176 to health education) with subthreshold depressive symptoms or prior depressive episodes. Participants were recruited from primary care settings in the United States. The primary outcome was the occurrence of a depressive episode, determined by the Depression Symptom Rating. The secondary outcome was functioning, measured by the Global Assessment Scale.

**Results:** In intention-to-treat analyses, the adjusted hazard ratio favoring Competent Adulthood Transition with Cognitive Behavioral Humanistic and Interpersonal Training for first depressive episode was not statistically significant at 12 months (hazard ratio 0.77, 95% CI 0.42-1.40,  $P=.39$ ) and 24 months (hazard ratio 0.87, 95% CI 0.52-1.47,  $P=.61$ ). Competent Adulthood Transition with Cognitive Behavioral Humanistic and Interpersonal Training provided preventive benefit for first depressive episode for those with mild hopelessness or at least moderate paternal monitoring at baseline. Global Assessment Scale scores improved comparably in both groups (intention-to-treat).

**Conclusions:** A technology-based intervention for adolescent depression prevention implemented in primary care did not have additional benefit at 12 or 24 months. Further research is necessary to determine whether internet interventions have long-term benefit.

**Trial Registration:** ClinicalTrials.gov NCT01893749; <http://clinicaltrials.gov/ct2/show/NCT01893749>.

(*J Med Internet Res* 2020;22(10):e16802) doi:[10.2196/16802](https://doi.org/10.2196/16802)

## KEYWORDS

adolescent; depression; prevention; scalable; eHealth

## Introduction

As many as 7%-13% of adolescents living in the United States and other high-income countries experience minor or major depressive episodes each year [1]. Adolescents with depressive episodes, compared to those without depression, have a higher incidence of social adjustment challenges [2,3] and are at risk for future suicide attempts and recurrent depressive episodes throughout life [4]. Unfortunately, only a minority of depressed adolescents receive treatment [5]. The negative consequences of untreated depression, delayed treatment, and extensive lack of access to treatment call for evidence-based preventive interventions that can be disseminated widely.

Recent systematic reviews and meta-analyses [6,7] found that evidence-based interventions (eg, cognitive behavioral therapy) for adolescent depression prevention and treatment could be successfully delivered via the internet [6], indicating their potential to increase access to effective preventive programs. While digital interventions can circumvent geographic barriers, time constraints, and stigma by allowing users to access interventions from their personal devices, their effects on adolescent depression prevention are mixed [7,8], and more intensive interventions with human support or guidance may have superior results [7,8].

To this point, most digital-based preventive interventions have primarily focused on short-term reduction in depressive symptoms and have not focused on how the interventions impact the risk of developing depressive episodes over time. In a review [9] examining 83 studies targeting depressive outcomes in youth, of which 8 were delivered online, only one-third (32/83) specifically focused on reducing the risk of depressive episodes rather than symptoms; small effect size reductions of depressive symptoms were found postintervention, and there was a modest absolute risk reduction of depressive episodes [9]. This limitation, in part, results from a lack of prospective studies examining the long-term outcomes necessary to detect depression onset [7,10]. A prior randomized clinical trial of

CATCH-IT [11-13], the internet-based depression prevention intervention evaluated in this study, demonstrated lower depressive symptoms at 6, 12, and 30 months, relative to baseline. Despite these long-term outcomes, because randomization was based on primary care delivery models of CATCH-IT, all participants in that study received CATCH-IT rather than an active comparator.

This randomized clinical trial compared the efficacy of the CATCH-IT intervention to an online general health education program in preventing the onset of depressive episodes at 12 and 24 months in an intermediate to high-risk group of adolescents recruited from primary care settings [14]. We previously reported that at 6 months, CATCH-IT demonstrated preventive benefit against depressive episodes compared to that offered by health education for adolescents with higher depressive symptoms at baseline but not for the entire sample [15]. Given our prior 6-month results, showing significant effects for moderation based on baseline adolescent characteristics and documenting increased benefit of psychotherapy after termination (sleeping effects) [16-18], we hypothesized that a depression prevention benefit might emerge for the CATCH-IT group, and significant moderation relationships (in prespecified theory-based covariates) might be demonstrated in intention-to-treat analysis at 12 and 24 months [14].

## Methods

### Study Design and Setting

To evaluate the efficacy of CATCH-IT in a scalable setting, we used a hybrid type 1 efficacy-implementation design in which we tested the efficacy of the clinical intervention while simultaneously collecting information for large-scale implementation [19]. Focus groups tested the prototype of the system for usability and accessibility in urban minority adolescents. Adolescent participants welcomed the platform and expressed preference for an online program that is engaging

and private with some face-to-face interaction to express their feelings.

The trial model was implemented in 8 major US health systems (31 primary care sites including more than 41,000 adolescents) in a population health approach (screen all youth, offer intervention, assessment, refer those in need of treatment) with over 1200 primary care staff consenting and being trained. Data were collected from urban and suburban clinics located in Boston, Massachusetts and surrounding areas, and in Chicago, Illinois and surrounding areas, including northern Indiana. Participating primary care clinics were recruited by study staff and health care providers at the study sites. Research staff worked closely with office nursing staff conducting in-person adolescent screening and primary health care providers trained in motivational interviewing [14,20].

A more detailed description of our protocol, which follows the CONSORT statement, and a detailed description of the study design and implementation process appear elsewhere [14,20]. Institutional review board approval was received from the institutional review board of record (University of Illinois at Chicago) and local institutional review boards. A data safety and monitoring board reviewed the trial and results twice per year.

### Eligibility Criteria

Eligible adolescents were aged 13 to 18 years and had either an elevated score on the Center for Epidemiologic Studies Depression (CES-D) scale [21] (8-17 on 10-item scale at screening or  $\geq 16$  on a 20-item scale at baseline, indicating

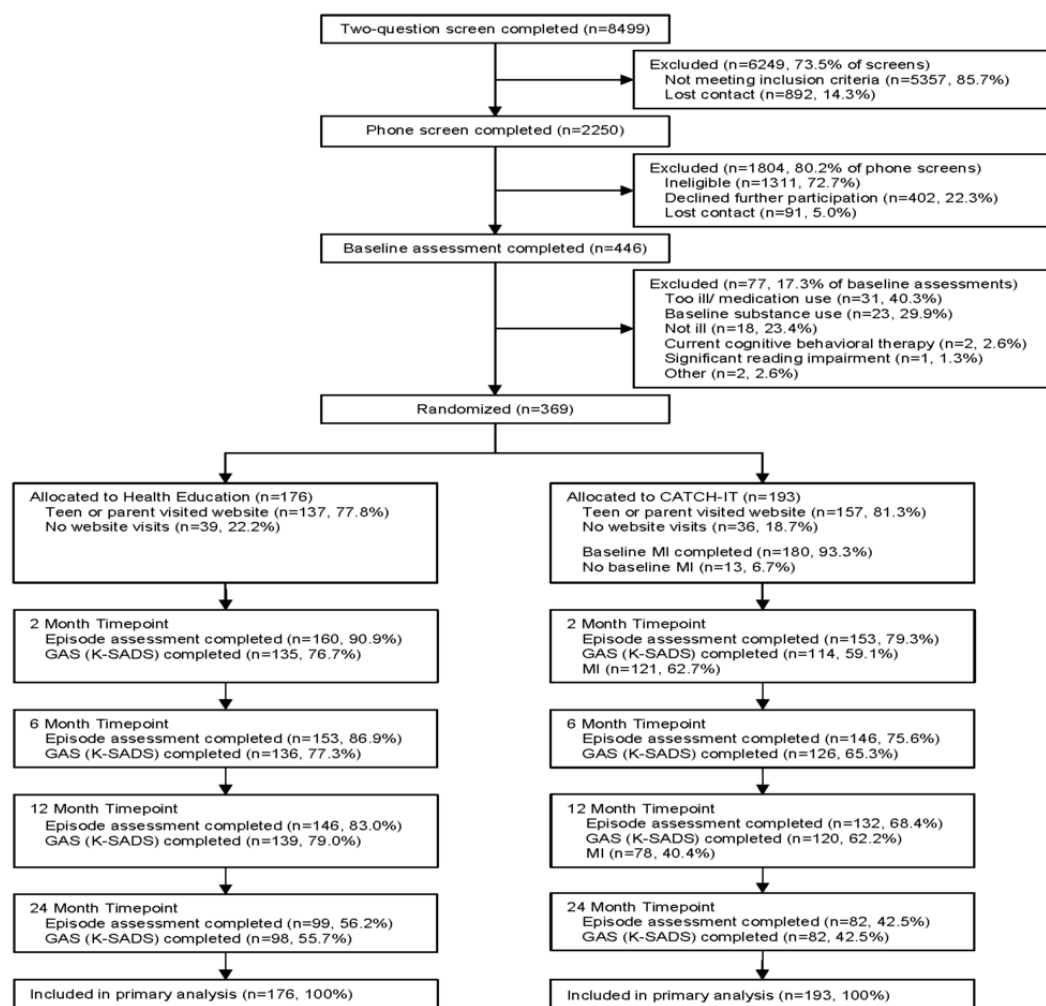
clinically significant risk for depression) [22] or a prior history of a major depressive episode or dysthymia [21,23].

Adolescents were excluded if they had a current DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) diagnosis of major depressive disorder or were currently being treated for major depressive disorder. Adolescents with schizophrenia, psychosis, bipolar disorder, significant reading impairment or developmental disability, current drug or alcohol abuse, imminent risk of suicide, or who were in treatment for serious medical conditions were also excluded.

### Participant Recruitment and Enrollment

Adolescents were recruited from February 2012 to July 2016 via screening, posted flyers, recruitment letters (sent to all families) and by information offered directly to families by their primary care physician during clinic visits. Initial screening for risk of depression was administered either during a face-to-face interview or over the phone using the Patient Health Questionnaire-2 [11,24]. A positive screen was the presence of any cardinal depression symptom (depressed mood, irritability, or anhedonia over the prior 2 weeks). Eligible participants and parents were then referred for an enrollment assessment at the office of their primary care physician. During this appointment, eligibility was formally confirmed through a full assessment (using Kiddie Schedule for Affective Disorders Scale, KSADS; CES-D10), accessibility to a computer and internet was confirmed, and baseline informed consent was obtained (see Figure 1 for CONSORT diagram). Participants were provided with the study website URL and personal log-in information.

**Figure 1.** CONSORT Diagram. CATCH-IT: Competent Adulthood Transition with Cognitive Behavioral Humanistic and Interpersonal Training; GAS: Global Assessment Scale; K-SADS: Kiddie Schedule for Affective Disorders and Schizophrenia; MI: motivational interview.



## Implementation

We previously published a report [20] on the implementation of this study in 30 participating clinics with 42,310 adolescents (data were not available on adolescents from 1 additional clinic). A total of 369 adolescents were randomized into the intervention or control program. The average reach (the proportion of target audience exposed to the intervention) for screening across the Chicago study sites was 0.216, and the average reach value for enrollment across all study sites was 0.181. We describe the effect of internal barriers to recruiting adolescents (eg, screening adolescents, transferring screening results to the research team) on the screening and enrollment process by reporting reach; these internal barriers, which were within the practices, are related to the implementation of universal screening [20].

## Randomization and Masking

Random assignment to CATCH-IT or health education was achieved via a computer-generated sequence blocked by site and time of entry. Participants were also stratified according to their gender, age (13-14 years; 15-18 years) [25,26], and the severity of their risk of depressive episode (current CES-D score or prior episode of major depressive disorder). Allocation to CATCH-IT versus health education was 1:1. Although condition assignment was kept from participants, families, clinicians, and

investigators until all initial assessments and preliminary data gathering were completed, it was not possible to blind the participants to their assignment, nor could the primary care clinician be blinded, as CATCH-IT implementation required clinicians to perform 3 motivational interviews over the course of the study. However, the 2 arms were presented to adolescents and parents as different approaches to improving adolescent health (eg, coping skills training or physical health improvement). Assessors were blinded for the duration of the study, and investigators were blinded until all 12-month data were collected.

## Treatment Arms

### The CATCH-IT Intervention

CATCH-IT, an internet-based depression prevention intervention delivered in primary care settings, sought to address gaps in the field of adolescent depression prevention: (1) the implementation of a low cost and low burden intervention in primary care where youth and their families usually receive care and (2) the efficacy of an intervention in reducing diagnosed depressive episodes (versus only symptom reduction) across a 24-month follow-up period [27,28]. CATCH-IT, originally developed in the United States, incorporates cognitive behavioral therapy and interpersonal psychotherapy and was developed to



have high levels of interactivity using previously published methods [27,28].

CATCH-IT integrates a set of sequential online modules with a coaching component consisting of direct interviews and periodic phone calls. There are 15 adolescent modules (14 modules on depression and an optional anxiety module with relaxation technique) designed to develop coping strategies protective against depression and 5 parent modules [29] (described in prior publications in detail) [14,15,29]. Each module takes approximately 15-20 minutes. Adapted from the Coping with Depression Adolescent Course, in combination with behavioral activation strategies (for example, behavioral scheduling of pleasurable activities) [30] and techniques used in interpersonal psychotherapy (for example, effective social problem solving, building and engaging social support) [31], CATCH-IT encourages the user to adopt a proactive approach toward mental health maintenance. Motivational support includes 3 formal motivational interviews delivered at baseline, 2 months, and 12 months by the participant's primary care physician, and 1 to 3 coaching phone calls by study staff. A separate CATCH-IT parent program, based on Beardslee and Gladstone's Preventive Intervention Project [32], included 4 modules that provided information about youth depression and ways to promote resilience in adolescents, and included a separate module for parents who think they may be depressed.

### ***The Health Education Intervention***

Health education, a 14-module online health educational tool, was used as an active control. Health education addresses adolescent health, wellness, and general safety. Module 14 discusses the identification of mood symptoms, the importance of seeking treatment for mental health concerns, and stigma associated with mental health treatment that might act as a barrier towards seeking professional help [33]; 4 modules for parents were included.

### ***Shared Elements Across Both Treatment Arms***

Both intervention arms provided care consistent with Guidelines for Adolescent Depression in Primary Care (GLAD-PC) [34], closely related to the chronic care model. This included screening adolescents for depression; training clinicians in the identification, diagnosis, and treatment of depression; establishing a treatment plan; creating referral relationships; and developing a safety plan [34]. Up to 3 check-in calls were made during the first 3 weeks of intervention to ensure participants had gained access to the website. We provide estimates of nonintervention time related to assessment and follow-up. These were developed based on total number of calls and sampling of call duration (40 randomly selected files to determine mean duration of calls and assessments).

### ***Outcomes***

The primary outcome was occurrence of first depressive episode, where a Depression Symptom Rating (DSR) score indicating at least subthreshold major depression ( $DSR \geq 3$ ) was considered a depressive episode. The secondary outcome was mental health functioning measured by GAS scores ([Multimedia Appendix 1: Table S10](#)).

## **Measurements**

### ***Demographics***

At baseline and during follow-up assessments, demographic information, educational status, and institution were collected from the participants and their families.

### ***Fidelity and Adherence***

Fidelity to the intervention was based on completion of the motivation interviews, quality of the motivation interviews as rated by the Motivational Interview Treatment Integrity (4.2.1) which ranges from 1 (low quality) to 5 (high quality) [35] and number of characters typed into the CATCH-IT website. We also measured modules completed and total time on the website by adolescents and parents. Additionally, at the 2-year time point, we asked 11 adolescents and parents 4 questions regarding satisfaction with the study and intervention.

### ***Depressive Episodes***

The 2-question screener was based on the Patient Health Questionnaire-Adolescent [24]. The K-SADS [36,37], a semistructured interview assessing current and lifetime psychiatric diagnoses in participants under age 18, was administered with parents and adolescents by licensed mental health professionals [36,38]. DSR scores were obtained from the K-SADS and the Kiddie Longitudinal Interval Follow-up Evaluation [38], which was used to identify each week of the follow-up interval's onset and offset of diagnoses based on recall, and GAS ratings were ascertained at each assessment. The K-SADS and GAS were completed at baseline and at 2, 6, 12, and 24 months. To test for robustness of findings, we examined outcomes using DSR cut-points of  $\geq 3$ ,  $\geq 4$ , and  $\geq 5$ , indicating probable major depressive episode [39]. We selected  $DSR \geq 3$  (at least subthreshold major depressive episode) versus  $DSR \geq 4$  (probable or definite major depressive episode) as our primary outcome due to the lower than expected occurrence of  $DSR \geq 4$  in the entire sample, identified during data safety monitoring board reviews of data. For assessment of presence of current or prior episodes (secondary versus primary prevention), we conducted an exploratory analysis of whether presence or absence of  $DSR \geq 3$  or  $\geq 4$  affected subsequent outcomes.

### ***Covariates in Moderation Analyses***

We have proposed and published the technology-based behavioral vaccine model as an integrated conceptual framework for understanding the prevention of depression across the lifespan through internet-based interventions in community settings [10,29,40]. Based on this model, we propose moderators within this conceptualization: (1) a life course schedule that is theory-driven and that includes booster doses operationalized as patient or participant factors influencing response over the life course, including demographics such as gender, race, ethnicity, age, parental education, and site; vulnerability factors (CES-D10, Adolescent Life Events Questionnaire, Beck Hopelessness Scale, Child Report of Parental Behavior Inventory); child comorbid psychopathology (Child Anxiety Related Emotional Disorders; Disruptive Behaviors Disorder Scale; substance abuse, using CRAFFT [Car, Relax, Alone, Forget, Friends, Trouble]); and parent psychopathology



(CES-D10); (2) effective components of information and training to encode responses to future threats that can then be deployed at some future points using information (operationalized as internet modules completed, time on site); (3) a motivational framework to boost response to behavior prescription (intrinsic motivation operationalized as Theory of Planned Behavior Scale, Transtheoretical Model Scale) as well as personal relevance (Sociocultural Relevance Scale); and (4) a structured implementation strategy as, to intervene effectively, we must address or provide proper conditions for successful participation in the intervention such as positive primary care relationships (operationalized Physician Relationship Scale) [11,14,30,41].

## Statistical Analysis

We estimated incidence rates by calculating the number of depressive episodes per 10,000 person-weeks of follow-up. Kaplan-Meier curves were used to estimate the time to first episode distribution for each intervention under 6 different treatment allocations: intention-to-treat, modified intention-to-treat (CATCH-IT adolescents who completed the baseline motivational interview), as treated (visited website at least once), per-protocol 2, per-protocol 4, and per-protocol 7 (adolescent and parent completed a total of 2, 4, 7 modules, respectively) (Multimedia Appendix 1: Table S1-2). Cox proportional hazard regression was used to estimate the hazard ratio comparing CATCH-IT to health education, adjusted for covariates. We present hazard ratios and coefficients adjusted for gender, ethnicity (Hispanic or non-Hispanic), race (White or non-White), baseline age, site (Boston or Chicago), and baseline CES-D10 score. We conducted exploratory analysis of the moderating effects of theory-based covariates by including interaction terms in the Cox models. We conducted the analysis of CES-D10 as a moderator in both the full sample and the subsample with elevated baseline CES-D10 scores. To examine whether the sizable number of moderator analyses we tested reveal significant effects more than by chance, we plotted all the ordered *P* values that were observed against what would be observed by chance in a quantile-quantile plot (Multimedia Appendix 1: Figure S1). We used linear mixed-effect growth models with random intercept and slope to examine differences between groups in change over time in GAS. We transformed the time scale in our growth models toward linearity by using a logarithm transformation of time. For the exploratory analysis of the impact of prior episodes on outcomes, standard descriptive statistics were used. Original sample size calculations to achieve 80% power were based on randomization of 200 into each arm and 28% incidence in the control condition at 12 months, a constant hazard ratio of 0.62, and total attrition of 36% of the sample [14]. Analyses were conducted using R (version 3.3.1), SAS (version 9.4; SAS Institute), and Mplus (version 8) [42].

## Results

### Participants

Of the 369 participants, 193 were randomized to CATCH-IT, and 176 were randomized to health education (Figure 1); depressive episode assessments (K-SADS) were available for 278 participants (75.3%) at 12 months and 181 (49.1%) at 24

months or greater (Multimedia Appendix 1: Table S3). The mean time observed was 524.8 days (SD 314.8 days) from baseline to final K-SADS. Predictors of missing episode assessment data were consistent at 12 and 24 months and included Chicago site, CATCH-IT group, older age, and lower parental education (Multimedia Appendix 1: Table S4-S5).

Adolescents were 13 to 18 years (mean 15.4 years, SD 1.5); 62% (233/3654) of adolescents had at least a prior subthreshold depressive episode ( $DSR \geq 3$ ), and 40% (147/3654) had at least a prior probable depressive episode ( $DSR \geq 4$ ). At baseline, the mean CES-D10 (0-30 scale) score was 9.4 (SD 4.6), the mean GAS score was 78.1 (SD 9.4), and the mean score on the Beck Hopelessness Scale was 4.7 (SD 3.6). Of the 369 participants, 16 (4.3%) adolescents reported recurrent thoughts (subthreshold or threshold) of death at baseline, 2 (0.5%) adolescents reported subthreshold suicidal ideation, and 2 (0.5%) adolescents reported subthreshold nonsuicidal self-harm thoughts. For parents at baseline, the mean CES-D10 score was 6.4 (SD 5.0), and 22.9% (72/314) were at least moderately depressed ( $CES-D20 \geq 16$ ). Adolescents were diverse in self-reported race and ethnicity: 159/369 (43.1%) identified as non-Hispanic White; 94/369 (25.5%) identified as non-Hispanic Black; and 77/369 (20.9%) identified as Hispanic. The parents of about half of the adolescents (mothers: 144/359, 40.1%; fathers: 157/336, 46.7%) had earned a high school diploma or less, and 39.4% (143/363) were not married. Relative to Chicago, the Boston site had more adolescents who identified as White ( $P < .001$ ), a lower percentage of ethnic minority adolescents ( $P = .01$ ), greater levels of parent education ( $P = .03$  for mothers and  $P < .001$  for fathers), more parents who were married at baseline ( $P = .03$ ), and parents who were older ( $P < .001$ ) (Multimedia Appendix 1). A complete description of the cohort data at baseline is provided in a prior publication [15].

### Fidelity and Adherence

Tests of fidelity and adherence have been described elsewhere [15] and are shown in Table 1 and Multimedia Appendix 1: Tables S11-S12. The median number of adolescent and parent modules completed was 4 in the CATCH-IT group and 8 in the health education group. Number of characters typed was used to measure active use of the modules, with adolescents and parents (combined) in the CATCH-IT arm typing a mean of 3713 (median 1899) characters. Adolescents spent more time on the CATCH-IT site than on the health education site (CATCH-IT: median 39.6 minutes; health education: median 8.4 minutes) but completed more modules on health education (CATCH-IT: median 1.0; health education: median 4.0). For CATCH-IT, 14 adolescents (14/193, 7.3%) completed all 15 modules; for health education, 74 adolescents (74/176, 42.0%) completed all 14 modules.

The median number of motivational interviews completed was 2 for both adolescents and parents, with 94.8% (183/193) of adolescents completing at least 1 and 34.2% (66/193) completing all 3 (Table 1 and Multimedia Appendix 1: Table S11). Motivational interviews were assessed for fidelity by 2 trained raters using the Motivation Interview Treatment Integrity 4.2.1 coding manual. The mean length of motivation interview was 7.7 minutes (SD 4.0), and mean lengths of technical and

relational global ratings were 3.0 (SD 0.5) and 2.9 (SD 0.6), respectively ([Multimedia Appendix 1](#): Table S11). Based on a log recording telephone and email contacts with Chicago participants ([Multimedia Appendix 1](#): Table S12), a mean of 17.0 (SD 8.9) and 14.1 (SD 6.5) contacts were made with CATCH-IT and health education participants, respectively. We estimated mean total time required to complete assessment and follow-up common to both arms with GLAD-PC model and total time specific for the intervention (101 minutes for

assessment and follow-up common to both arms, 171 minutes for CATCH-IT and 31 minutes for health education, and in total contact time, assessment and follow-up plus intervention time, 272 minutes versus 132 minutes for CATCH-IT and health education, respectively) ([Multimedia Appendix 1](#): Table S13). With regard to the end of study satisfaction questions, adolescents noted they valued the periodic assessment calls above all other study elements, regardless of study arm.

**Table 1.** Website use, motivational interviews, and participant contacts.

Outcomes	CATCH-IT <sup>a</sup> , mean (SD)	Health education, mean (SD)	CATCH-IT, median (IQR)	Health education, median (IQR)	P value <sup>b</sup>
<b>Website use</b>					
<b>Adolescents<sup>c</sup></b>					
Modules completed	3.4 (4.7)	6.8 (6.5)	1 (4)	4 (14)	.003
Total minutes on site	100.2 (143.1)	22.8 (31.0)	39.6 (149.2)	8.4 (35.1)	<.001
Days visited site	3.7 (4.5)	1.4 (1.6)	2 (4)	1 (2)	<.001
Total characters typed	3071 (4572)	N/A <sup>d</sup>	923 (4469)	N/A	N/A
<b>Parents<sup>e</sup></b>					
Modules completed	2.1 (2.0)	2.2 (1.9)	2 (4)	4 (4)	.80
Total minutes on site	32.6 (37.3)	8.6 (10.0)	22.4 (51.9)	5.6 (14.9)	<.001
Days visited site	1.6 (1.6)	0.9 (1.1)	1 (2)	1 (1)	<.001
Total characters typed	716 (977)	N/A	101 (1205)	N/A	N/A
<b>Adolescents and parents</b>					
Modules completed	5.3 (5.8)	8.8 (7.3)	4 (8)	8 (17)	<.001
Total minutes on site	130.6 (157.9)	30.6 (35.6)	75.8 (192.2)	18.9 (40.8)	<.001
Days visited site	5.2 (5.2)	2.2 (2.2)	4 (6)	2 (2)	<.001
Total characters typed	3713 (4932)		1899 (5792)		
<b>Motivational interviews<sup>f</sup></b>					
Adolescents	2.0 (0.9)	N/A	2 (2)	N/A	N/A
Parents	1.9 (0.9)	N/A	2 (2)	N/A	N/A
Adolescents and parents	3.8 (1.8)	N/A	4 (4)	N/A	N/A
Participant contacts <sup>g</sup>	17 (9)	N/A	18 (13)	N/A	.004

<sup>a</sup>CATCH-IT: Competent Adulthood Transition with Cognitive Behavioral Humanistic and Interpersonal Training.

<sup>b</sup>Medians were compared using Wilcoxon rank-sum test.

<sup>c</sup>n=193 for the CATCH-IT group; n=176 for the health education group.

<sup>d</sup>N/A: not applicable.

<sup>e</sup>n=165 for the CATCH-IT group; n=157 for the health education group.

<sup>f</sup>CATCH-IT adolescents and parents were offered 3 motivational interviews; see [Multimedia Appendix 1](#) for additional information.

<sup>g</sup>n=248 telephone calls and emails with Chicago participants, including administrative contacts such as scheduling calls in addition to coaching calls, safety checks, and motivational interview calls. See [Multimedia Appendix 1](#) for additional information.

## Time to Depressive Episode

In intention-to-treat analyses, the hazard ratio favored CATCH-IT for occurrence of first depressive episode but was statistically significant neither at 12 months (hazard ratio 0.77, 95% CI 0.42-1.40,  $P=.39$ ; [Multimedia Appendix 1](#): Table S1) nor at 24 months (hazard ratio 0.87, 95% CI 0.52-1.47,  $P=.61$ ,

[Multimedia Appendix 1](#): Table S2). For per-protocol 2 analysis (at least 2 modules completed by adolescent and parent combined in either arm), the effect of CATCH-IT on depressive episode prevention was improved but remained nonsignificant at 12 months (hazard ratio 0.65, 95% CI 0.34-1.23,  $P=.18$ , [Multimedia Appendix 1](#): Table S1) and 24 months (hazard ratio 0.73, 95% CI 0.41-1.28,  $P=.27$ ; [Multimedia Appendix 1](#): Table

S2). Annual incidence of first at least subthreshold depressive episode ( $DSR \geq 3$ ) was 12.7% for CATCH-IT and 13.8% for health education; incidence of probable major depressive episode ( $DSR \geq 4$ ) was 4.6% for CATCH-IT and 5.8% for health education ([Multimedia Appendix 1: Table S6](#)).

### Functional Status Improvement

GAS scores improved in each group from baseline to 24 months ([Multimedia Appendix 1: Table S17](#)). For GAS with time logarithm transformed, the slope was 2.964 (SE 0.285;  $P < .001$ ), and adjusted change from baseline to 24 months was 9.5 (SE 0.9) for CATCH-IT, compared to a slope of 3.076 (SE 0.278;  $P < .001$ ) and change of 9.9 (SE 0.9) for health education. There was no significant difference in slopes between groups ( $P = .78$ ) ([Multimedia Appendix 1: Table S17](#)).

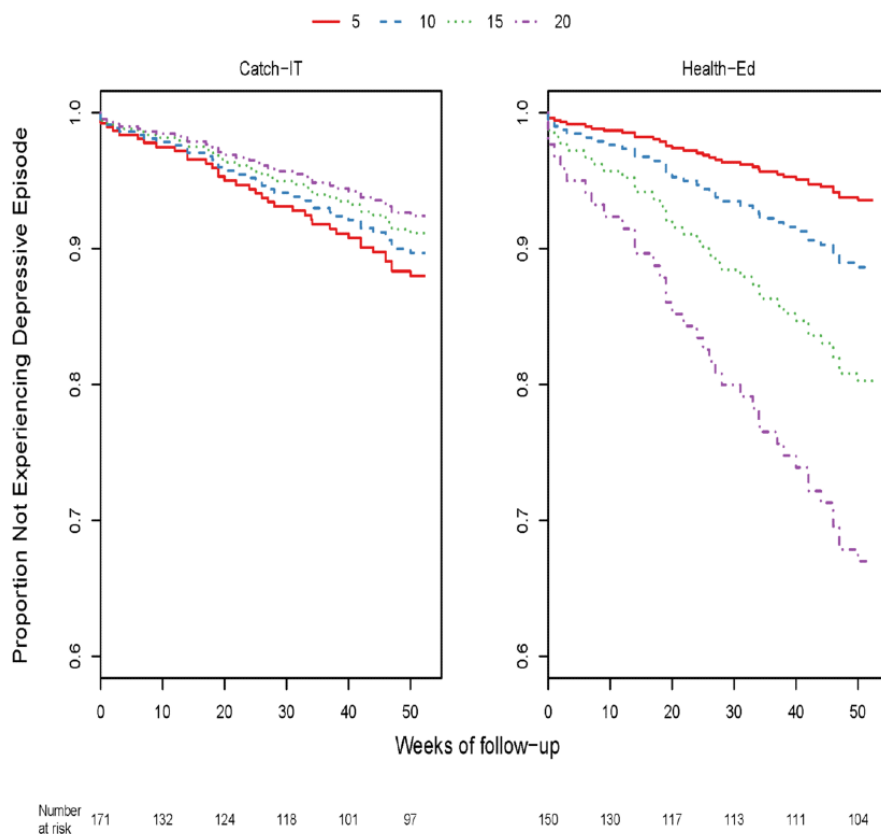
### Moderation Analyses

Through 12 months, within the group of adolescents who enrolled based on elevated CES-D (321/369, 86.9%) with or without a prior depressive episode, those with higher baseline CES-D10 scores exhibited a marginally stronger preventive effect of CATCH-IT on time to event ( $DSR \geq 3$ ) relative to those with lower baseline scores (0.15 decline in the logarithm of hazard ratio for each CES-D10 point increase,  $P = .054$ ) ([Multimedia Appendix 1: Table S7, Figure 2](#)). Those who also visited the website at least once (as treated) (253/369, 68.6%) showed a similar effect (0.18 decline in the logarithm of hazard ratio for each CES-D10 point increase,  $P = .03$ ). Baseline CES-D10 was no longer a significant moderator at 24 months, either in the sample as a whole or in the subgroups tested. However, 2 additional moderators were identified as significant on time to event through 24 months ( $DSR \geq 3$ ). Specifically, there was a complex interaction between baseline hopelessness (270/369, 73.2%) whereby, at lower Beck Hopelessness Scale scores, CATCH-IT demonstrated a preventive effect, but results

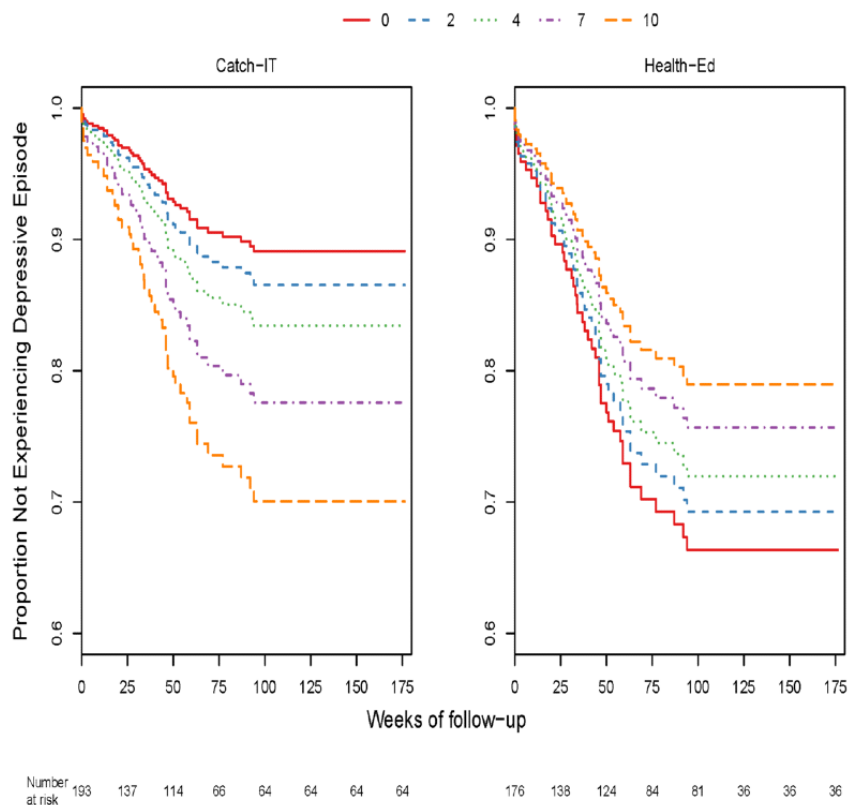
favoring health education at higher scores ( $P = .04$ ) ([Multimedia Appendix 1: Table S8, Figure 3](#)). Similarly, there was a complex interaction between baseline paternal monitoring (170/369, 46.1%) and group assignment, such that at higher levels of paternal monitoring at baseline, CATCH-IT demonstrated preventive effects on time to event ( $DSR \geq 3$ ) but favored health education at lower paternal monitoring ( $P = .048$ , [Multimedia Appendix 1: Table S9, Figure 4](#)). Site, race, ethnicity, and gender did not moderate outcomes. We conducted an overall comparison of the 12-month and 24-month moderation effects to see whether there was any signal evident beyond what one would expect from a null distribution. [Multimedia Appendix 1: Figure S1](#) provides a comparison of all the ordered  $P$  values against this standard. The moderation effects for hopelessness and paternal monitoring did not stand out against the null distribution. No GAS moderation analyses yielded statistically significant results after time transformation (all  $P > .05$  after logarithm transformation, [Multimedia Appendix 1: Table S18](#)).

Assessment of presence of current or prior episodes (secondary versus primary prevention) showed there was no meaningful variation in the presence of current or past  $DSR \geq 3$  or  $\geq 4$  across groups at baseline ([Multimedia Appendix 1: Table S14](#)). With the main outcome threshold of  $DSR \geq 3$ , about two-thirds of the sample could be characterized as secondary prevention and one-third primary prevention ([Multimedia Appendix 1: Table S14](#)). However, most of the episodes of  $DSR \geq 3$  during follow-up were new episodes; only 4 episodes were present or persisting at 2 months in each study arm ([Multimedia Appendix 1: Table S15](#)). Adolescents with past or current episodes at baseline were statistically equally likely to have a follow-up episode regardless of group assignment. For adolescents with a current or past episode with  $DSR \geq 3$  at baseline, 17.9% of CATCH-IT and 19.8% of health education participants had an episode during the follow-up period ([Multimedia Appendix 1: Table S16](#)).

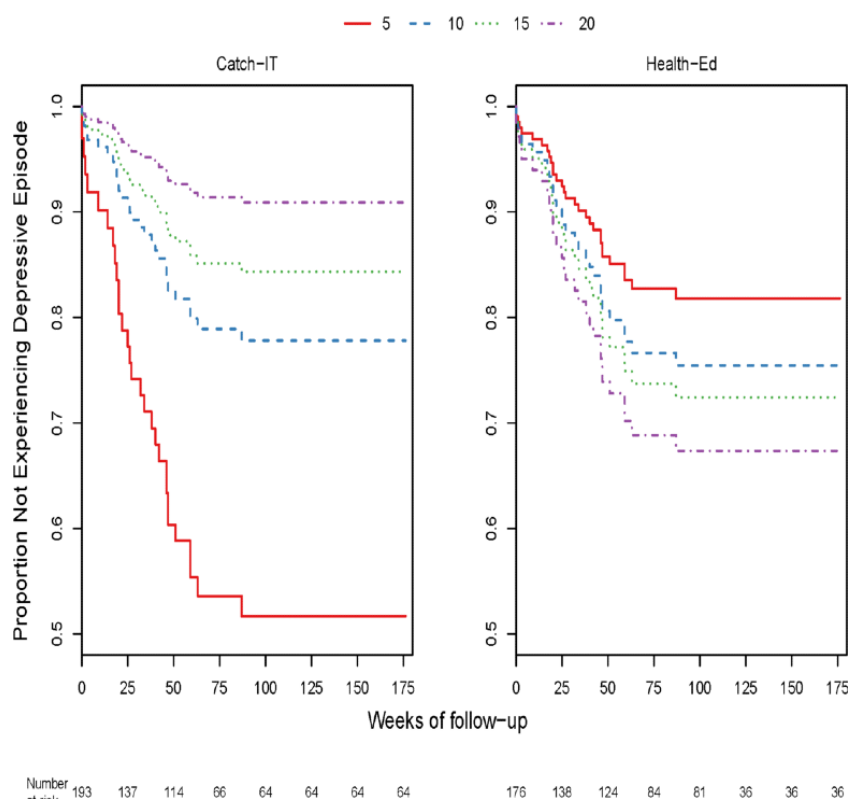
**Figure 2.** Adolescent CES-D Scale score at baseline and time to Depression Symptom Rating $>3$ —event at 12 months, excluding those who enrolled with no depressed mood.



**Figure 3.** Adolescent Beck Hopelessness Scale score at baseline and time to Depression Symptom Rating $>3$ —event at 24 months.



**Figure 4.** Adolescent Child Report of Parental Behaviour Inventory (Father Monitoring Subscale) score at baseline and time to Depression Symptom Rating $\geq 3$ —event at 24 months.



## Discussion

This hybrid Type 1 effectiveness-implementation clinical trial [19] conducted in primary care practice settings (31 sites) with a diverse sample is the only study, to our knowledge, with over 12 months of follow-up for adolescents enrolled in an internet-based study [13,43,44]. At 12 and 24 months postintervention using Cox proportional hazard regression analysis, we found that CATCH-IT was not superior to health education in preventing the onset of depressive episodes. The annual incidence of major depressive episodes was fairly low for both groups (DSR $\geq 4$ : 4.6% CATCH-IT, 5.8% health education; DSR $\geq 3$ : 13.7% CATCH-IT, 12.7%). This compares favorably with the 12-month incidence of major depressive episodes (DSR $\geq 4$ ) in the United States, which is 8.3% [1] in general samples and >15% in similar risk-adjusted samples; these rates are roughly half the rates for minor plus major depressive episodes in high-risk samples (DSR $\geq 3$ ) [1,21,23]. Moderated analyses suggested that adolescents who enroll with higher CES-D scores, lower levels of hopelessness, and higher paternal monitoring may obtain preventive benefit from CATCH-IT for 12 months, and possibly for as long as two years postenrollment.

There are several potential explanations for why this trial failed to show a main effect of the CATCH-IT intervention. First, the incidence of depressive episodes was much lower than anticipated for this high-risk sample, and remained lower than expected even after lowering the threshold of primary outcome from probable depressive episode (DSR  $\geq 4$ ) to subthreshold depressive episode (DSR $\geq 3$ ). An important factor may be the

amount of assessment follow-up time common to both arms (150 minutes) within the framework of the GLAD-PC guidelines [36], which was nearly equal to the total intervention time for CATCH-IT (170 minutes), rather than any particular effect of the health education intervention (30 minutes). It is possible that this level of assessment and follow-up suppressed the level of depressive episodes in both arms, depriving the CATCH-IT intervention of some of the potential to prevent episodes. The demonstrated intervention of a Chinese language CATCH-IT at 12 months which had only self-assessments but did have an attention control is consistent with this explanation [45], as is the significant improvement in functional status over time in both groups. Another explanation would be that, in fact, the active control reduced between group differences since most adolescent studies with active controls have not demonstrated between group differences [46,47]. The recruitment of slightly less than the target sample size (92%), higher attrition (>50% versus 36% anticipated), weaker intervention effect (hazard ratio 0.87 versus predicted hazard ratio 0.62), and lower dose (hazard ratio more favorable for those with higher dose) could all play roles in the nonsignificant results in intention-to-treat analyses. Perhaps a dose threshold of 2-4 modules (including cognitive restructuring, behavioral activation) is essential, as occurred in the Chinese CATCH-IT trial [45] and other studies [48-50]. Additionally, the fact that some teens seem to respond more favorably to the health education intervention (those with no current symptoms), while others to CATCH-IT (those with symptoms), reduced between group differences.

Many have called for examination of moderators for treatment tailoring and to contribute to future meta-analyses parsing out



differential treatment outcomes [51]. Moderated results suggest that CATCH-IT may benefit a significant portion of youth, such as groups of youth with elevated depressive symptoms. For example, adolescents who were enrolled with no current depressed mood but with a prior depressive episode fared worse with CATCH-IT, while those with depressed mood may benefit for as long as 12 months. Similarly, adolescents with lower hopelessness may benefit from CATCH-IT, while those with higher levels may not benefit from such cognitive behavioral therapy-based interventions [52]. It is possible that the context of social support may matter, such that CATCH-IT, which requires youth effort (eg, completion of cognitive behavioral therapy homework), may offer preventive benefit for those who report adequate paternal monitoring, consistent with Mohr's supportive accountability model [8]. One might speculate that more paternal involvement in the health education intervention may be overbearing because less accountability is needed. While female gender predicted higher risk of episodes, results did not vary by gender nor did they vary by ethnicity. The results of this trial along with those from prior trials [7] suggest that interventions can prevent depressive episode onset and worsening of depressive symptoms in youth subpopulations.

This study has important limitations and strengths. While it lacked an inactive or wait-list control, wait-list control is less ethical, as effective interventions are available. We cannot deconstruct the effects of motivation interview versus the online modules, but depressive symptoms decreased for adolescents who used CATCH-IT even without a motivation interview in our pilot study [12]. Differential loss to follow-up for those in the CATCH-IT group and at the Chicago site is concerning. However, predictors of increased risk of episodes, such as depressed mood or female gender, did not predict loss to follow-up, and make it unlikely that hazard ratios or incidence of episodes were affected by dropout patterns. Measurement of depressive episodes is challenging, and measurement issues could explain lower than expected rates. However, both sites were trained by an experienced evaluation team with methods

used in prior clinical trials [39]. Despite these limitations, this study has several strengths. The standardized implementation across several primary care settings supports feasibility of this model of care, though it will be important to determine potential logistical barriers such as check-in calls to participants. The examination of depressive episodes, long-term follow up, active control, and blinding of outcome assessors address methodological limitations of prior studies.

Future investigators should exercise caution in the design of low-intensity technology-based preventive interventions when considering active controls and extensive assessment batteries. Rather than an active control, an informational brochure may be a better option for a control; likewise, rather than using the KSADS to gather diagnostic information, the much shorter Mini-International Neuropsychiatric Interview for Children and Adolescents could be considered. Similarly, careful stakeholder-grounded intervention design (additional incentives or complementary design features, for example, online discussion groups or animation) may be needed to induce adolescent-parent pairs to complete more modules and thus strengthen the preventive effect to ensure minimum dose levels in the target population. Moderated results suggest that life course factors (depressed mood) and family factors (paternal monitoring) may be related to the potential for efficacy beyond 6 months. For example, a general health promotion model intervention may be more appropriate for adolescents at risk for a major depressive episode but currently with minimal symptoms, or conversely, CATCH-IT may not benefit adolescents with high levels of hopelessness and low levels of paternal monitoring. In short, we have a great deal to learn about how to implement technology-based depression prevention interventions and for whom they will work, in primary care and community settings. Future research should investigate how to best tailor online interventions to the characteristics of teens at risk of depression in order to optimize outcomes and prevent the development of major depression.

## Acknowledgments

In memoriam, CB passed away on August 2, 2019. An internationally renowned pioneer of community mental health, CB had a lifelong commitment to the south side of Chicago. This paper, which is one of his more than 500 scholarly texts, adds to his seminal work on the power of resilience in the face of adversity and his advocacy of prevention. CB was a role model for many, including the other authors of this paper who will strive to carry on his legacy. CB mentored the senior author for 16 years; without the insight, support, and wisdom of CB, this project could never have been completed. CB is sorely missed.

Research reported in this article was supported by the National Institute of Mental Health of the National Institutes of Health under award number R01MH090035. The implementation process was developed with funding from Robert Wood Johnson Foundation. The funders had no involvement in study design, data collection and analysis, and preparation and submission of the manuscript. The interest of Asok K Ray, MD FRCS (Edin), Purnima Ray, and Malika Ray in this manuscript and research is appreciated.

Ethical bodies that approved this study include institutional review boards from Wellesley College, University of Illinois, Advocate Health Care, Franciscan St. Mary, Northwestern, and NorthShore University Health System.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Supplemental tables and figure.

[\[DOCX File , 191 KB - jmir\\_v22i10e16802\\_app1.docx\]](#)

## References

1. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med* 2011 Jul 26;9:90 [FREE Full text] [doi: [10.1186/1741-7015-9-90](https://doi.org/10.1186/1741-7015-9-90)] [Medline: [21791035](https://pubmed.ncbi.nlm.nih.gov/21791035/)]
2. Janssens KAM, Rosmalen JGM, Ormel J, van Oort FVA, Oldehinkel AJ. Anxiety and depression are risk factors rather than consequences of functional somatic symptoms in a general population of adolescents: the TRAILS study. *J Child Psychol Psychiatry* 2010 Mar;51(3):304-312. [doi: [10.1111/j.1469-7610.2009.02174.x](https://doi.org/10.1111/j.1469-7610.2009.02174.x)] [Medline: [19788552](https://pubmed.ncbi.nlm.nih.gov/19788552/)]
3. Pettit JW, Lewinsohn PM, Roberts RE, Seeley JR, Monteith L. The long-term course of depression: development of an empirical index and identification of early adult outcomes. *Psychol Med* 2009 Mar;39(3):403-412 [FREE Full text] [doi: [10.1017/S0033291708003851](https://doi.org/10.1017/S0033291708003851)] [Medline: [18606049](https://pubmed.ncbi.nlm.nih.gov/18606049/)]
4. Weissman MM, Wolk S, Goldstein RB, Moreau D, Adams P, Greenwald S, et al. Depressed adolescents grown up. *JAMA* 1999 May 12;281(18):1707-1713. [doi: [10.1001/jama.281.18.1707](https://doi.org/10.1001/jama.281.18.1707)] [Medline: [10328070](https://pubmed.ncbi.nlm.nih.gov/10328070/)]
5. Rushton J, Bruckman D, Kelleher K. Primary care referral of children with psychosocial problems. *Arch Pediatr Adolesc Med* 2002 Jun;156(6):592-598. [doi: [10.1001/archpedi.156.6.592](https://doi.org/10.1001/archpedi.156.6.592)] [Medline: [12038893](https://pubmed.ncbi.nlm.nih.gov/12038893/)]
6. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. *J Child Psychol Psychiatry* 2017 Apr;58(4):474-503. [doi: [10.1111/jcpp.12663](https://doi.org/10.1111/jcpp.12663)] [Medline: [27943285](https://pubmed.ncbi.nlm.nih.gov/27943285/)]
7. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. *J Youth Adolesc* 2015 Jan;44(1):90-113. [doi: [10.1007/s10964-014-0165-0](https://doi.org/10.1007/s10964-014-0165-0)] [Medline: [25115460](https://pubmed.ncbi.nlm.nih.gov/25115460/)]
8. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *J Med Internet Res* 2011 Mar 10;13(1):e30 [FREE Full text] [doi: [10.2196/jmir.1602](https://doi.org/10.2196/jmir.1602)] [Medline: [21393123](https://pubmed.ncbi.nlm.nih.gov/21393123/)]
9. Hetrick SE, Cox GR, Witt KG, Bir JJ, Merry SN. Cognitive behavioural therapy (CBT), third-wave CBT and interpersonal therapy (IPT) based interventions for preventing depression in children and adolescents. *Cochrane Database Syst Rev* 2016 Aug 09(8):CD003380. [doi: [10.1002/14651858.CD003380.pub4](https://doi.org/10.1002/14651858.CD003380.pub4)] [Medline: [27501438](https://pubmed.ncbi.nlm.nih.gov/27501438/)]
10. Abuwalla Z, Clark MD, Burke B, Tannenbaum V, Patel S, Mitacek R, et al. Long-term telemental health prevention interventions for youth: A rapid review. *Internet Interv* 2018 Mar;11:20-29 [FREE Full text] [doi: [10.1016/j.invent.2017.11.006](https://doi.org/10.1016/j.invent.2017.11.006)] [Medline: [30135756](https://pubmed.ncbi.nlm.nih.gov/30135756/)]
11. Van Voorhees BW, Fogel J, Reinecke MA, Gladstone T, Stuart S, Gollan J, et al. Randomized clinical trial of an Internet-based depression prevention program for adolescents (Project CATCH-IT) in primary care: 12-week outcomes. *J Dev Behav Pediatr* 2009 Feb;30(1):23-37 [FREE Full text] [doi: [10.1097/DBP.0b013e3181966c2a](https://doi.org/10.1097/DBP.0b013e3181966c2a)] [Medline: [19194326](https://pubmed.ncbi.nlm.nih.gov/19194326/)]
12. Saulsberry A, Marko-Holguin M, Blomeke K, Hinkle C, Fogel J, Gladstone T, et al. Randomized clinical trial of a primary care internet-based intervention to prevent adolescent depression: one-year outcomes. *J Can Acad Child Adolesc Psychiatry* 2013 May;22(2):106-117 [FREE Full text] [Medline: [23667356](https://pubmed.ncbi.nlm.nih.gov/23667356/)]
13. Richards K, Marko-Holguin M, Fogel J, Anker L, Ronayne J, Van Voorhees BW. Randomized clinical trial of an internet-based intervention to prevent adolescent depression in a primary care setting (CATCH-IT): 2.5-year outcomes. *J Evid Based Psychother* 2016 Sep;16(2):113-134 [FREE Full text] [Medline: [30345428](https://pubmed.ncbi.nlm.nih.gov/30345428/)]
14. Gladstone TG, Marko-Holguin M, Rothberg P, Nidetz J, Diehl A, DeFrino DT, et al. An internet-based adolescent depression preventive intervention: study protocol for a randomized control trial. *Trials* 2015 May 01;16:203 [FREE Full text] [doi: [10.1186/s13063-015-0705-2](https://doi.org/10.1186/s13063-015-0705-2)] [Medline: [25927539](https://pubmed.ncbi.nlm.nih.gov/25927539/)]
15. Gladstone T, Terrizzi D, Stinson A, Nidetz J, Canel J, Ching E, et al. Effect of internet-based cognitive behavioral humanistic and interpersonal training vs. internet-based general health education on adolescent depression in primary care: a randomized clinical trial. *JAMA Netw Open* 2018 Nov;1(7):1-15 [FREE Full text] [doi: [10.1001/jamanetworkopen.2018.4278](https://doi.org/10.1001/jamanetworkopen.2018.4278)] [Medline: [30533601](https://pubmed.ncbi.nlm.nih.gov/30533601/)]
16. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Rounsaville BJ. Enduring effects of a computer-assisted training program for cognitive behavioral therapy: a 6-month follow-up of CBT4CBT. *Drug Alcohol Depend* 2009 Feb 01;100(1-2):178-181 [FREE Full text] [doi: [10.1016/j.drugalcdep.2008.09.015](https://doi.org/10.1016/j.drugalcdep.2008.09.015)] [Medline: [19041197](https://pubmed.ncbi.nlm.nih.gov/19041197/)]
17. Hartmann A, von Wietersheim J, Weiss H, Zeeck A. Patterns of symptom change in major depression: classification and clustering of long term courses. *Psychiatry Res* 2018 Sep;267:480-489. [doi: [10.1016/j.psychres.2018.03.086](https://doi.org/10.1016/j.psychres.2018.03.086)] [Medline: [29980128](https://pubmed.ncbi.nlm.nih.gov/29980128/)]
18. von Brachel R, Hirschfeld G, Berner A, Willutzki U, Teismann T, Cwik JC, et al. Long-term effectiveness of cognitive behavioral therapy in routine outpatient care: a 5- to 20-year follow-up study. *Psychother Psychosom* 2019;88(4):225-235. [doi: [10.1159/000500188](https://doi.org/10.1159/000500188)] [Medline: [31121580](https://pubmed.ncbi.nlm.nih.gov/31121580/)]
19. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care* 2012 Mar;50(3):217-226 [FREE Full text] [doi: [10.1097/MLR.0b013e3182408812](https://doi.org/10.1097/MLR.0b013e3182408812)] [Medline: [22310560](https://pubmed.ncbi.nlm.nih.gov/22310560/)]

20. Mahoney N, Gladstone T, DeFrino D, Stinson A, Nidetz J, Canel J, et al. Prevention of adolescent depression in primary care: barriers and relational work solutions. *Calif J Health Promot* 2017 Aug;15(2):1-12 [[FREE Full text](#)] [Medline: [30393470](#)]
21. Radloff LS. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. *J Youth Adolescence* 1991 Apr;20(2):149-166. [doi: [10.1007/bf01537606](#)]
22. Vilagut G, Forero CG, Barbaglia G, Alonso J. Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): a systematic review with meta-analysis. *PLoS One* 2016;11(5):e0155431 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0155431](#)] [Medline: [27182821](#)]
23. Van Voorhees BW, Paunesku D, Gollan J, Kuwabara S, Reinecke M, Basu A. Predicting future risk of depressive episode in adolescents: the Chicago Adolescent Depression Risk Assessment (CADRA). *Ann Fam Med* 2008;6(6):503-511. [doi: [10.1370/afm.887](#)] [Medline: [19001302](#)]
24. Johnson JG, Harris ES, Spitzer RL, Williams JBW. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. *J Adolesc Health* 2002 Mar;30(3):196-204. [doi: [10.1016/s1054-139x\(01\)00333-0](#)] [Medline: [11869927](#)]
25. Pastore DR, Juszczak L, Fisher MM, Friedman SB. School-based health center utilization: a survey of users and nonusers. *Arch Pediatr Adolesc Med* 1998 Aug;152(8):763-767. [doi: [10.1001/archpedi.152.8.763](#)] [Medline: [9701135](#)]
26. Piantadosi S. *Clinical Trials: A Methodologic Perspective*. New York, NY: John Wiley & Sons; 2017.
27. Gagne R, Briggs L, Wager W. *Principles of Instructional Design*. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers; 1992.
28. Van Voorhees BW, Watson N, Bridges JFP, Fogel J, Galas J, Kramer C, et al. Development and pilot study of a marketing strategy for primary care/internet-based depression prevention intervention for adolescents (the CATCH-IT intervention). *Prim Care Companion J Clin Psychiatry* 2010;12(3):e1-e9 [[FREE Full text](#)] [doi: [10.4088/PCC.09m00791blu](#)] [Medline: [20944776](#)]
29. Van Voorhees BW, Gladstone T, Cordel S, Marko-Holguin M, Beardslee W, Kuwabara S, et al. Development of a technology-based behavioral vaccine to prevent adolescent depression: A health system integration model. *Internet Interv* 2015 Sep;2(3):303-313 [[FREE Full text](#)] [doi: [10.1016/j.invent.2015.07.004](#)] [Medline: [30473992](#)]
30. Jacobson N, Martell C, Dimidjian S. Behavioral activation treatment for depression: returning to contextual roots. *Clinical Psychology-Science and Practice* 2001;8(3):255-270. [doi: [10.1093/clipsy/8.3.255](#)]
31. Robertson M. *Interpersonal Psychotherapy-A Clinician's Guide*. London: CRC Press; 2003.
32. Beardslee WR, Gladstone TRG, Wright EJ, Cooper AB. A family-based approach to the prevention of depressive symptoms in children at risk: evidence of parental and child change. *Pediatrics* 2003 Aug;112(2):e119-e131. [doi: [10.1542/peds.112.2.e119](#)] [Medline: [12897317](#)]
33. Van Voorhees BW, Fogel J, Houston TK, Cooper LA, Wang NY, Ford DE. Beliefs and attitudes associated with the intention to not accept the diagnosis of depression among young adults. *Ann Fam Med* 2005;3(1):38-46. [doi: [10.1370/afm.273](#)] [Medline: [15671189](#)]
34. Zuckerbrot RA, Cheung A, Jensen PS, Stein REK, Laraque D, GLAD-PC STEERING GROUP. Guidelines for Adolescent Depression in Primary Care (GLAD-PC):part i. practice preparation, identification, assessment, and initial management. *Pediatrics* 2018 Dec;141(3):e20174081. [doi: [10.1542/peds.2017-4081](#)] [Medline: [29483200](#)]
35. Moyers TB, Rowell LN, Manuel JK, Ernst D, Houck JM. The Motivational Interviewing Treatment Integrity Code (MITI 4): rationale, preliminary reliability and validity. *J Subst Abuse Treat* 2016 Jun;65:36-42 [[FREE Full text](#)] [doi: [10.1016/j.jsat.2016.01.001](#)] [Medline: [26874558](#)]
36. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 1997 Jul;36(7):980-988. [doi: [10.1097/00004583-199707000-00021](#)] [Medline: [9204677](#)]
37. Kaufman J, Birmaher B, Brent DA, Ryan ND, Rao U. K-SADS-PL. *J Am Acad Child Adolesc Psychiatry* 2000 Oct;39(10):1208. [doi: [10.1097/00004583-200010000-00002](#)] [Medline: [11026169](#)]
38. Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, McDonald-Scott P, et al. The Longitudinal Interval Follow-up Evaluation. A comprehensive method for assessing outcome in prospective longitudinal studies. *Arch Gen Psychiatry* 1987 Jun;44(6):540-548. [doi: [10.1001/archpsyc.1987.01800180050009](#)] [Medline: [3579500](#)]
39. Garber J, Clarke GN, Weersing VR, Beardslee WR, Brent DA, Gladstone TRG, et al. Prevention of depression in at-risk adolescents: a randomized controlled trial. *JAMA* 2009 Jun 03;301(21):2215-2224 [[FREE Full text](#)] [doi: [10.1001/jama.2009.788](#)] [Medline: [19491183](#)]
40. Van Voorhees BW, Mahoney N, Mazo R, Barrera AZ, Siemer CP, Gladstone TRG, et al. Internet-based depression prevention over the life course: a call for behavioral vaccines. *Psychiatr Clin North Am* 2011 Mar;34(1):167-183 [[FREE Full text](#)] [doi: [10.1016/j.psc.2010.11.002](#)] [Medline: [21333846](#)]
41. Van Voorhees BW, Vanderplough-Booth K, Fogel J, Gladstone T, Bell C, Stuart S, et al. Integrative internet-based depression prevention for adolescents: a randomized clinical trial in primary care for vulnerability and protective factors. *J Can Acad Child Adolesc Psychiatry* 2008 Nov;17(4):184-196 [[FREE Full text](#)] [Medline: [19018321](#)]

42. Muthén LK, Muthén BO. Statistical analysis with latent variables. Mplus. URL: [https://statmodel.com/download/usersguide/MplusUserGuideVer\\_8.pdf](https://statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf) [accessed 2020-09-01]
43. Booth KVP, Paunesku D, Msall M, Fogel J, Van Voorhees BW. Using population attributable risk to help target preventive interventions for adolescent depression. *Int J Adolesc Med Health* 2008;20(3):307-319. [doi: [10.1515/ijamh.2008.20.3.307](https://doi.org/10.1515/ijamh.2008.20.3.307)] [Medline: [19097569](https://pubmed.ncbi.nlm.nih.gov/19097569/)]
44. Kruger J, Kim P, Iyer V, Marko-Holguin M, Fogel J, DeFrino D, et al. Evaluation of protective and vulnerability factors for depression following an internet-based intervention to prevent depression in at-risk adolescents. *International Journal of Mental Health Promotion* 2017 Apr 19;19(2):69-84. [doi: [10.1080/14623730.2017.1308264](https://doi.org/10.1080/14623730.2017.1308264)]
45. Ip P, Chim D, Chan KL, Li TMH, Ho FKW, Van Voorhees BW, et al. Effectiveness of a culturally attuned Internet-based depression prevention program for Chinese adolescents: A randomized controlled trial. *Depress Anxiety* 2016 Dec;33(12):1123-1131. [doi: [10.1002/da.22554](https://doi.org/10.1002/da.22554)] [Medline: [27618799](https://pubmed.ncbi.nlm.nih.gov/27618799/)]
46. Rasing SPA, Creemers DHM, Janssens JMAM, Scholte RHJ. Depression and anxiety prevention based on cognitive behavioral therapy for at-risk adolescents: a meta-analytic review. *Front Psychol* 2017;8:1066 [FREE Full text] [doi: [10.3389/fpsyg.2017.01066](https://doi.org/10.3389/fpsyg.2017.01066)] [Medline: [28701980](https://pubmed.ncbi.nlm.nih.gov/28701980/)]
47. Weersing VR, Jeffreys M, Do MT, Schwartz KTG, Bolano C. Evidence base update of psychosocial treatments for child and adolescent depression. *J Clin Child Adolesc Psychol* 2017;46(1):11-43 [FREE Full text] [doi: [10.1080/15374416.2016.1220310](https://doi.org/10.1080/15374416.2016.1220310)] [Medline: [27870579](https://pubmed.ncbi.nlm.nih.gov/27870579/)]
48. Clarke G, Reid E, Eubanks D, O'Connor E, DeBar LL, Kelleher C, et al. Overcoming depression on the Internet (ODIN): a randomized controlled trial of an internet depression skills intervention program. *J Med Internet Res* 2002 Dec;4(3):E14 [FREE Full text] [doi: [10.2196/jmir.4.3.e14](https://doi.org/10.2196/jmir.4.3.e14)] [Medline: [12554545](https://pubmed.ncbi.nlm.nih.gov/12554545/)]
49. Christensen H, Griffiths K, Groves C, Korten A. Free range users and one hit wonders: community users of an Internet-based cognitive behaviour therapy program. *Aust N Z J Psychiatry* 2006 Jan;40(1):59-62. [doi: [10.1080/j.1440-1614.2006.01743.x](https://doi.org/10.1080/j.1440-1614.2006.01743.x)] [Medline: [16403040](https://pubmed.ncbi.nlm.nih.gov/16403040/)]
50. Gellatly J, Bower P, Hennessy S, Richards D, Gilbody S, Lovell K. What makes self-help interventions effective in the management of depressive symptoms? meta-analysis and meta-regression. *Psychol Med* 2007 Sep;37(9):1217-1228. [doi: [10.1017/S0033291707000062](https://doi.org/10.1017/S0033291707000062)] [Medline: [17306044](https://pubmed.ncbi.nlm.nih.gov/17306044/)]
51. Ebert DD, Buntrock C, Reins JA, Zimmermann J, Cuijpers P. Efficacy and moderators of psychological interventions in treating subclinical symptoms of depression and preventing major depressive disorder onsets: protocol for an individual patient data meta-analysis of randomised controlled trials. *BMJ Open* 2018 Mar 16;8(3):e018582. [doi: [10.1136/bmjopen-2017-018582](https://doi.org/10.1136/bmjopen-2017-018582)] [Medline: [29549201](https://pubmed.ncbi.nlm.nih.gov/29549201/)]
52. Raes F, Williams JMG, Hermans D. Reducing cognitive vulnerability to depression: a preliminary investigation of Memory Specificity Training (MEST) in inpatients with depressive symptomatology. *J Behav Ther Exp Psychiatry* 2009 Mar;40(1):24-38. [doi: [10.1016/j.jbtep.2008.03.001](https://doi.org/10.1016/j.jbtep.2008.03.001)] [Medline: [18407245](https://pubmed.ncbi.nlm.nih.gov/18407245/)]

## Abbreviations

**CATCH-IT:** Competent Adulthood Transition with Cognitive Behavioral Humanistic and Interpersonal Training  
**CES-D:** Center for Epidemiologic Studies - Depression  
**CONSORT:** Consolidated Standards of Reporting Trials  
**DSR:** Depression Symptom Rating  
**GAS:** Global Assessment Scale  
**GLAD-PC:** Guidelines for Adolescent Depression in Primary Care  
**K-SADS:** Kiddie Schedule for Affective Disorders Scale

*Edited by G Eysenbach; submitted 28.10.19; peer-reviewed by B Zima, C Moberg, M Deady; comments to author 03.03.20; revised version received 28.07.20; accepted 16.08.20; published 28.10.20.*

### *Please cite as:*

Van Voorhees B, Gladstone TRG, Sobowale K, Brown CH, Aaby DA, Terrizzi DA, Canel J, Ching E, Berry AD, Cantorna J, Eder M, Beardslee W, Fitzgibbon M, Marko-Holguin M, Schiffer L, Lee M, de Forest SA, Sykes EE, Suor JH, Crawford TJ, Burkhouse KL, Goodwin BC, Bell C

24-Month Outcomes of Primary Care Web-Based Depression Prevention Intervention in Adolescents: Randomized Clinical Trial  
*J Med Internet Res* 2020;22(10):e16802

URL: <https://www.jmir.org/2020/10/e16802>

doi: [10.2196/16802](https://doi.org/10.2196/16802)

PMID: [33112254](https://pubmed.ncbi.nlm.nih.gov/33112254/)

©Benjamin Van Voorhees, Tracy R G Gladstone, Kunmi Sobowale, C Hendricks Brown, David A Aaby, Daniela A Terrizzi, Jason Canel, Eumene Ching, Anita D Berry, James Cantorna, Milton Eder, William Beardslee, Marian Fitzgibbon, Monika Marko-Holguin, Linda Schiffer, Miae Lee, Sarah A de Forest, Emily E Sykes, Jennifer H Suor, Theodore J Crawford, Katie L Burkhouse, Brady C Goodwin, Carl Bell. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Original Paper

# Effectiveness of a Peer-Led Web-Based Intervention to Improve General Self-Efficacy in Using Dating Apps Among Young Adults: Randomized Clustered Trial

William CW Wong<sup>1,2</sup>, MD, MB ChB; Wai Han Sun<sup>3</sup>, MPhil, MSSc, BSocSci; Shu Ming Cheryl Chia<sup>2</sup>, MSc; Joseph D Tucker<sup>4,5</sup>, MD, DPhil; William PH Mak<sup>6</sup>, MSc; Lin Song<sup>7</sup>, DPhil; Kitty Wai Ying Choi<sup>3</sup>, BSocSci; Stephanie Tsz Hei Lau<sup>2</sup>, MPH; Eric Yuk Fai Wan<sup>2,8</sup>, DPhil

<sup>1</sup>Department of General Practice, University of Hong Kong-Shenzhen Hospital, Shenzhen, China

<sup>2</sup>Department of Family Medicine and Primary Care, University of Hong Kong, Hong Kong, China

<sup>3</sup>Sticky Rice Love Ltd., Hong Kong, China

<sup>4</sup>University of North Carolina at Chapel Hill - Project China, Guangzhou, China

<sup>5</sup>London School of Hygiene and Tropical Medicine, London, United Kingdom

<sup>6</sup>Vocational Training Council, Hong Kong, China

<sup>7</sup>Department of Communication, Faculty of Social Sciences, University of Macau, Macau, Macao

<sup>8</sup>Department of Pharmacology and Pharmacy, University of Hong Kong, Hong Kong, China

**Corresponding Author:**

William CW Wong, MD, MB ChB

Department of Family Medicine and Primary Care

University of Hong Kong

3/F, Ap Lei Chau Clinic, 161 Main Street, Ap Lei Chau,

Hong Kong

China

Phone: 852 64883866

Email: [wongwcw@hku.hk](mailto:wongwcw@hku.hk)

## Abstract

**Background:** Online dating apps are popular platforms for seeking romance and sexual relationships among young adults. As mobile apps can easily gain access to a pool of strangers (“new friends”) at any time and place, it leads to heightened sexual health risks and privacy concerns.

**Objective:** This study aimed to evaluate the effectiveness of a peer-led web-based intervention for online dating apps to prepare Chinese college students so that they have better self-efficacy when using dating apps.

**Methods:** An open clustered randomized controlled trial was conducted among students from three colleges (The University of Hong Kong, Hang Seng University of Hong Kong, and Yijin Programme of Vocational Training College) in Hong Kong. Students aged 17 to 27 years who attended common core curriculum or general education were randomized into intervention and control groups. The intervention material, developed with high peer engagement, included four short videos, an interactive scenario game, and a risk assessment tool. An existing website promoting physical activities and healthy living was used as a control. Using the information, motivation, and behavioral skills (IMB) approach to design the evaluation, questionnaires covering participants’ sociodemographics and dating app characteristics, as well as the general self-efficacy scale (GSE) as the primary outcome and the risk propensity scale (RPS) as the secondary outcome were administered before, immediately after, and at 1 month after the intervention. Intention-to-treat analysis was adopted, and between-group differences were assessed using the Mann-Whitney *U* test. A post-hoc multiple linear regression model was used to examine the correlates of the GSE and RPS.

**Results:** A total of 578 eligible participants (290 in the intervention group and 288 in the control group) participated in the study with 36 lost to follow-up. There were more female participants (318/542, 58.7%) than male participants in the sample, reflecting the distribution of college students. Over half of the participants (286/542, 52.8%) reported the following reasons for using dating apps: being curious (170/498, 34.1%), trying to make new friends (158/498, 31.7%), and finding friends with similar interests (121/498, 24.3%). Overall, the participants in the intervention group reported favorable experiences when compared with the finding in the control group. There was significant improvement in the GSE score and reduction in the RPS score ( $P<.001$ ) in

the intervention group. University of Hong Kong students were more susceptible to risk reduction after the intervention when compared with students from the other two institutions.

**Conclusions:** The online intervention was effective in improving general self-efficacy and reducing risk tendency among young students. Future work is needed to determine if this approach is cost-effective and such behavioral change is sustainable.

**Trial Registration:** ClinicalTrials.gov NCT03685643; <https://clinicaltrials.gov/ct2/show/NCT03685643>.

**International Registered Report Identifier (IRRID):** RR2-10.1186/s13063-018-3167-5

(*J Med Internet Res* 2020;22(10):e16378) doi:[10.2196/16378](https://doi.org/10.2196/16378)

## KEYWORDS

internet; sexual health; self-efficacy; young adult; risk assessment

## Introduction

### Background of Online Dating Apps

The use of online dating apps has become popular for seeking friendship and romantic and sexual relationships owing to the advancement of mobile technology and increased internet accessibility through smartphones. Many dating apps are free to download and embed a geolocation function that allows users to find nearby like-minded individuals. Through virtual communication platforms, users can easily meet friends and gain access to a wide pool of potential sexual partners at any place and time.

In the United States, there was an increase in the use of online or mobile dating apps by 4% in only 2 years (from 11% in 2013 to 15% in 2015) [1]. Young adults (aged 18-24 years) and older adults (aged 50-60 years) contributed most to the increase in dating app usage [1]. A more recent survey conducted in 2017 showed that 30% of adults aged 18 to 29 years were using dating apps [2]. Sumter and Vandenbosch found that 17% of young adults had been using popular mobile dating apps, such as Tinder, Grindr, and Coffee meets Bagel in the past 2 years [3]. Online dating apps are also gaining popularity among Chinese youth [4]. It was reported that as many as two-fifth of millennials used online dating apps in Hong Kong [5]. Young adults (including university students) have been identified as the largest population group engaged in online dating.

Prior research found an increased level of risky sexual behavior among individuals who sought multiple partners through online domains [6]. According to Sawyer et al, young adults accounted for an estimated 50% of the 20 million new sexually transmitted infection (STI) cases per year, and less than half of them used condoms [7]. A local study found that dating app users were more likely to have multiple sexual partners and unprotected sex, resulting in an increased risk of STI [6]. Impulsivity has been consistently associated with risky sexual behavior, which contributes to the increasing rates of STI among young adults [8]. The characteristics of impulsive individuals usually involve living in the moment and less likelihood of delaying gratification. Hence, they are more likely to engage in activities that are seen as “fun” and “exciting,” which are the characteristics of meeting someone online. Dating apps with built-in geolocation and instant messaging functions facilitate inner impulsivity among young adults to have social and even sexual interactions. In the case of dating apps, the lack of prescreening and instilling essential knowledge, attitudes, and

behaviors among users could cause emerging concerns beyond public health. In our previous study, we found that dating app users could experience other nonhealth-related concerns, such as financial scams and privacy intrusion [9]. As sex education is not compulsory in Hong Kong schools but is provided on an ad hoc basis or as an extracurricular activity [10], the lack of training and awareness regarding dating app use could further expose these young people to the aforementioned risks.

### Peer-Led Approach and Crowdsourcing in Intervention Development

Nowadays, many young adults gain sexual knowledge from their peers or the internet. Therefore, peer-led education, which is defined as “teaching or sharing of health information, values, and behaviors between individuals with shared characteristics such as social status, experience, and cultural backgrounds” [11], could be an effective strategy for disseminating sex education. It is an integrative approach that involves delivery of an intervention by peers of similar age in settings, such as community centers and youth clubs, using pedagogical or “diffusional” methods. Our systematic review and meta-analysis on peer-led sex education in developed countries demonstrated that peer-led education could effectively change sexual health knowledge and attitudes among young people [12].

In this study, a crowdsourcing method was adopted to form the backbone of this peer-led approach. Crowdsourcing is defined as having a group to solve a problem [13]. It is a tool that uses a bottom-up approach to facilitate and engage the community. Crowdsourcing allows a diverse group of individuals, including laypeople and experts, to contribute to public health interventions [14]. In addition to innovation, crowdsourcing discourages cognitive fixation, a process where people become focused upon others’ ideas in a group setting, but forms a basis to allow development of promotional videos and images in a cooperative manner.

In collaboration with the research team and a nongovernmental organization (NGO) that specializes in online sex education (*StickyRiceLove*), a crowdsourcing contest called “‘Hi, Stranger!’ Dating Apps Education Design Contest” was conducted from December 2017 to April 2018 among students from the University of Hong Kong (HKU) enrolled in a common core course (Sexuality and Culture), as well as the members of a local NGO network called *Sexuality Education Alliance*. They were first asked to generate and submit ideas in the form of texts, images, videos, and websites. They were later evaluated

based on the content, interactivity, creativity, and credibility by a team of experts made up of public health physicians and sociologists using a peer-vetted creative production approach [15]. The following five key risk domains were identified: sexual harassment, privacy, monetary issues, legal issues, and mental well-being. The synthesized results were used to develop the content and modes of delivery of the intervention in a design workshop by the peer volunteers from *StickyRiceLove*. Full details of the intervention development are published elsewhere [9].

### Study Aim and Hypothesis

The aim of this study was to assess the effectiveness of a peer-led web-based intervention among young adults enrolled in three tertiary educational institutions in Hong Kong. The specific objectives were to evaluate the effectiveness and the participants' acceptability of the peer-led online intervention based on the changes in participants' emotions, awareness, attitudes, and behavioral skills in using dating apps. In addition, this study explored participants' sociodemographic factors and dating app characteristics as moderators of the primary outcome of the intervention (ie, general self-efficacy scale [GSE] score) and the secondary outcome (risk propensity scale [RPS] score). It was hypothesized that there would be a statistical difference between the intervention and control groups, in which the participants in the intervention group would have higher self-efficacy and a lower tendency to take risks 1 month after the intervention. In addition, differences in the education environment and mode of study among the tertiary institutions would affect the GSE and RPS scores.

## Methods

### Study Design

An open, clustered, randomized controlled trial (RCT) involving a peer-led safer sex intervention and control (1:1 allocation) was conducted among college students between September 2018 and May 2019. Allocation concealment was achieved through the sealed envelope method [16], and class clusters were randomly assigned to the two groups by an independent researcher. The study protocol and its design were approved by the Institutional Review Board of Hong Kong West Cluster and HKU (UW 18-369).

### Sampling Methods

The inclusion criteria were as follows: (1) age 17 to 27 years and (2) first- or second-year college students enrolled into different tutorial groups or classes in the common core curriculum or general education courses in the three main tertiary educational institutions based in Hong Kong, namely the HKU, Hang Seng University of Hong Kong (HSUHK), and

Yijin Programme of Vocational Training College. Participants were excluded if they were color blind, had no access to a computer/internet literacy, and could not read/write Chinese.

The HKU is the oldest tertiary educational institution in Hong Kong and is regarded as one of the most prestigious universities in Asia, with five out of six applicants admitted being in the top 10% achievement ranking in the Hong Kong Examinations and Assessment Authority public examination. HSUHK is a nonprofit self-financed private university that offers undergraduate and taught postgraduate degrees, with five schools in the areas of business, communication, decision sciences, humanities and social science, and translation. The Yijin Programme is an alternative pathway for college education of generic skills including language, interpersonal skills, and communication offered to individuals who have completed high school or adult learners aged 21 years or above. Its qualification is accepted by government agencies as meeting academic entry requirements for civil service posts.

Based on the primary outcome of self-efficacy, a previous cluster RCT on school-delivered rights-based sexuality education for adolescents revealed a standard deviation of 0.56 and effect size of 0.20 [17]. Considering a minimal cluster size of 10 students, level of significance of .05, power of 0.8, and intracluster correlation coefficient of 0.007, at least 338 students were required for this study, with a 30% assumed attrition rate.

### Intervention

The final intervention involved four (2-4 min) short videos, an interactive scenario game, and a risk assessment tool (Multimedia Appendix 1, Multimedia Appendix 2, Multimedia Appendix 3, and Multimedia Appendix 4). The videos aimed to address the risks and benefits of using dating apps and encouraged the viewers to reflect on their perception of dating apps [9]. The first video illustrated similarities between meeting people on dating apps and in real life, providing examples of being misled by profile characteristics and monetary scams. The second video mainly targeted privacy concerns. The third demonstrated risky sexual behaviors associated with dating apps, and the fourth explained the legal issues and risks of sexual assault, including precautionary steps and available resources.

The scenario game depicted in Figure 1 is a first-person simulation game where the participant is presented with multiple choices when faced with real-life scenarios [9]. The game has been designed with various algorithms that result in both positive and adverse outcomes. A brief rationale for how player choices resulted in the outcomes would be indicated. Lastly, a risk assessment evaluation consisting of 14 questions would appear, and it would give the participant a score to infer the risk level for adverse events when using dating apps.

**Figure 1.** Screenshots of the online game.

### Intervention Procedure

According to the protocol, participants were informed about the voluntary nature of the assessment and that their refusal to participate would not affect their grades [18]. The participating students were asked not to share or discuss the interventions at the time of enrollment. A quasianonymous approach was used (ie, no login when accessing the intervention on the computer/mobile phone). Technical or logistical measures (eg, cookies, email confirmation, and phone calls) were not used to detect or prevent access.

After obtaining written consent, participants completed a preintervention online questionnaire, a postintervention questionnaire immediately after the intervention, and a follow-up questionnaire that encompassed questions relating to risk assessment 1 month later. The questionnaires were constructed and administered using a password-secured survey tool called SurveyGizmo to be accessed by a QR code using participants' mobile devices at their institutions. A reminder was sent to the students who did not complete the follow-up questionnaire after 1 week, and the questionnaire was closed 2 weeks after delivery. Upon completion, participants were given a HK \$50 (US \$1=HK \$7.8) coffee coupon as gratitude for participation.

In contrast, an existing Hong Kong government website promoting physical activity and healthy living was used in the control group. The "health exercise for all" campaign website has a similar modality as that in the intervention group (ie, videos and interactive games) to illustrate the different forms of exercise for different age groups and settings [19]. The game allows players to input personal information, including weight, gender, sports preference, and time spent on an individual sport with a personalized training diary and exercise tips.

### Evaluation

The effectiveness of the intervention was guided by the information (facts, heuristics, and implicit theories), motivation (personal and social), and behavioral skills (self-efficacy and objective skills) (IMB) model [20]. The GSE includes 10 items for assessing one's ability to cope with a range of difficulties in daily life, correlated to one's self-belief, emotion, optimism, and work satisfaction [21]. The RPS rates general risk-taking tendencies with nine items on a 9-point Likert scale. A higher score indicates higher risk-taking tendencies [22]. Patient Health Questionnaire-2 (PHQ-2) is a two-item scale to assess the frequency of depressed mood and anhedonia over the previous 2 weeks, with preliminary screening for depressive symptoms [23]. In our study, the self-reported GSE was considered an appropriate primary outcome measure, as the GSE not only measures one's beliefs about their capabilities to control life events, such as finding a relationship or sexual partner online and achieving desirable outcomes, but is also related to social anxiety such that a negative GSE score indicates vulnerability to higher levels of stress when facing a difficult situation [24].

The preintervention questionnaire consisted of sociodemographic factors and dating app characteristics, such as age, gender, sexual orientation, relationship status, housing type, social media use, reasons to engage in dating apps, addiction risk, and unhappy encounters associated with dating apps (Multimedia Appendix 5). The postintervention questionnaire encompassed questions on acceptability of the intervention (ie, expressed interest, appropriateness of content, sufficient examples and explanation provided, cultivation of the participants' skills, continual use, and likelihood of recommendation to friends) (Multimedia Appendix 6). The follow-up questionnaire included questions matching those in



the preintervention questionnaire, with additional questions regarding visitation frequency and actual recommendation to friends (Multimedia Appendix 7).

## Statistical Methods

The demographics were described, and differences in baseline were assessed using chi-square tests for categorical variables (eg, age, gender, and sexual orientation) and *t* tests for scaled outcomes. In order to conduct a chi-square test for the evaluation of the intervention's acceptability among the participants, five scaled options were categorized into two groups as follows: "strongly agree, agree, and neutral" and "strongly disagree and disagree." Students who did not complete the questionnaire were included for the analysis as per intention to treat (ITT), with "last observation carried forward" values adopted. Owing to the skewness distribution of the outcomes, the differences of intervention testing for mainly the GSE, RPS, and PHQ-2 at follow-up were assessed using the Mann-Whitney *U* test. Univariate linear regression and multiple linear regression

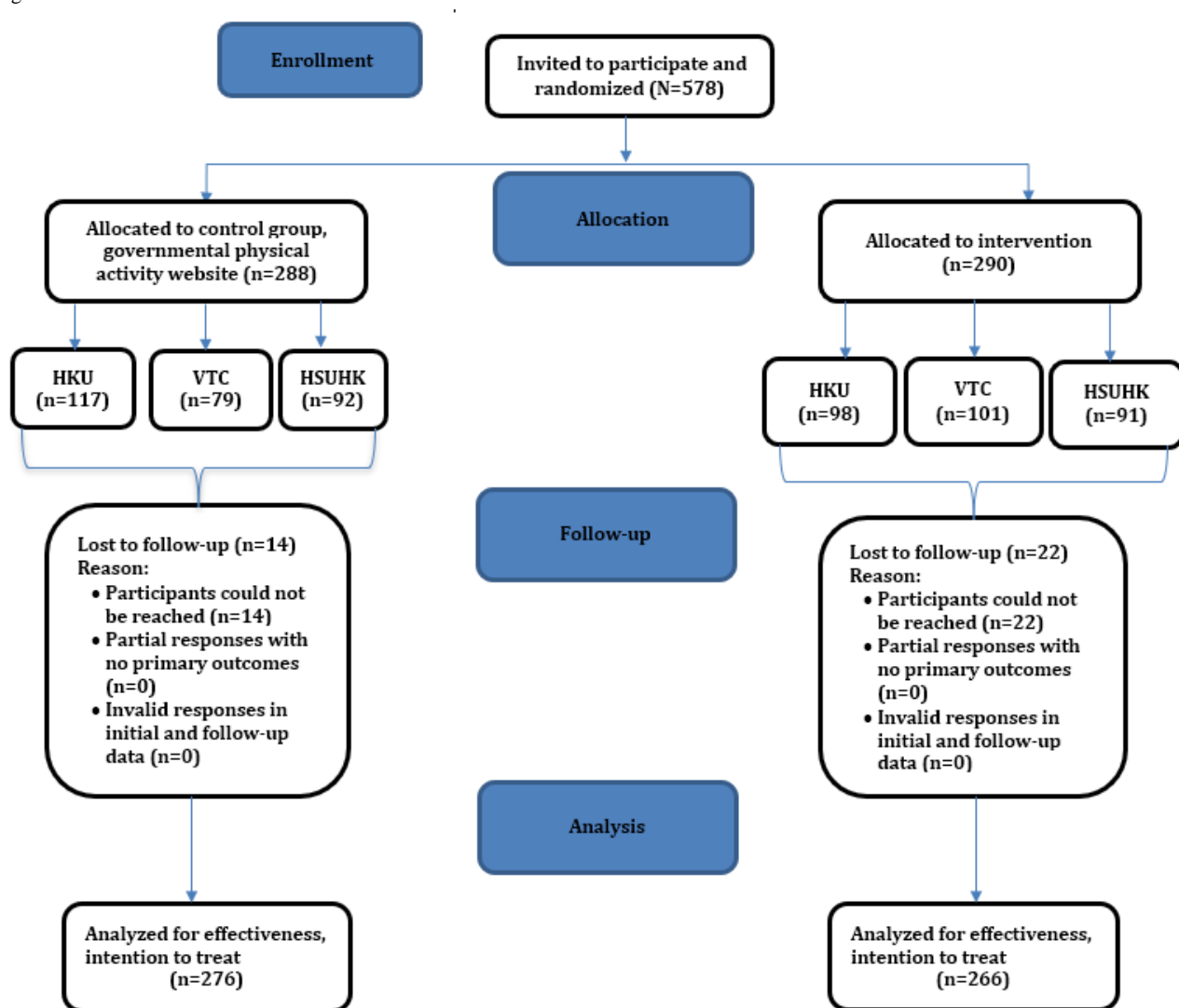
models were used to measure the association among the different confounding factors (eg, gender, relationship status, and reasons for using dating apps) against the primary and secondary outcomes. All data were entered, cleaned, and analyzed using SPSS (version 25; IBM Corp), and statistical significance was set at a *P* value <.05.

## Results

### Demographics and Dating App Statistics

A total of 578 participants who completed the preintervention questionnaire were randomized into the intervention (*n*=290) and control (*n*=288) groups. Since common core curriculum and general education classes were compulsory courses, the response rate was generally very high (215/240, 89.5% for HKU; 183/200, 91.5% for HSUHK; and 180/208 for Yijin Programme, 86.5%). A total of 36 participants were lost to follow-up, giving rise to a dropout rate of 6.2% (Figure 2).

**Figure 2.** CONSORT flow diagram. HKU: University of Hong Kong; HSUHK: Hang Seng University of Hong Kong; VTC: Vocational Training College.



Participants in the intervention and control groups had a similar age profile (ie, mean 19.83 years, SD 1.9 years and mean 19.88 years, SD 1.8 years, respectively) (Table 1). There were more

females (318/542, 58.7%) than males (224/542, 41.3%), reflecting the gender ratio in tertiary education (53.3% females and 46.7% males pursuing undergraduate studies in 2017/2018,



University Grants Committee). Most of the participants were heterosexual (490/542, 90.4%). Two-thirds of the recruited participants were single (353/542, 65.1%), and the remaining (189/542, 34.9%) were in some form of relationship. In addition, about one-third stayed at private accommodation (208/542, 38.4%) or public housing (186/542, 34.3%).

**Table 2** presents the gathered responses pertaining to the participants' characteristics regarding dating apps. Nearly half of the recruited participants (286/578, 49.5%) were found to have visited dating websites. The participants suggested that internet (258/606, 42.6%) and friends (266/606, 43.9%) were the most popular means to gain information about dating apps according to a multiple response setting where the participants could select more than one choice per question. In addition, social media was highlighted by the responses as the most popular social networking platform across both the intervention and control groups (96/218, 44.0% and 101/225, 44.9%, respectively).

Being curious (170/498, 34.1%), trying to make new friends (158/498, 31.7%), and finding friends with similar interests (121/498, 24.3%) were popular selections made by the participants regarding what attracted them to engage in dating apps. Most of the dating app users responded that they did not face any discontent (207/286, 72.4%), except for a minority that faced privacy intrusion (34/286, 11.9%) and sexual assaultment (23/286, 8.0%) before. More than half of the responses (169/285, 59.3%) highlighted that using dating apps to establish friendship required a longer time than anticipated. The participants perceived that dating apps had affected their work or school life (106/285, 37.2%), sleep (125/285, 43.9%), and relationship with their friends/family (75/285, 26.3%). At baseline, no relevant differences were found among the outcomes across the intervention and control groups ( $P=.09$  to  $P=.85$ ).

**Table 1.** Baseline sociodemographic characteristics of the participating college students.

Variable	Intervention group (N=290), n (%) or mean (SD)	Control group (N=288), n (%) or mean (SD)
<b>Gender</b>		
Female	139 (47.9%)	179 (62.2%)
Male	137 (47.2%)	87 (30.2%)
Not specified	14 (4.8%)	22 (7.6%)
Age (years)	19.83 (1.9)	19.88 (1.8)
<b>Type of housing</b>		
Private apartment	107 (36.9%)	101 (35.1%)
Public housing estates	104 (35.9%)	82 (28.5%)
Subsidised housing	34 (11.7%)	39 (13.5%)
Student hostel	19 (6.6%)	29 (10.1%)
Village	7 (2.4%)	15 (5.2%)
Short interim	5 (1.7%)	0 (0%)
Not specified	14 (4.8%)	22 (7.6%)
<b>Relationship status</b>		
Single	183 (63.1%)	170 (59.0%)
Others (dating, married, or cohabitating)	93 (32.1%)	96 (33.3%)
Not specified	14 (4.8%)	22 (7.6%)
<b>Sexual orientation</b>		
Heterosexual	250 (86.8%)	240 (83.3%)
Others (gay, lesbian, or bisexual)	24 (8.3%)	26 (9.0%)
Not specified	14 (4.9%)	22 (7.6%)

**Table 2.** Characteristics of the responses from participants about their dating app use.

Variable	Intervention group (N=290), n (%) or mean (SD)	Control group (N=288), n (%) or mean (SD)	P value <sup>a</sup>
<b>Ever used a dating website</b>			.51
Yes	139 (47.9%)	147 (51.0%)	
No	151 (52.1%)	141 (49.0%)	
<b>Source of information about dating apps</b>			
Internet	118 (34.8%)	140 (52.4%)	<.001
Friends	147 (43.3%)	119 (44.6%)	.82
Teachers or school staff	52 (15.3%)	4 (1.5%)	<.001
Magazine or fashion materials	3 (0.9%)	2 (0.7%)	.75
Family/parents	19 (5.6%)	2 (0.7%)	<.001
<b>Social networking platforms</b>			
Social media platform (eg, Facebook and Instagram)	96 (44.0%)	101 (44.7%)	.49
Dating apps	42 (19.3%)	77 (62.1%)	<.001
Instant chatline (eg, WhatsApp and WeChat)	78 (35.8%)	43 (34.7%)	<.001
Others (eg, school and online game)	2 (0.9%)	4 (3.2%)	.03
<b>Reasons for using dating apps</b>			
Making new friends	73 (29.9%)	85 (33.5%)	.38
Out of curiosity	83 (34.0%)	87 (35.6%)	.77
Finding friends with similar interest	61 (25.0%)	60 (24.5%)	.54
Sexual relationship	11 (4.5%)	9 (3.7%)	.59
Long-term relationship	16 (6.6%)	13 (5.3%)	.39
<b>Encounter of unhappiness while using dating apps</b>			.77
Yes	38 (27.3%)	41 (27.9%)	
No	101 (72.6%)	106 (72.1%)	
<b>Types of problem encountered with dating apps</b>			
Financially scammed	7 (18.4%)	5 (12.5%)	.65
Cyberbully/blackmail	5 (12.3%)	4 (10.0%)	.74
Privacy issues	15 (39.5%)	19 (47.5%)	.86
Sexual assault	11 (28.9%)	12 (30.0%)	.73
<b>Dating apps and their addiction implications</b>			
<b>Time spent on dating app is longer than anticipated</b>			.20
Disagree	61 (44.2%)	55 (37.4%)	
Agree	77 (55.8%)	92 (62.6%)	
<b>Affect work or school life</b>			.30
Disagree	83 (60.1%)	96 (65.3%)	
Agree	55 (39.9%)	51 (34.7%)	
<b>Affect sleep</b>			.83
Disagree	77 (55.8%)	83 (56.5%)	
Agree	61 (44.2%)	64 (43.5%)	
<b>Affect relationship with friends/family</b>			.67
Disagree	104 (25.2%)	107 (72.8%)	
Agree	35 (74.8%)	40 (27.2%)	

Variable	Intervention group (N=290), n (%) or mean (SD)	Control group (N=288), n (%) or mean (SD)	<i>P</i> value <sup>a</sup>
<b>Outcomes</b>			
GSE <sup>b</sup>	23.95 (5.25)	22.65 (4.85)	.27
RPS <sup>c</sup>	5.96 (0.95)	5.97 (0.90)	.85
PHQ-2 <sup>d</sup>	3.86 (1.33)	4.10 (1.38)	.09

<sup>a</sup>Chi-square test is used to compare the statistical difference in categorical variables between groups. Mann-Whitney *U* test is also used.

<sup>b</sup>GSE: general self-efficacy scale.

<sup>c</sup>RPS: risk propensity scale.

<sup>d</sup>PHQ-2: Patient Health Questionnaire.

### Intervention Evaluation Outcomes

As presented in Table 3, the responses received indicated that the proportion of participants who became interested in the dating program was significantly higher in the intervention group (162/410, 39.5%) than in the control group (88/410, 21.5%;  $P=.002$ ). In addition, the proportion of participants who responded that the intervention provided sufficient explanations and examples was significantly higher in the intervention group (231/410, 56.3%) than in the control group (129/410, 31.5%;  $P<.001$ ). Moreover, the proportion of participants who responded that they would recommend the approach to their friends was significantly higher in the intervention group (211/410, 51.5%) than in the control group (125/410, 30.5%;  $P=.04$ ). Furthermore, one-quarter of participants from the intervention group (114/497, 23.0%) and one-tenth of participants from the control group (47/497, 9.5%) had introduced the intervention to their friends ( $P<.001$ ).

After adjustment for the baseline GSE score, a significant difference was found between the intervention and control

groups at the 1-month follow-up ( $P<.001$ ). A similar result trend is elicited especially in HSUHK, where a significant difference was observed ( $P<.001$ ). In addition, a small but significant ( $P=.04$ ) decrease in the GSE score by  $-0.03$  for every 1 unit increase could be observed in the control group (Table 4).

Using the stepwise univariate regression analysis, dating apps and their addiction implications were found to have a positive association for affecting the RPS score ( $P<.001$ ) (Table 5). On the other hand, other factors having positive associations for affecting the GSE score were relationship status ( $P=.02$ ), reasons to use dating apps, in particular, seeking sexual relationships ( $P=.002$ ) and long term relationships ( $P=.014$ ), information about dating apps disseminated by teachers/school staff ( $P=.02$ ), and differences in gender ( $P=.04$ ). These significant factors were further analyzed using a backward multiple regression analysis, and it was found that there was a significant difference in the treatment group for the GSE score ( $P<.001$ ) and RPS score ( $P<.001$ ) (Table 6). In addition, HKU was found to be a significant confounding factor ( $P=.006$ ) negatively associated with the RPS score.

**Table 3.** Feedback responses at postintervention.

Variable	Intervention group (N=276), n (%)	Control group (N=266), n (%)	P value <sup>a</sup>
<b>This intervention caused me to be interested in the dating program.</b>			
Disagree	86 (34.6%)	74 (45.7%)	.002
Agree	162 (65.3%)	88 (54.3%)	
<b>The content of the intervention is appropriate.</b>			
Disagree	17 (6.9%)	18 (11.1%)	.14
Agree	231 (93.1%)	144 (88.9%)	
<b>This intervention provided sufficient explanations and examples.</b>			
Disagree	17 (6.9%)	33 (20.4%)	<.001
Agree	231 (93.1%)	129 (79.6%)	
<b>This intervention has cultivated my ability and skills.</b>			
Disagree	44 (17.7%)	33 (20.4%)	.52
Agree	204 (82.3%)	129 (79.6%)	
<b>I will continue to use the intervention.</b>			
Disagree	39 (9.5%)	37 (9.0%)	.07
Agree	209 (50.1%)	125 (30.5%)	
<b>I will recommend this intervention to my friends.</b>			
Disagree	37 (15.7%)	37 (22.8%)	.04
Agree	211 (84.3%)	125 (77.2%)	
<b>Are there any beneficial effects with respect to the program?</b>			
Yes	209 (90.1%)	188 (81.7%)	.16
No	23 (9.9%)	42 (18.3%)	
<b>Are there any improvements to be made to the program?</b>			
Yes	123 (60.6%)	141 (61.5%)	.72
No	80 (39.4%)	88 (38.4%)	
<b>Have you visited the HKU<sup>b</sup> intervention website after the study?</b>			
Yes	153 (56.0%)	130 (57.0%)	.09
No	120 (43.9%)	98 (43.0%)	
<b>Have you recommended any friends after the intervention?</b>			
Yes	114 (42.7%)	47 (20.4%)	<.001
No	153 (57.3%)	183 (79.6%)	

<sup>a</sup>Chi-square test is used to compare the statistical difference between groups.

<sup>b</sup>HKU: University of Hong Kong.

**Table 4.** Effectiveness of the intervention in participating college students at the 1-month follow-up.

Scale	Intervention (N=290), mean (SD)			Control (N=288), mean (SD)			Control vs intervention <i>P</i> value			Total	Univariate regression analysis <i>P</i> value (95% CI)
	HKU <sup>a</sup> (n=98)	HSUHK <sup>b</sup> (n=91)	VTC <sup>c</sup> (n=101)	HKU (n=117)	HSUHK (n=92)	VTC (n=79)	HKU	HSUHK	VTC		
GSE <sup>d</sup>	24.55 (5.57)	24.68 (4.85)	23.59 (5.88)	23.01 (4.71)	22.32 (5.14)	23.37 (5.87)	.12 <sup>e</sup>	<.001 <sup>e</sup>	.53 <sup>e</sup>	<.001 <sup>f</sup>	.04 (-0.03 to -0.01)
RPS <sup>g</sup>	5.57 (0.99)	5.67 (1.01)	4.72 (1.55)	5.73 (1.09)	5.65 (1.13)	5.31 (1.23)	.10 <sup>e</sup>	.97 <sup>e</sup>	.32 <sup>e</sup>	.12 <sup>f</sup>	.06 (0.02 to 0.09)
PHQ-2 <sup>h</sup>	4.24 (1.41)	3.78 (1.34)	4.55 (1.52)	4.09 (1.31)	3.77 (1.15)	4.30 (1.51)	.47 <sup>e</sup>	.80 <sup>e</sup>	.26 <sup>e</sup>	.19 <sup>f</sup>	.23 (-0.05 to 0.01)

<sup>a</sup>HKU: University of Hong Kong.<sup>b</sup>HSUHK: Hang Seng University of Hong Kong.<sup>c</sup>VTC: Vocational Training College.<sup>d</sup>GSE: general self-efficacy scale.<sup>e</sup>Mann-Whitney *U* test is used to compare the statistical difference between groups within institutions.<sup>f</sup>Mann-Whitney *U* test is used to compare the statistical difference between groups overall.<sup>g</sup>RPS: risk propensity scale.<sup>h</sup>PHQ-2: Patient Health Questionnaire-2.



**Table 5.** Factors affecting the general self-efficacy scale and risk propensity scale scores using a univariate regression model.

Factor	General self-efficacy scale score		Risk propensity scale score	
	Coefficient $\beta$ (95% CI)	<i>P</i> value <sup>a</sup>	Coefficient $\beta$ (95% CI)	<i>P</i> value <sup>a</sup>
<b>Sociodemographic factors</b>				
Age	0.00 (–0.23 to 0.23)	.98	–0.02 (–0.08 to 0.03)	.34
<b>Gender</b> (ref <sup>b</sup> : male)				
Female	–0.93 (–1.79 to –0.06)	.04	0.11 (–0.09 to 0.32)	.28
<b>Housing type</b> (ref: public housing)				
Subsidized housing	0.13 (–1.11 to 1.36)	.84	0.19 (–0.11 to 0.48)	.22
Private housing	–0.51 (–1.38 to 0.36)	.25	–0.20 (–0.41 to 0.01)	.06
Short interim	1.40 (–3.88 to 6.66)	.60	–0.10 (–1.49 to 1.29)	.87
Student hostel	–0.35 (–1.88 to 1.18)	.66	0.21 (–0.11 to 0.56)	.19
Village	2.08 (–0.06 to 4.21)	.06	–0.25 (–0.75 to 0.26)	.34
<b>Sexual orientation</b> (ref: heterosexual)				
Gay/lesbian/bisexual	–0.09 (–1.56 to 1.38)	.90	0.02 (–0.33 to 0.37)	.15
<b>Relationship status</b> (ref: single)				
Dating/marriage/cohabitating	0.10 (0.12 to 1.88)	.03	–0.15 (–0.36 to 0.06)	.37
<b>Dating characteristics</b>				
<b>Ever used dating apps</b> (ref: yes)				
No	–0.69 (–1.53 to 0.14)	.10	0.13 (–0.06 to 0.33)	.18
<b>Types of platforms to know friends</b> (ref: dating apps)				
Social media platforms and instant chatline	–1.37 (–2.91 to 0.18)	.08	–0.01 (–0.40 to 0.38)	.96
<b>Reasons for using dating apps</b> (ref: curiosity)				
New friends	–0.62 (–1.87 to 0.63)	.33	–0.11 (–0.41 to 0.19)	.47
Similar interest	–0.32 (–1.53 to 0.89)	.60	–0.23 (–0.07 to 0.53)	.14
Sexual relationship	3.58 (1.36 to 5.80)	.002	0.23 (–0.36 to 0.81)	.45
Long-term relationship	2.38 (0.46 to 4.27)	.01	–0.13 (–0.62 to 0.37)	.61
<b>Types of media</b> (ref: friends)				
Internet	–0.01 (–0.90 to 0.87)	.98	0.05 (–0.16 to 0.26)	.63
Teachers or school staff members	–1.72 (–3.17 to –0.27)	.02	–0.06 (–0.41 to 0.28)	.72
Magazine or fashion materials	0.11 (–4.28 to 4.50)	.96	0.66 (–0.48 to 1.79)	.26
Family/parents/government initiation	0.18 (–2.04 to 2.39)	.88	–0.01 (–0.63 to 0.43)	.71
<b>Dating apps and their online implications</b>				
Affect sleep	–0.12 (–0.62 to 0.39)	.65	0.26 (0.12 to 0.40)	<.001
Spent a longer time than usual on a dating app	–0.25 (–0.76 to 0.26)	.84	0.35 (0.16 to 0.54)	<.001
Affect work or school	–0.31 (–0.87 to 0.25)	.28	0.31 (0.13 to 0.49)	<.001
Affect relationship with friends and family	–0.61 (–1.21 to 0.00)	.05	0.37 (0.15 to 0.56)	<.001

<sup>a</sup>*P* value is analyzed using univariate analysis.<sup>b</sup>ref: reference.

**Table 6.** Factors affecting the general self-efficacy scale and risk propensity scale scores using a multiple linear regression model.

	General self-efficacy scale score		Risk propensity scale score	
	Coefficient $\beta$ (95% CI)	<i>P</i> value <sup>a</sup>	Coefficient $\beta$ (95% CI)	<i>P</i> value <sup>a</sup>
Intervention group	–2.79 (–4.65 to –0.92)	.004	0.453 (0.099 to 0.81)	.01
<b>Gender</b> (ref <sup>b</sup> : male)				
Female	–1.56 (–3.42 to 0.30)	.98	N/A <sup>c</sup>	N/A
<b>School</b> (ref: VTC <sup>d</sup> )				
HSUHK <sup>e</sup>	–0.03 (–4.86 to 4.81)	.99	–0.37 (–0.79 to 0.06)	.09
HKU <sup>f</sup>	0.35 (–2.93 to 3.63)	.83	–0.62 (–1.06 to –0.18)	.006
<b>Baseline outcomes</b>				
General self-efficacy scale	–0.51 (–0.68 to –0.34)	<.001	N/A	N/A
Risk propensity scale	N/A	N/A	–0.39 (–0.55 to –0.22)	<.001

<sup>a</sup>*P* value is analyzed using multiple linear regression analysis.

<sup>b</sup>ref: reference.

<sup>c</sup>N/A: not applicable.

<sup>d</sup>VTC: Vocational Training College.

<sup>e</sup>HSUHK: Hang Seng University of Hong Kong.

<sup>f</sup>HKU: University of Hong Kong.

## Discussion

### Principal Findings

To the best of our knowledge, there has been limited research focusing on the awareness of online dating app risks among young adults. Most sex-related apps and dating apps contain no information about sexual health promotion [25]. This finding is consistent with the finding of another local study that the use of dating websites was common among Hong Kong college students, mostly for curiosity and seeking new friends [26]. The participants from the intervention group disclosed having a greater interest toward the peer-led intervention compared with participants from the control group. They felt that clear and sufficient examples were given as illustrations compared with participants from the control group. For outcome evaluation, the findings supported that the intervention could elicit significant changes in participants' GSE skills ( $P=.04$ ), as well as significant changes in risk-taking tendencies ( $P=.06$ ).

In our study, participants' GSE scores increased after the intervention, although the absolute difference was small. The small increase in the GSE score was important as an indication that the web-based intervention changed their understanding of perceived control over the usage of dating apps in a short duration of time. A positive GSE finding may also indicate that participants were less anxious about being “judged” by potential partners. They may feel less anxious owing to their ability to hold positive appraisals [24]. Young adults with higher GSE scores would feel more efficacious and less anxious when using dating apps. The promising result of our study regarding enhanced general self-efficacy toward online dating app use is consistent with the findings of a systematic review stipulating that high levels of peer participation in these types of interventions could greatly improve sex knowledge, attitude,

general self-efficacy, and social norms [12]. The positive outcome exhibited could be related to the introduction of gamification in nongaming situations that made the intervention unique and appealing to the targeted youth. The use of real-life scenarios and humor in games and videos may have attracted and retained the participants' attention [27]. As a result, this motivated and improved users' experience and engagement [28]. This was in contrast to a slight deterioration in the GSE score by –0.02 for every 1 unit increase in the control group, suggesting that the impertinent content relating to physical activity was unable to provide knowledge and skills in relation to safer dating app use.

A significant negative change was found in the risk propensity tendency among the participants overall, especially among HKU participants ( $P=.006$ ) (Table 6), suggesting that the web-based intervention may influence young adults who already have strong interest in sexuality and sexual health or that more academically capable students may be more amicable to the intervention. Coupled with the strong emphasis of the potential adverse sexual health outcomes and the precautionary steps from the web-based intervention [25], the intervention may help students to be more cautious toward online dating apps. Such results were not observed among students from other colleges for whom the motivation for unsafe behavior could be attributed more strongly to benefits than to risk perception. On the other hand, it is well known that risk perception itself is insufficient to explain the risky behavior [29].

According to analysis of the online search threads, it was found that young adults may have a higher tendency to engage with the digital world to search for sexual health concerns; topics including sexual pleasure, puberty, menstruation, and transmission; and mechanisms of infection [27]. Therefore, with the successful implementation of our intervention, we hope that

people of all ages, not just young adults, could benefit from this new innovative way of training on how to use dating apps safely.

### Strengths and Limitations

The major strengths of the study include its randomized design, low attrition rate, and implementation of validated outcome measures to assess the effectiveness of our peer-led web-based intervention. The game stimulation allows participants to create a personal profile that can drive tailoring functions and personalized messages. With the incorporation of risks and benefits in this peer-led web-based intervention, it could serve as a practical approach to shape and instill online dating app users with right values.

In contrast, the participants recruited in the selected universities may not be representative of the whole youth population. Further, peer-led education is effective in establishing safer sex norms and attitudes than adult-led interventions, but less so in imparting factual knowledge [30]. Although self-reported measures are common and practical methods to obtain information in behavioral health studies, self-reporting bias may be introduced, especially under-reporting the sensitive nature of sexual behavior [31]. Multiple entries were still possible despite the prohibition of duplicate responses from identical IP addresses. Lastly, the limited timeframe and exposure of the intervention may make it unable to capture any long-term changes in the intensity, frequency, or context of participants' emotions. Therefore, a longer intervention period is required for significant and sustainable changes in outcomes at the risk of high attrition rates. In theory, contamination could occur as participants from the control and intervention groups attended the same class, which could result in undermining of the effect magnitude of the intervention.

### Implications and Recommendations

With the increasing popularity of mobile technology and the internet, this multifaceted intervention could serve as a starting

point for novel interventions to increase safety awareness when seeking sexual partners online. Continuing to document dating app users' experiences through testimonials or videos demonstrated on the intervention website may increase general self-efficacy and reduce the risk tendency toward dating app use. The results obtained from the study could be insightful for public health practitioners and app companies to improve the safety profile of dating apps. Furthermore, novel delivery methods, such as social media and community-based centers, could be considered to reach out to an extensive youth population beyond the college setting. Health professionals could also share and update any new research findings pertaining to the detrimental health impacts associated with casual sex to deter users from risky sexual behavior through dating apps and to create a long-lasting unfavorable outlook about it. Lastly, trial-based health economic evaluation was not conducted in this study, and a previous study showed that a peer-led approach could be potentially more expensive than an adult-led initiative (i.e. €28.2 [US \$33.8] per target student in the peer-led group vs €11.6 [US \$13.9] in the teacher-led arm) [32]. Therefore, it is important to further investigate the cost-effectiveness, barriers, opportunities, and supporting measures to sustain the intervention in the long term by placing more emphasis on feedback from all stakeholders (eg, students, teachers, and members of the school) to allow the development of programs that are culturally and socially appropriate.

### Conclusion

This study demonstrated that a peer-led intervention could improve short-term self-efficacy and reduce risk propensity in using dating apps among young Chinese adults. It is important to evaluate the motivations for using dating apps and attempt to understand the underlying mechanism between using dating apps and associated risks. Future work is needed to determine how to maintain behavioral change over a longer duration, how to reach underserved populations using different means, and how to disseminate such an intervention on a large scale.

### Acknowledgments

The authors wish to thank the study participants and staff from The University of Hong Kong and StickyRiceLove organization, who made this work possible. We are thankful to the students of Hang Seng University of Hong Kong and Yijin Programme of Vocational Training College who participated in this study. We also want to thank Miss Chow Long Ching Bellavista for being part of the team that assisted in the delivery of the intervention. This study was funded by The Healthcare and Promotion Fund Scheme of the Government of Hong Kong (grant number: 09160275). The funding body was not involved in the design of the study or data collection. It was also not involved in the analysis, interpretation of data, or writing of the manuscript.

### Authors' Contributions

WW designed the study, applied for funding, oversaw the whole implementation, supervised the data analysis, edited the manuscript, and provided feedback on the manuscript. LS, CKWY, and LSTH developed and carried out the intervention and evaluation of the tools. SWH and MWP recruited the participants and followed up with them. SWH applied for funding. CSMC conducted the statistical analysis under WYF's guidance. CSMC drafted and edited the manuscript. TJD provided advice on the crowdsourcing materials. All authors read and approved the manuscript.

### Conflicts of Interest

None declared.

## Multimedia Appendix 1

Video illustrating similarities between meeting people on dating apps and in real life. Examples of being misled by profile characteristics and monetary scams.

[[MOV File , 49616 KB - jmir\\_v22i10e16378\\_app1.mov](#) ]

## Multimedia Appendix 2

Video depicting privacy concerns related to dating apps.

[[MOV File , 37196 KB - jmir\\_v22i10e16378\\_app2.mov](#) ]

## Multimedia Appendix 3

Video depicting risky sexual behavior.

[[MOV File , 45974 KB - jmir\\_v22i10e16378\\_app3.mov](#) ]

## Multimedia Appendix 4

Video depicting legal issues and risks of sexual assault, including precautionary steps and available resources.

[[MOV File , 42172 KB - jmir\\_v22i10e16378\\_app4.mov](#) ]

## Multimedia Appendix 5

Preintervention questionnaire.

[[DOCX File , 30 KB - jmir\\_v22i10e16378\\_app5.docx](#) ]

## Multimedia Appendix 6

Postintervention questionnaire.

[[DOCX File , 19 KB - jmir\\_v22i10e16378\\_app6.docx](#) ]

## Multimedia Appendix 7

Follow-up intervention questionnaire.

[[DOCX File , 27 KB - jmir\\_v22i10e16378\\_app7.docx](#) ]

## References

- Tran A, Suharlim C, Mattie H, Davison K, Agénor M, Austin SB. Dating app use and unhealthy weight control behaviors among a sample of U.S. adults: a cross-sectional study. *J Eat Disord* 2019 May 31;7(1):16 [[FREE Full text](#)] [doi: [10.1186/s40337-019-0244-4](#)] [Medline: [31164984](#)]
- Yi S, Te V, Pengpid S, Peltzer K. Social and behavioural factors associated with risky sexual behaviours among university students in nine ASEAN countries: a multi-country cross-sectional study. *SAHARA J* 2018 Dec;15(1):71-79 [[FREE Full text](#)] [doi: [10.1080/17290376.2018.1503967](#)] [Medline: [30058474](#)]
- Sumter S, Vandenbosch L. Dating gone mobile: Demographic and personality-based correlates of using smartphone-based dating applications among emerging adults. 2018 Oct 20;21(3):655-673 [[FREE Full text](#)] [doi: [10.1177/1461444818804773](#)]
- Four in ten Hong Kongers have used internet dating Internet. YouGov. 2017. URL: <https://hk.yougov.com/en-hk/news/2017/11/23/internet-dating/> [accessed 2019-06-10]
- Sexual Behaviour. The Government of the Hong Kong Special Administrative Region (Centre for Health Protection). 2017. URL: <https://www.chp.gov.hk/en/statistics/data/10/757/5520.html> [accessed 2019-06-01]
- Choi EP, Wong JY, Lo HH, Wong W, Chio JH, Fong DY. The Impacts of Using Smartphone Dating Applications on Sexual Risk Behaviours in College Students in Hong Kong. *PLoS One* 2016 Nov 9;11(11):e0165394 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0165394](#)] [Medline: [27828997](#)]
- Sawyer A, Smith E, Benotsch E. Dating Application Use and Sexual Risk Behavior Among Young Adults. *Sex Res Soc Policy* 2017 Sep 6;15(2):183-191 [[FREE Full text](#)] [doi: [10.1007/s13178-017-0297-6](#)]
- Hayaki J, Anderson B, Stein M. Sexual risk-taking mediates the association between impulsivity and acquisition of sexually transmitted infections among hazardously drinking incarcerated women. *Am J Addict* 2012 Nov;21 Suppl 1(1):S63-S71 [[FREE Full text](#)] [doi: [10.1111/j.1521-0391.2012.00296.x](#)] [Medline: [23786513](#)]
- Wong W, Song L, See C, Lau S, Sun W, Choi W, et al. Using Crowdsourcing to Develop a Peer-Led Intervention for Safer Dating App Use: Pilot Study. *JMIR Form Res* 2020 Apr 21;4(4):e12098 [[FREE Full text](#)] [doi: [10.2196/12098](#)] [Medline: [32314975](#)]
- Che FS. A study of the implementation of sex education in Hong Kong secondary schools. *Sex Education* 2005 Aug;5(3):281-294 [[FREE Full text](#)] [doi: [10.1080/14681810500171458](#)]
- MacDowall W, Bonnell C, Davies M. Health Promotion Practice. London, UK: Open University Press; 2006.

12. Sun WH, Miu HY, Wong CK, Tucker JD, Wong WC. Assessing Participation and Effectiveness of the Peer-Led Approach in Youth Sexual Health Education: Systematic Review and Meta-Analysis in More Developed Countries. *J Sex Res* 2018 Jan 29;55(1):31-44. [doi: [10.1080/00224499.2016.1247779](https://doi.org/10.1080/00224499.2016.1247779)] [Medline: [27898248](https://pubmed.ncbi.nlm.nih.gov/27898248/)]
13. Tucker J, Day S, Tang W, Bayus B. Crowdsourcing in medical research: concepts and applications. *PeerJ* 2019;7:e6762 [FREE Full text] [doi: [10.7717/peerj.6762](https://doi.org/10.7717/peerj.6762)] [Medline: [30997295](https://pubmed.ncbi.nlm.nih.gov/30997295/)]
14. Tang W, Wei C, Cao B, Wu D, Li KT, Lu H, et al. Crowdsourcing to expand HIV testing among men who have sex with men in China: A closed cohort stepped wedge cluster randomized controlled trial. *PLoS Med* 2018 Aug;15(8):e1002645 [FREE Full text] [doi: [10.1371/journal.pmed.1002645](https://doi.org/10.1371/journal.pmed.1002645)] [Medline: [30153265](https://pubmed.ncbi.nlm.nih.gov/30153265/)]
15. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. *Am J Prev Med* 2014 Feb;46(2):179-187. [doi: [10.1016/j.amepre.2013.10.016](https://doi.org/10.1016/j.amepre.2013.10.016)] [Medline: [24439353](https://pubmed.ncbi.nlm.nih.gov/24439353/)]
16. Rohrbach LA, Berglas NF, Jerman P, Angulo-Olaiz F, Chou C, Constantine NA. A Rights-Based Sexuality Education Curriculum for Adolescents: 1-Year Outcomes From a Cluster-Randomized Trial. *J Adolesc Health* 2015 Oct;57(4):399-406. [doi: [10.1016/j.jadohealth.2015.07.004](https://doi.org/10.1016/j.jadohealth.2015.07.004)] [Medline: [26403840](https://pubmed.ncbi.nlm.nih.gov/26403840/)]
17. Dettori J. The random allocation process: two things you need to know. *Evid Based Spine Care J* 2010 Dec;1(3):7-9 [FREE Full text] [doi: [10.1055/s-0030-1267062](https://doi.org/10.1055/s-0030-1267062)] [Medline: [22956922](https://pubmed.ncbi.nlm.nih.gov/22956922/)]
18. Lau ST, Choi KW, Chen J, Mak WP, Yeung HK, Tucker J, et al. Study protocol for a peer-led web-based intervention to promote safe usage of dating applications among young adults: a cluster randomized controlled trial. *Trials* 2019 Feb 06;20(1):102 [FREE Full text] [doi: [10.1186/s13063-018-3167-5](https://doi.org/10.1186/s13063-018-3167-5)] [Medline: [30728064](https://pubmed.ncbi.nlm.nih.gov/30728064/)]
19. Healthy exercise for all campaign. Leisure and Cultural Services Department. 2014. URL: <https://www.lcsd.gov.hk/en/healthy/> [accessed 2019-06-01]
20. Fisher WA, Fisher JD, Rye BJ. Understanding and promoting AIDS-preventive behavior: Insights from the theory of reasoned action. *Health Psychology* 1995;14(3):255-264. [doi: [10.1037/0278-6133.14.3.255](https://doi.org/10.1037/0278-6133.14.3.255)]
21. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. In: Weinman J, Wright S, Johnston M, editors. *Measures in health psychology: A user's portfolio*. Windsor, UK: NFER-Nelson; 1995:35-37.
22. Meertens R, Lion R. Measuring an Individual's Tendency to Take Risks: The Risk Propensity Scale. *J Appl Social Psychol* 2008 Jun;38(6):1506-1520 [FREE Full text] [doi: [10.1111/j.1559-1816.2008.00357.x](https://doi.org/10.1111/j.1559-1816.2008.00357.x)]
23. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003 Nov;41(11):1284-1292. [doi: [10.1097/01.MLR.0000093487.78664.3C](https://doi.org/10.1097/01.MLR.0000093487.78664.3C)] [Medline: [14583691](https://pubmed.ncbi.nlm.nih.gov/14583691/)]
24. Pitcho-Prelentzos S, Heckel C, Ring L. Predictors of social anxiety among online dating users. *Computers in Human Behavior* 2020 Sep;110:106381 [FREE Full text] [doi: [10.1016/j.chb.2020.106381](https://doi.org/10.1016/j.chb.2020.106381)]
25. Huang ET, Williams H, Hocking JS, Lim MS. Safe Sex Messages Within Dating and Entertainment Smartphone Apps: A Review. *JMIR Mhealth Uhealth* 2016 Nov 08;4(4):e124 [FREE Full text] [doi: [10.2196/mhealth.5760](https://doi.org/10.2196/mhealth.5760)] [Medline: [27826133](https://pubmed.ncbi.nlm.nih.gov/27826133/)]
26. Chan LS. Who uses dating apps? Exploring the relationships among trust, sensation-seeking, smartphone use, and the intent to use dating apps based on the Integrative Model. *Computers in Human Behavior* 2017 Jul;72:246-258. [doi: [10.1016/j.chb.2017.02.053](https://doi.org/10.1016/j.chb.2017.02.053)]
27. Bailey J, Mann S, Wayal S, Hunter R, Free C, Abraham C, et al. Sexual health promotion for young people delivered via digital media: a scoping review. *Public Health Res* 2015 Nov;3(13):1-120. [doi: [10.3310/phr03130](https://doi.org/10.3310/phr03130)] [Medline: [26583166](https://pubmed.ncbi.nlm.nih.gov/26583166/)]
28. Wakefield M, Loken B, Hornik R. Use of mass media campaigns to change health behaviour. *The Lancet* 2010 Oct;376(9748):1261-1271 [FREE Full text] [doi: [10.1016/s0140-6736\(10\)60809-4](https://doi.org/10.1016/s0140-6736(10)60809-4)]
29. Hanoch Y, Johnson JG, Wilke A. Domain specificity in experimental measures and participant recruitment: an application to risk-taking behavior. *Psychol Sci* 2006 Apr 06;17(4):300-304. [doi: [10.1111/j.1467-9280.2006.01702.x](https://doi.org/10.1111/j.1467-9280.2006.01702.x)] [Medline: [16623686](https://pubmed.ncbi.nlm.nih.gov/16623686/)]
30. Mellanby A, Newcombe R, Rees J, Tripp J. A comparative study of peer-led and adult-led school sex education. *Health Educ Res* 2001 Aug;16(4):481-492 [FREE Full text] [doi: [10.1093/her/16.4.481](https://doi.org/10.1093/her/16.4.481)] [Medline: [11525394](https://pubmed.ncbi.nlm.nih.gov/11525394/)]
31. Schroder KE, Carey MP, Vanable PA. Methodological challenges in research on sexual risk behavior: II. Accuracy of self-reports. *Ann Behav Med* 2003 Oct;26(2):104-123 [FREE Full text] [doi: [10.1207/S15324796ABM2602\\_03](https://doi.org/10.1207/S15324796ABM2602_03)] [Medline: [14534028](https://pubmed.ncbi.nlm.nih.gov/14534028/)]
32. Borgia P, Marinacci C, Schifano P, Perucci CA. Is peer education the best approach for HIV prevention in schools? Findings from a randomized controlled trial. *J Adolesc Health* 2005 Jun;36(6):508-516. [doi: [10.1016/j.jadohealth.2004.03.005](https://doi.org/10.1016/j.jadohealth.2004.03.005)] [Medline: [15901516](https://pubmed.ncbi.nlm.nih.gov/15901516/)]

## Abbreviations

**GSE:** general self-efficacy scale  
**HKU:** University of Hong Kong  
**HSUHK:** Hang Seng University of Hong Kong  
**NGO:** nongovernmental organization  
**PHQ-2:** Patient Health Questionnaire-2  
**RCT:** randomized controlled trial



**RPS:** risk propensity score

**STI:** sexually transmitted infection

*Edited by G Eysenbach; submitted 26.09.19; peer-reviewed by M Lim, C Li; comments to author 18.11.19; revised version received 25.02.20; accepted 26.07.20; published 30.10.20.*

*Please cite as:*

Wong WCW, Sun WH, Chia SMC, Tucker JD, Mak WPH, Song L, Choi KKY, Lau STH, Wan EYF

*Effectiveness of a Peer-Led Web-Based Intervention to Improve General Self-Efficacy in Using Dating Apps Among Young Adults: Randomized Clustered Trial*

*J Med Internet Res* 2020;22(10):e16378

URL: <http://www.jmir.org/2020/10/e16378/>

doi: [10.2196/16378](https://doi.org/10.2196/16378)

PMID: [33124987](https://pubmed.ncbi.nlm.nih.gov/33124987/)

©William CW Wong, Wai Han Sun, Shu Ming Cheryl Chia, Joseph D Tucker, William PH Mak, Lin Song, Kitty Wai Ying Choi, Stephanie Tsz Hei Lau, Eric Yuk Fai Wan. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Effects of a Mobile and Web App (Thought Spot) on Mental Health Help-Seeking Among College and University Students: Randomized Controlled Trial

David Wiljer<sup>1,2,3,4</sup>, PhD; Jenny Shi<sup>2</sup>, MPH; Brian Lo<sup>1,2,3,5</sup>, MHI; Marcos Sanches<sup>6</sup>, MSc; Elisa Hollenberg<sup>2</sup>, MSW; Andrew Johnson<sup>2</sup>, BA; Alexxa Abi-Jaoudé<sup>2</sup>, MPH; Gloria Chaim<sup>4,7</sup>, MSW; Kristin Cleverley<sup>4,7,8</sup>, RN, PhD; Joanna Henderson<sup>4,7</sup>, PhD; Wanrudee Isaranuwatthai<sup>3,9</sup>, PhD; Andrea Levinson<sup>4,10</sup>, MSc, MD; Janine Robb<sup>10</sup>, RN, MSc; Howard W Wong<sup>2,3,11</sup>, MHI; Aristotle Voineskos<sup>4,11</sup>, MD, PhD

<sup>1</sup>UHN Digital, University Health Network, Toronto, ON, Canada

<sup>2</sup>Office of Education, Centre for Addiction and Mental Health, Toronto, ON, Canada

<sup>3</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

<sup>4</sup>Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

<sup>5</sup>Information Management Group, Centre for Addiction and Mental Health, Toronto, ON, Canada

<sup>6</sup>Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada

<sup>7</sup>Margaret and Wallace McCain Centre for Child, Youth & Family Mental Health, Centre for Addiction and Mental Health, Toronto, ON, Canada

<sup>8</sup>Lawrence S Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

<sup>9</sup>Centre for Excellence in Economic Analysis Research, St Michael's Hospital, Toronto, ON, Canada

<sup>10</sup>Health and Wellness, University of Toronto, Toronto, ON, Canada

<sup>11</sup>Slaight Family Centre for Youth in Transition, Centre for Addiction and Mental Health, Toronto, ON, Canada

## Corresponding Author:

David Wiljer, PhD

UHN Digital

University Health Network

190 Elizabeth Street

R. Fraser Elliott Building RFE 3-411

Toronto, ON

Canada

Phone: 1 416 340 4800 ext 6322

Email: [david.wiljer@uhn.ca](mailto:david.wiljer@uhn.ca)

## Abstract

**Background:** Mental health disorders are the most prevalent health issues among postsecondary students, yet few solutions to this emerging crisis exist. While mobile health technologies are touted as promising solutions for the unmet mental health needs of these students, the efficacy of these tools remains unclear. In response to these gaps, this study evaluates Thought Spot, a mobile and web app created through participatory design research.

**Objective:** The goal of the research is to examine the impact of Thought Spot on mental health and wellness help-seeking intentions, behaviors, attitudes, self-stigma, and self-efficacy among postsecondary students in Canada.

**Methods:** A 2-armed randomized controlled trial involving students from three postsecondary institutions was conducted. Students were eligible if they were aged 17 to 29 years, enrolled in full-time or part-time studies, functionally competent in English, and had access to a compatible digital device. The usual care group received a mental health services information pamphlet. The intervention group received the Thought Spot app on their digital device. Thought Spot is a standalone app that allows users to add, review, and search crowdsourced information about nearby mental health and wellness services. Users can also track their mood on the app. Outcomes were self-assessed through questionnaires collected at baseline and 3 and 6 months. The primary outcome was change in formal help-seeking intentions from baseline to 6 months, measured by the General Help-Seeking Questionnaire. A mixed-effects model was used to compare the impact of usual care and intervention on the primary outcome (formal help-seeking intentions). Secondary outcomes included changes in informal help-seeking intentions and help-seeking behaviors, help-seeking attitudes, self-stigma, and self-efficacy.

**Results:** A total of 481 students were randomized into two groups: 240 to usual care, and 241 to the intervention group. There were no significant differences in help-seeking intentions between the usual care and intervention groups over 6 months ( $F_{2,877}=0.85$ ;  $P=.43$ ,  $f=0.04$ ). Both groups demonstrated similar increases in formal help-seeking intentions at 3 and 6 months ( $F_{2,877}=23.52$ ;  $P<.001$ ,  $f=0.21$ ). Compared with males, females sought more help from formal resources (OR 1.86; 95% CI 1.22 to 2.83,  $P=.001$ ). Females were less likely to seek help from informal sources than males (OR 0.80; 95% CI 0.22 to 0.73,  $P<.001$ ).

**Conclusions:** Prompting postsecondary students about mental health and help-seeking appears to increase help-seeking intentions. mHealth interventions may be as effective as information pamphlets in increasing formal help-seeking but may confer a small advantage in driving help-seeking from informal sources. Although there is enthusiasm, developers and health policy experts should exercise caution and thoroughly evaluate these types of digital tools. Future studies should explore the cost-effectiveness of digital interventions and develop strategies for improving their efficacy.

**Trial Registration:** ClinicalTrials.gov NCT03412461; <https://clinicaltrials.gov/ct2/show/NCT03412461>

**International Registered Report Identifier (IRRID):** RR2-10.2196/resprot.6446

(*J Med Internet Res* 2020;22(10):e20790) doi:[10.2196/20790](https://doi.org/10.2196/20790)

## KEYWORDS

crowdsourcing; help-seeking behavior; mental health; mobile applications; randomized controlled trial; school mental health services; social support; young adult

## Introduction

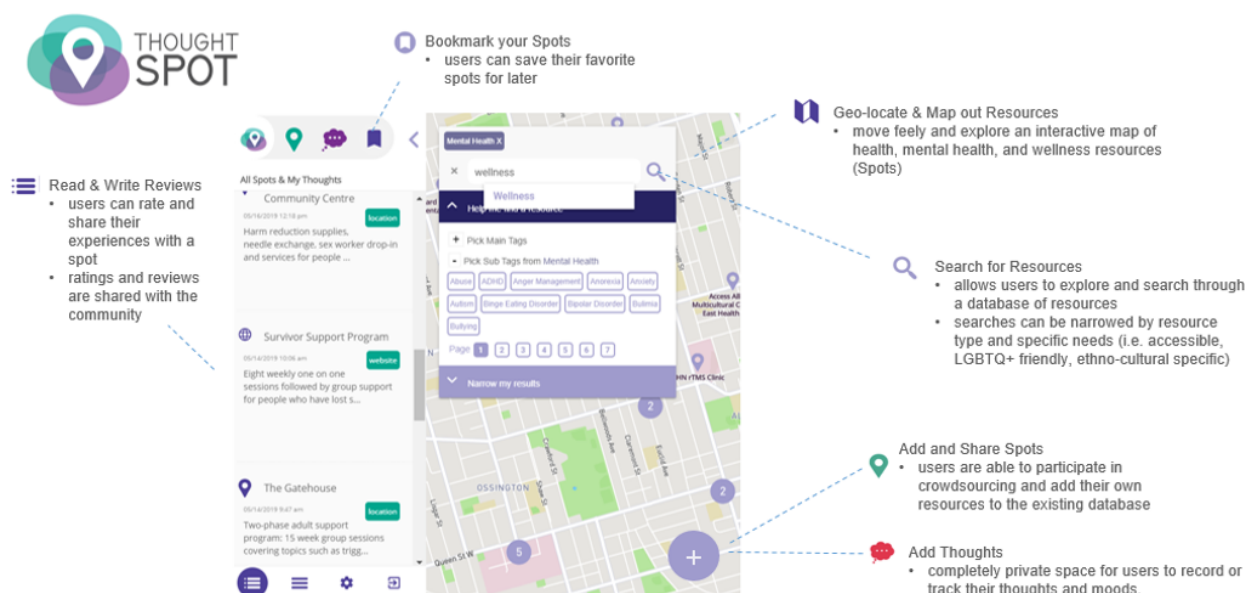
Mental health disorders among postsecondary students are a global public health concern [1-4]. Youth aged 16 to 29 years face many challenges in the transition from childhood to adulthood, and 70% of all mental health conditions have their onset during this period [5-8]. Postsecondary education is an experience that can pose many challenges for transition-aged youth due to social, financial, and academic stressors [3,9-11]. Over the last 10 years, rates of mental health disorders among postsecondary students have increased [12]. Yet 35% of youth who are experiencing mental health problems do not seek formal help (eg, clinical services, health professionals) [13] or informal help (eg, friends and family, religious leaders, self-help) [13] due to barriers such as perceived stigma, difficulty expressing their concerns, difficulty accessing help, and a preference for self-reliance [14,15]. Amid the “campus mental health crisis” [1], there is a desperate need for interventions that facilitate help-seeking and access to mental health and wellness services [16].

The proliferation of mobile devices and their ubiquity in the lives of transition-aged youth has encouraged developers, postsecondary institutions, and health care organizations to focus their efforts on online and mobile health (mHealth) interventions that address the mental health challenges faced by this population [17-20]. The evidence to support mHealth solutions has not been established, but they remain a major hope for transition-aged youth, given the extraordinary challenges

and costs of addressing the mental health needs of postsecondary students [21-23]. To date, assessments of most mHealth interventions have been limited to case studies, pilot studies, and randomized controlled trials (RCTs) with small sample sizes [22]. Few studies involve user-centered design processes or end users as co-creators, evidence-based strategies that help mHealth tools achieve sustained engagement, effectiveness, and behavior change [24].

Applying principles from participatory action and participatory design research [25,26], the research team produced Thought Spot, an mHealth intervention co-created with transition-aged youth that aims to improve help-seeking behavior related to mental health services for postsecondary settings [27,28]. Thought Spot serves as a map-based database (via a mobile app and website) that allows users to search and geolocate health, mental health, and wellness resources (Figure 1). Additional features of the app include mood and thought tracking, reviews about nearby resources and services, and the ability to bookmark searched information. Thought Spot users can also participate in crowdsourcing by adding new resources and writing reviews. Accordingly, the main objective of this RCT was to assess the impact of Thought Spot on intentions to seek formal help. The secondary objective was to examine the impact on intentions to seek help from informal sources and on help-seeking behaviors, help-seeking attitudes, self-stigma, and self-efficacy [17]. The research team hypothesized that Thought Spot would be superior to school-specific mental health services information pamphlets in increasing formal help-seeking intentions.

**Figure 1.** Features on the Thought Spot mobile app and online platform, designed by students for students to allow users to find and share health, mental health, and wellness resources (spots) using a map-based database of crowdsourced resources and a self-contained search feature.



## Methods

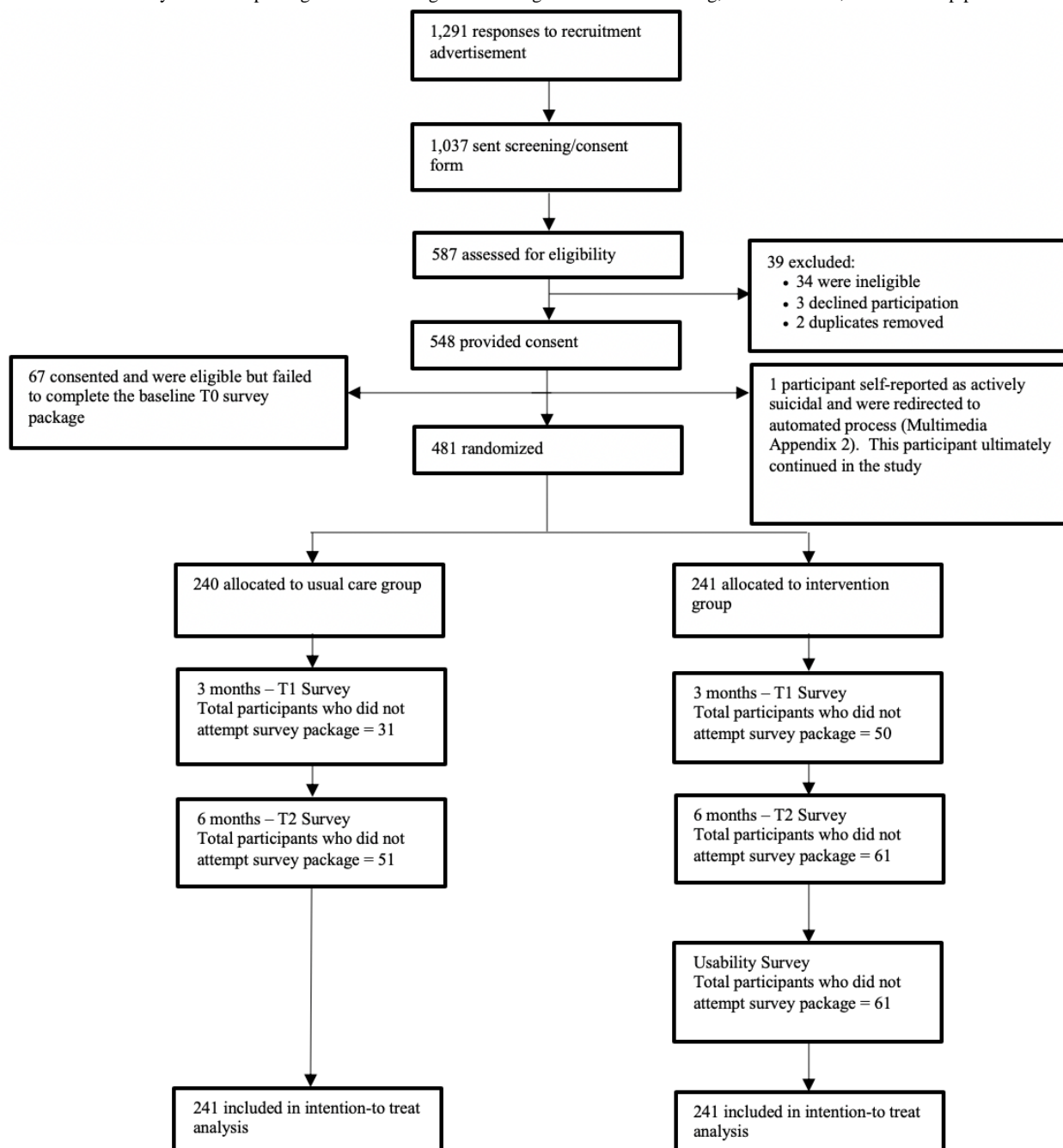
### Study Design

The team conducted a 2-armed RCT using participants who were students at three postsecondary institutions in the Greater Toronto Area (University of Toronto, Ryerson University, and George Brown College). The study staff and biostatistician were blinded throughout the study. The protocol was approved by the research ethics boards of each participating postsecondary institution and the Centre for Addiction and Mental Health and has been previously published ([Multimedia Appendix 1](#)) [27]. The trial was registered at ClinicalTrials.gov [NCT03412461]. Digital informed consent was obtained from each participant through Research Electronic Data Capture (REDCap) before participants could be included in the trial [29].

### Recruitment

Participants were recruited using the most effective methods identified by two focus groups and the Thought Spot youth advisory committee during earlier stages of the project [28]. Participants were recruited through institutional and student-related listservs, bulletin boards, websites, social media, and class presentations.

Full-time and part-time students aged 17 to 29 years who were enrolled at any of the three postsecondary institutions, who were functionally competent in English, and who had access to a digital device compatible with the intervention were eligible to participate in the study. Active suicidality was the sole exclusion criterion, but no participants met this criterion during the screening process ([Figure 2](#); [Multimedia Appendix 2](#)). Participants also did not have knowledge of or access to Thought Spot prior to participating in the study.

**Figure 2.** Consolidated System of Reporting Trials flow diagram detailing the order of screening, randomization, and follow-up procedures.

## Randomization and Treatment

Participants were randomly assigned using a 1:1 allocation ratio to either receive access to Thought Spot or to receive a school-specific mental health services information pamphlet. Randomization and allocation was performed using REDCap, a secure, browser-based, electronic data capture system [29]. A randomization module within REDCap provided computer-generated random allocation of participants in blocks of 10. All participants were enrolled by EH. All authors were blind to the allocation except for JS, who oversaw the randomization process. REDCap was configured to collect data through online questionnaires at baseline, 3 months and 6 months. We asked participants in both groups to use their intervention (Thought Spot or pamphlet) as needed over 6 months. Intervention group participants received detailed

instructions via email on how to download and access the Thought Spot app on their personal electronic devices. Participants were also sent two study reminders, at 6 weeks and 18 weeks, in anticipation of the 3-month and 6-month questionnaires. At the end of the study, participants in the intervention group completed an adapted version of the Usefulness, Satisfaction, and Ease of Use questionnaire [30] and were invited to participate in qualitative interviews to evaluate Thought Spot's usability and user experience. Full details on the randomization process and interventions are provided in the trial protocol in [Multimedia Appendix 1](#). There were no changes to methods or major changes to the features and functionalities of the app. The content on the app changed by design because participants could add their own crowdsourced information about mental health resources and



services during the trial. There were multiple bug fixes to address usability issues, such as malfunctioning buttons, slowness of search features, and log-in credentials not being saved. There were no major downtimes during the study period. Participants were provided with a small honorarium (up to CAD \$40 [USD \$30.08]) for participating in the study.

### Assessment Measures

The primary outcome was a change in formal help-seeking intentions from baseline to 6 months measured by the General Help-Seeking Questionnaire (GHSQ) [31]. This scale comprises ten 7-point Likert scale questions that ask participants how likely they are to seek help from different sources, including friends, family members, mental health professionals, and helplines [31]. The formal help-seeking intentions scores were calculated by summing the scores of the 3 questions measuring formal help-seeking intentions [31]. Secondary outcomes of the study included changes in informal help-seeking intentions and help-seeking behaviors, help-seeking attitudes, self-stigma, and self-efficacy, measured, respectively, by the Actual Help-Seeking Questionnaire (AHSQ) [31], Attitudes Toward Seeking Professional Psychological Help Scale–Short Form (ATSPPH-SF) [32], Self-Stigma of Seeking Help Scale (SSOSH) [33], and the Youth Efficacy/Empowerment Scale–Mental Health (YES-MH) [34].

### Statistical Analysis

Statistical analysis was performed by a biostatistician (MS) who was blinded to randomization. To determine the required sample size, power calculations were conducted using the primary outcome, the average formal help-seeking score on the GHSQ

[27]. Monte Carlo simulations were conducted in SAS 9.4 for Windows (SAS Institute Inc) using the means, standard deviations, and within-subject associations reported in previous research using the GHSQ [35,36]. Monte Carlo simulations were chosen because our hypothesis required a model that accounted for dropouts, random effects, and the specific use of linear contrast to compare pre with post. The calculation accounted for a 40% dropout rate and assumed the use of a mixed-effects model to test the primary hypothesis at a significance level of .05 (2-tailed). A sample of 472 participants (236 per arm) was found to provide 80% power to detect a small effect size (Cohen  $d=0.25$ ), which is equivalent to a 15% difference in change in GHSQ score from baseline to 6 months.

Univariate analyses were conducted to describe groups at baseline (Table 1) and compare completers and noncompleters of the questionnaires. Completers were defined as participants who finished the survey at all 3 points of the study [27]. To evaluate the primary hypothesis, the research team used mixed-effects models using the intention-to-treat method, where time, group, and gender were entered as fixed effects and participants were entered as random effects. A linear contrast of estimated marginal means from the fixed-effect interaction between time and group was conducted to compare changes from baseline to 6 months across groups. Missing values were handled by maximum likelihood estimation, which is able to incorporate all available information in the data under the missing at random assumptions [37]. The team conducted sensitivity analyses, in which we added model variables found to have significant bivariate associations with dropouts, defined as participants who were randomized but who did not complete all 3 surveys, at a significance level of  $\alpha=.20$ .

**Table 1.** Baseline characteristics of participants who underwent randomization<sup>a</sup>.

Characteristics	Control (n=240)	Treatment (n=241)
<b>Gender, n (%)</b>		
Female	188 (78.3)	190 (78.6)
Male	47 (19.6)	44 (18.9)
Nonbinary	5 (2.1)	7 (2.5)
Age in years, mean (SD)	23.2 (3.1)	22.9 (3.4)
<b>Student status, n (%)</b>		
Domestic student	207 (86.3)	207 (85.9)
International student	31 (12.9)	33 (13.7)
Don't know	2 (0.8)	1 (0.4)
<b>Education level, n (%)</b>		
High school diploma	144 (60.0)	140 (58.1)
College diploma	23 (9.6)	9 (3.7)
Bachelor's degree	59 (24.6)	74 (30.7)
Master's degree	11 (4.6)	9 (4.2)
Doctoral degree	0	0
Other	3 (1.3)	9 (3.7)
<b>Type of postsecondary institution<sup>b</sup>, n (%)</b>		
College	110 (45.8)	114 (47.3)
University	130 (54.2)	125 (51.9)
Did not answer	0	2 (0.8)
<b>Experience with mental health or substance use concerns, n (%)</b>		
Yes	168 (70.0)	172 (70.8)
No	55 (23.0)	63 (26.1)
Don't know	16 (6.7)	6 (4.6)
Did not answer	1 (0.4)	0
GHSQ <sup>c</sup> formal sources, mean (SD)	8.6 (3.9)	8.3 (4.1)
GHSQ informal sources, mean (SD)	36.6 (8.5)	36.1 (9.4)

<sup>a</sup>Percentages may not total 100 due to rounding. There was no significant difference between the trial groups.

<sup>b</sup>Postsecondary institution was missing for two participants.

<sup>c</sup>GHSQ: General Help-Seeking Questionnaire.

Similar models were conducted to explore the secondary outcomes of this study except for help-seeking behavior. This outcome was measured using the binary-scaled AHSQ and analyzed using a mixed binomial logistic regression model. The *P* values for the secondary analysis were not adjusted for multiple comparison. In addition to the prespecified analysis plan, the research team conducted an exploratory per-protocol analysis of the primary and secondary outcomes. Participants who logged on to Thought Spot more than once during the study were considered compliant.

For *F* tests, the study report standardized effect sizes using Cohen's *f* [38], reported as *f*, which is calculated by the R package sjstats (R Foundation for Statistical Computing) [39]. For contrasts that represent simple changes or difference in changes, Cohen's *d* was used, which is reported as *d*. The respective *d*

values are calculated using estimated marginal means and two formulas described by Morris and DeShon [40]. The biostatistician applied their recommended formula for repeated-measures design that focuses on change within a person relative to the variability of change scores and another formula used for independent groups pretest/posttest designs (Multimedia Appendix 3) [40].

## Results

### Recruitment and Participant Characteristics

From March 2018 to January 2019, 481 participants were randomized from three Canadian postsecondary institutions into a 6-month trial on a rolling basis. Of these participants, 240 were assigned to the control group and 241 were assigned to

the intervention group (Figure 2). The trial ended after all participants received their 6-month follow-up on June 30, 2019. Table 1 shows a comparison of study groups at baseline. Prior experience with mental health concerns was reported by 70.7% (340/481) of participants, and 61.3% (294/480) of participants reported having experienced suicidal ideation in their lifetime. There were no reported adverse events or harms during the trial.

### Primary Outcome: Change in Formal Help-Seeking Intentions

The mixed-effects model found a significant time effect ( $F_{2,877}=23.52$ ;  $P<.001$ ,  $f=0.21$ ) but not a significant group-by-time interaction ( $F_{2,877}=0.85$ ;  $P=.43$ ,  $f=0.04$ ). Linear contrasts exploring the change in formal help-seeking intentions between baseline and 6 months did not find significant differences between the intervention and control group (estimate mean difference in change 0.39;  $t_{894}=1.05$ ; 95% CI  $-0.34$  to  $1.12$ ;  $P=.29$ ;  $d=0.09$ ). Both the control group (estimated mean change 1.32; 95% CI 0.81 to 1.84,  $d=0.33$ ) and the intervention group (estimated mean difference 0.93, 95% CI 0.41 to 1.45,  $d=0.24$ ) showed a significant increase in the intention to seek

formal sources of help. A sensitivity analysis revealed that group-by-time interaction was not significant after controlling for factors that were associated with dropout ( $F_{2,145}=1.18$ ,  $P=.31$ ,  $f=0.11$ ). There was also no difference in changes between groups (estimate 1.35, 95% CI  $-0.51$  to  $3.21$ ,  $d=0.32$ ).

### Secondary Outcomes

An analysis of secondary outcomes found no significant group-by-time interactions for help-seeking intentions from informal sources (GHSQ), help-seeking behavior from formal sources (AHSQ), help-seeking attitudes toward professional help (ATSPPH), self-efficacy (SSOSH) and self-stigma (YES-MH; Table 2). However, some evidence of a group-by-time interaction was found for help-seeking from informal sources (Wald  $\chi^2=4.5$ ,  $P=.11$ ), but it is not significant (Table 3). Linear contrasts of the interaction indicated a decrease in help-seeking behavior related to informal sources in the control group, whereas an increase from 3 months to 6 months was found for the intervention group (odds ratio [OR] 0.86; 95% CI 0.71 to 1.02,  $P=.09$ ), but it did not meet the significance threshold (Multimedia Appendix 4).

**Table 2.** Main results from mixed models.

Participant questionnaires	Linear contrast <sup>a</sup>			Group-by-time interaction <sup>b</sup>		
	Estimate <sup>a</sup>	95% CI	P value	F statistic	DF <sup>c</sup>	P value
<b>GHSQ<sup>d</sup> (intention to treat)</b>						
Formal resources	0.39	$-0.34$ to $1.12$	.30	0.85	2/877	.43
Informal resources	0.36	$-1.21$ to $1.92$	.65	0.70	2/811	.50
GHSQ (per-protocol) – formal resources	0.18	$-0.62$ to $0.98$	.66	0.14	2/757	.87
GHSQ (dropout) <sup>b</sup> – formal resources	1.35	$-0.51$ to $3.21$	.15	1.18	2/145	.31
GHSQ (controlled for site) <sup>e</sup> – formal resources	0.39	$0.34$ to $1.13$	.29	0.85	2/877	.43
ATSPPH <sup>f</sup>	$-0.15$	$-0.91$ to $0.61$	.70	1.39	2/850	.25
SSOSH <sup>g</sup>	0.08	$-1.13$ to $1.28$	.90	0.03	2/876	.97
YES-MH <sup>h</sup> : self	$-0.09$	$-0.73$ to $0.55$	.79	0.12	2/801	.89
YES-MH: service	0.47	$-0.40$ to $1.34$	.29	0.60	2/823	.55
YES-MH: system	0.62	$-0.26$ to $1.50$	.17	1.31	2/818	.27
YES-MH: total	0.16	$-1.52$ to $1.83$	.85	0.08	2/738	.92

<sup>a</sup>The linear contrast tests the change from baseline to end of the trial across groups (primary outcome); (6 months – baseline) in control minus intervention. Positive values indicate larger increases in the intervention group.

<sup>b</sup>The group-by-time interaction tests for any difference in the group trajectories (not primary outcome). Sensitivity analysis controls for all variables at  $\alpha=.20$ . These variables are listed in Multimedia Appendix 7.

<sup>c</sup>DF: degree of freedom.

<sup>d</sup>GHSQ: General Help-Seeking Questionnaire.

<sup>e</sup>An additional sensitivity analysis controlling for site as fixed effects was conducted, but site was not significant and results were similar to those for the main model.

<sup>f</sup>ATSPPH: Attitudes Toward Seeking Professional Psychological Help Scale.

<sup>g</sup>SSOSH: Self-Stigma of Seeking Help Scale.

<sup>h</sup>YES-MH: Youth Efficacy/Empowerment Scale–Mental Health.

**Table 3.** Main results from binomial models.

Participant questionnaires	Linear contrast <sup>a</sup>			Group-by-time interaction <sup>b</sup>		
	Odds ratio	95% CI	P value	Wald $\chi^2$	DF <sup>c</sup>	P value
AHSQ <sup>d</sup> – formal resources	0.80	0.48-1.34	.39	1.3	2	.53
AHSQ – informal resources	0.86	0.71-1.02	.09	4.5	2	.11

<sup>a</sup>In a binomial model, the contrast is the ratio of odds ratios. A ratio lower than 1 indicates a larger increase in the probability of positive answers in the intervention group.

<sup>b</sup>The statistic used in binomial models is the chi-square statistic.

<sup>c</sup>DF: degree of freedom.

<sup>d</sup>AHSQ: Actual Help-Seeking Questionnaire.

The sex effect was significant for help-seeking behavior related to formal and informal sources, help-seeking attitudes toward professional help, and self-efficacy (Multimedia Appendix 5 and 6). Compared with males, females exhibited higher formal AHSQ scores (OR 1.86; 95% CI 1.22 to 2.83,  $P=.001$ ). Females also exhibited significantly lower AHSQ informal scores than males (OR 0.80; 95% CI 0.22 to 0.73,  $P<.001$ ). Similarly, for attitudes toward health care professionals, females had significantly lower ATSPPH scores than males (estimate 0.80; 95% CI 0.22 to 1.38,  $P=.003$ ).

### Per-Protocol Analysis

Of the participants in the intervention group, 70.1% (169/241) met the definition of compliance. In comparing these participants to the control group, the per-protocol analysis led to the same conclusion for all primary and secondary outcomes with nonsignificant linear contrast (difference in change 0.18,  $t_{711}=0.439$ ,  $P=.66$ ,  $d=0.03$ ). In the intervention group, compliance significantly moderated the GHSQ trajectory ( $P=.006$ ,  $F=0.15$ ), with compliant participants increasing their help-seeking more than noncompliant participants. This change mostly happened between baseline to 3 months.

### App Use Data

Of the 241 people randomized to the intervention group, 168 visited Thought Spot, resulting in 3696 clicks recorded between March 2018 and June 2019. Overall, users viewed 190 Spots, conducted 293 searches, and created 74 Thoughts. Spots are locations of mental health and wellness resources. Thoughts are users' personal journal entries on the app. Details of the user data will be published independently of these findings.

## Discussion

### Principal Findings

In the analysis of our primary outcome, there were no significant differences in the formal help-seeking intentions of postsecondary students between control and intervention groups. However, both groups experienced a similar increase in formal help-seeking intentions during the 6-month study period, as assessed by the GHSQ. These findings suggest that prompting youth about mental health, regardless of the delivery method (eg, information pamphlet or mHealth intervention), may increase help-seeking intentions. Our results are consistent with findings from previous RCTs involving online mental health services, in which the interventions did not lead to significant

differences in formal help-seeking compared with an active control group [41,42]. However, the similar increase in help-seeking in both arms observed in our study supports the emerging discussion in that these technologies may have a “role as another option in their toolkit” when in need of mental health and wellness resources [41].

Analyses of secondary outcomes revealed no significant group-by-time interactions for help-seeking intentions from informal sources, attitudes toward seeking professional help, self-efficacy, or self-stigma. Although not significant, there is some evidence suggesting a small increase in help-seeking behavior related to informal sources between 3 and 6 months in the intervention group, whereas a small decrease was seen in the control group. This difference contrasts with findings from previous studies of help-seeking interventions, which reported no effect on informal help-seeking [42]. Although the attitudes and intentions regarding informal help-seeking did not differ between groups, the sustained increase in informal help-seeking behavior observed between 3 and 6 months in the intervention group may indicate that Thought Spot is effective in converting intention into action [43]. The difference may also be due to previously reported advantages of online interventions over typical information pamphlets and formal sources of help, such as being anonymous, less stigmatizing, easier to access and use, and more trustworthy [41,44-49]. Nonetheless, these findings must be contextualized using data from qualitative interviews, usability data, and use data collected throughout the RCT.

When looking at the differences between AHSQ scores, the results suggest that females tend to seek help from formal sources more than males. However, when looking at the informal AHSQ scores, males sought more help from informal sources than females. This observation complements the existing literature on gender differences in transition-aged youths' help-seeking behaviors. Based on a cross-section survey study, Findlay and Sunderland [50] found that females reported contacting formal and informal resources more than males. However, the data from this RCT show that there may be a difference in preference between formal and informal resources between genders. Another interesting finding was related to the help-seeking behavior for the nonbinary gender group. While not significant due to the small sample size, the considerable odds ratio for AHSQ formal scores suggest there may be a small effect where participants in the nonbinary gender group have higher help-seeking behavior than males.

The per-protocol analysis indicated that compliant participants increased their help-seeking more than noncompliant participants, which suggests that repeated visits may contribute to changes in help-seeking intentions. While these exploratory findings should be approached with caution, it supports the emerging interest on the impact of repeated app use on study outcomes [51]. The app evaluation framework, developed by the American Psychiatric Association, highlights the importance of user engagement in the evaluation of apps for clinical use [52]. However, while use of the app is required to enable the expected benefits, the significance of the repeated use in the context of help-seeking and whether use can be representative as meeting their needs remains unclear. As such, this finding warrants further investigation into the use data and the experiences of users.

### Comparison With Prior Work

One of the noteworthy strengths of our trial is its standing as one of the few RCTs with large sample sizes that evaluate the effectiveness of mHealth interventions for help-seeking among youth in postsecondary settings [41]. The research team saw higher than predicted retention rates of at least 73% at both 3 and 6 months. The baseline characteristics between the control and intervention groups were balanced, and the research team controlled for variables associated with the outcome (eg, GHSQ). Another strength of this study is the use of participatory action research and co-creation methodology to engage transition-aged youth throughout the development and execution of the trial [53]. This methodology has been shown to increase participation in mental health care, better address youth concerns, and produce more relevant results [54,55]. The research team was successful in engaging youth across three postsecondary institutions in a longitudinal mental health study. Given the many barriers that make meaningful youth engagement a challenge [56], obtaining 1291 individual responses during our recruitment is notable in its magnitude, especially given the short time frame (relative to other mental health trials). The high level of engagement in the study suggests that our commitment to co-creating solutions resonates with the postsecondary student population [28]. It may also reflect students' enthusiasm for mental health solutions and mHealth-related research, suggesting a need for continued engagement with this population at a time when mental health concerns and suicide rates among youth continue to rise at postsecondary institutions [57,58].

### Limitations

The study had some limitations. Group assignment could not be blinded for participants. There were also software bugs that led to an inconsistent app environment and usability issues during the trial. These issues could have affected the level of user engagement and compliance with the intervention and ultimately the effectiveness of Thought Spot because some participants had difficulty accessing key functions of the app during certain points of the trial.

There was also a number of participants who did not complete the 3-month survey packages. There were no software bugs or reported issues that could have led to the drop in participation and the reason for this observation is unknown.

Additionally, the effect size for the change in informal help-seeking behavior was small and was noted only in the 3- to 6-month period. Similarly, due to sample size restrictions, we were unable to compare the impact of Thought Spot among international and domestic students. Given the unique barriers associated with attending college in an unfamiliar location [59], future investigations should explore whether these apps can facilitate help-seeking for international students. Further investigation and a longer follow-up period are required to evaluate the sustained impact of mHealth solutions on informal help-seeking. Finally, the development of Thought Spot and the findings from this trial are based on the unique experiences of transition-aged youth enrolled in Canadian postsecondary institutions and may not be generalizable to youth outside Canadian academic settings.

### Conclusions

In summary, there were no significant differences in formal help-seeking intentions between the control and intervention groups. Female participants sought help from formal resources more often than males, whereas males were more likely to seek help from informal sources than females. There was some evidence of a small increase in informal help-seeking behavior between 3 and 6 months in the intervention group. Both groups experienced a similar increase in formal help-seeking intentions over 6 months. These findings suggest a need to further explore the effectiveness of mHealth technologies in supporting the mental health help-seeking needs of transition-aged youth. It is increasingly important as a next step to compare the cost-effectiveness of Thought Spot and information pamphlets for understanding the feasibility and sustainability of mHealth tools compared with existing strategies [60].

### Acknowledgments

This project was supported by eHealth Innovation Partnership Program grant from the Canadian Institute for Health Research (EH1-143558). The Canadian Institute for Health Research was not involved in the study design, data analysis, data interpretation, or writing of the report.

We thank all students who participated in the trial and in our youth engagement activities. We also thank all Centre for Addiction and Mental Health (CAMH) staff, collaborators, youth advisory committee members, and practicum students who were involved and supported the Thought Spot project, including: Daryl Boshart, Almerinda Collela, Paulysha De Gannes, Dr Gunther Eysenbach, Genevieve Ferguson, Olivia Heffernan, Tyson Herzog, Dr Jeffrey Hoch, Dr Huan Jiang, Mande Kaur, Dr Marcus Law, Julia Roy, Shehab Sennah, Sarah Sharpe, Angie Solomon, Chelsea Stunden, Vivetha Thambinathan, Tim Tripp, Nicole VanHeerwaarden,



Julie Zasada, and Hema Zbogor. We thank the CAMH REDCap Operations Committee for their assistance in implementing the e-consent process and survey delivery procedures. QoC Health served as the software developers of the Thought Spot platform.

### Conflicts of Interest

None of the authors have conflicts of interest to disclose.

#### Multimedia Appendix 1

Thought Spot randomized controlled trial protocol and statistical analysis plan.

[PDF File (Adobe PDF File), 355 KB - [jmir\\_v22i10e20790\\_app1.pdf](#)]

#### Multimedia Appendix 2

Automated process for participants self-reporting active suicidality.

[DOCX File, 15 KB - [jmir\\_v22i10e20790\\_app2.docx](#)]

#### Multimedia Appendix 3

Formulas to calculate effect sizes for repeated measures and pretest/posttest research designs.

[DOCX File, 18 KB - [jmir\\_v22i10e20790\\_app3.docx](#)]

#### Multimedia Appendix 4

Group-by-time interaction for help-seeking behavior from informal sources.

[DOCX File, 90 KB - [jmir\\_v22i10e20790\\_app4.docx](#)]

#### Multimedia Appendix 5

Gender effects for help-seeking behaviors, intentions, and attitudes toward professional help and self-efficacy.

[DOCX File, 17 KB - [jmir\\_v22i10e20790\\_app5.docx](#)]

#### Multimedia Appendix 6

Gender effects for help-seeking behaviors, intentions, and attitudes toward professional help and self-efficacy when comparing females and nonbinary to males.

[DOCX File, 191 KB - [jmir\\_v22i10e20790\\_app6.docx](#)]

#### Multimedia Appendix 7

Control variables or interactions that were added to the dropout model aside from the baseline General Help Seeking Questionnaire and sex.

[DOCX File, 15 KB - [jmir\\_v22i10e20790\\_app7.docx](#)]

#### Multimedia Appendix 8

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1349 KB - [jmir\\_v22i10e20790\\_app8.pdf](#)]

### References

1. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, WHO WMH-ICS Collaborators. WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. *J Abnorm Psychol* 2018 Oct;127(7):623-638. [doi: [10.1037/abn0000362](#)] [Medline: [30211576](#)]
2. Ketchen Lipson S, Gaddis SM, Heinze J, Beck K, Eisenberg D. Variations in student mental health and treatment utilization across US colleges and universities. *J Am Coll Health* 2015;63(6):388-396. [doi: [10.1080/07448481.2015.1040411](#)] [Medline: [25942473](#)]
3. Cunningham S, Duffy A. Investing in our future: importance of postsecondary student mental health research. *Can J Psychiatry* 2019 Feb;64(2):79-81 [FREE Full text] [doi: [10.1177/0706743718819491](#)] [Medline: [30819000](#)]
4. Gunnell D, Kidger J, Elvidge H. Adolescent mental health in crisis. *BMJ* 2018 Jun 19;361:k2608. [doi: [10.1136/bmj.k2608](#)] [Medline: [29921659](#)]
5. The human face of mental health and mental illness in Canada. Ottawa: Government of Canada; 2006. URL: [https://www.phac-aspc.gc.ca/publicat/human-humain06/pdf/human\\_face\\_e.pdf](https://www.phac-aspc.gc.ca/publicat/human-humain06/pdf/human_face_e.pdf) [accessed 2020-09-28]
6. MacKinnon N, Colman I. Factors associated with suicidal thought and help-seeking behaviour in transition-aged youth versus adults. *Can J Psychiatry* 2016 Dec;61(12):789-796 [FREE Full text] [doi: [10.1177/0706743716667417](#)] [Medline: [27578771](#)]

7. Caspi A, Houts RM, Ambler A, Danese A, Elliott ML, Hariri A, et al. Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the dunedin birth cohort study. *JAMA Netw Open* 2020 Apr 01;3(4):e203221 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.3221](https://doi.org/10.1001/jamanetworkopen.2020.3221)] [Medline: [32315069](https://pubmed.ncbi.nlm.nih.gov/32315069/)]
8. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. *Lancet* 2016 Jun 11;387(10036):2423-2478. [doi: [10.1016/S0140-6736\(16\)00579-1](https://doi.org/10.1016/S0140-6736(16)00579-1)] [Medline: [27174304](https://pubmed.ncbi.nlm.nih.gov/27174304/)]
9. Zarrett N, Eccles J. The passage to adulthood: challenges of late adolescence. *New Dir Youth Dev* 2006(111):13-28. [doi: [10.1002/yd.179](https://doi.org/10.1002/yd.179)] [Medline: [17225644](https://pubmed.ncbi.nlm.nih.gov/17225644/)]
10. Ruberman L. Challenges in the transition to college: the perspective of the therapist back home. *Am J Psychother* 2014;68(1):103-115. [doi: [10.1176/appi.psychotherapy.2014.68.1.103](https://doi.org/10.1176/appi.psychotherapy.2014.68.1.103)] [Medline: [24818459](https://pubmed.ncbi.nlm.nih.gov/24818459/)]
11. Chan V, Moore J, Derenne J, Fuchs DC. Transitional age youth and college mental health. *Child Adolesc Psychiatr Clin N Am* 2019 Jul;28(3):363-375 [FREE Full text] [doi: [10.1016/j.chc.2019.02.008](https://doi.org/10.1016/j.chc.2019.02.008)] [Medline: [31076114](https://pubmed.ncbi.nlm.nih.gov/31076114/)]
12. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007-2017). *Psychiatr Serv* 2019 Jan 01;70(1):60-63. [doi: [10.1176/appi.ps.201800332](https://doi.org/10.1176/appi.ps.201800332)] [Medline: [30394183](https://pubmed.ncbi.nlm.nih.gov/30394183/)]
13. Rickwood D, Thomas K. Conceptual measurement framework for help-seeking for mental health problems. *Psychol Res Behav Manag* 2012;5:173-183 [FREE Full text] [doi: [10.2147/PRBM.S38707](https://doi.org/10.2147/PRBM.S38707)] [Medline: [23248576](https://pubmed.ncbi.nlm.nih.gov/23248576/)]
14. Salaheddin K, Mason B. Identifying barriers to mental health help-seeking among young adults in the UK: a cross-sectional survey. *Br J Gen Pract* 2016 Oct;66(651):e686-e692. [doi: [10.3399/bjgp16X687313](https://doi.org/10.3399/bjgp16X687313)] [Medline: [27688518](https://pubmed.ncbi.nlm.nih.gov/27688518/)]
15. DiPlacito-DeRango ML. Acknowledge the barriers to better the practices: support for student mental health in higher education. *Can J Scholarship Teaching Learning* 2016 Dec 11;7(2):1. [doi: [10.5206/cjsotl-rcacea.2016.2.2](https://doi.org/10.5206/cjsotl-rcacea.2016.2.2)]
16. Ng P, Padjen M. An overview of post-secondary mental health on campuses in Ontario: challenges and successes. *Int J Ment Health Addiction* 2018 Oct 24;17(3):531-541. [doi: [10.1007/s11469-018-0015-5](https://doi.org/10.1007/s11469-018-0015-5)]
17. Seko Y, Kidd S, Wiljer D, McKenzie K. Youth mental health interventions via mobile phones: a scoping review. *Cyberpsychol Behav Soc Netw* 2014 Sep;17(9):591-602. [doi: [10.1089/cyber.2014.0078](https://doi.org/10.1089/cyber.2014.0078)] [Medline: [25007383](https://pubmed.ncbi.nlm.nih.gov/25007383/)]
18. Bratu S. Fear of missing out, improper behavior, and distressing patterns of use. *Linguist Philos Invest* 2018;17:130-140. [doi: [10.22381/pi1720187](https://doi.org/10.22381/pi1720187)]
19. Roess A. The promise, growth, and reality of mobile health: another data-free zone. *N Engl J Med* 2017 Nov 23;377(21):2010-2011. [doi: [10.1056/NEJMp1713180](https://doi.org/10.1056/NEJMp1713180)] [Medline: [29116869](https://pubmed.ncbi.nlm.nih.gov/29116869/)]
20. Becker TD, Torous JB. Recent developments in digital mental health interventions for college and university students. *Curr Treat Options Psych* 2019 Jun 14;6(3):210-220. [doi: [10.1007/s40501-019-00178-8](https://doi.org/10.1007/s40501-019-00178-8)]
21. Supporting the mental health of emerging adults in Ontario's post-secondary system. Ontario University and College Health Association. 2017. URL: [http://oucha.ca/pdf/mental\\_health/Oucha-Paper-April-2017.pdf](http://oucha.ca/pdf/mental_health/Oucha-Paper-April-2017.pdf) [accessed 2020-09-28]
22. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? *PLoS Med* 2013;10(2):e1001382 [FREE Full text] [doi: [10.1371/journal.pmed.1001382](https://doi.org/10.1371/journal.pmed.1001382)] [Medline: [23424286](https://pubmed.ncbi.nlm.nih.gov/23424286/)]
23. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *Gen Hosp Psychiatry* 2013 Aug;35(4):332-338 [FREE Full text] [doi: [10.1016/j.genhosppsych.2013.03.008](https://doi.org/10.1016/j.genhosppsych.2013.03.008)] [Medline: [23664503](https://pubmed.ncbi.nlm.nih.gov/23664503/)]
24. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. *Biomed Instrum Technol* 2012;Suppl:49-56. [doi: [10.2345/0899-8205-46.s2.49](https://doi.org/10.2345/0899-8205-46.s2.49)] [Medline: [23039777](https://pubmed.ncbi.nlm.nih.gov/23039777/)]
25. Whyte W, Greenwood D, Lazes P. Participatory action research: through practice to science in social research. *Particip Action Res* 1991:19-55. [doi: [10.4135/9781412985383.n2](https://doi.org/10.4135/9781412985383.n2)]
26. Hagen P, Collin P, Metcalf A, Nicholas M, Rahilly K, Swainston N. Participatory Design of evidence-based online youth mental health promotion, prevention, early intervention and treatment. Melbourne: Young and Well Cooperative Research Centre. 2012. URL: [http://about.au.reachout.com/wp-content/uploads/2015/05/Young\\_and\\_Well\\_CRC\\_IM\\_PD\\_Guide.pdf](http://about.au.reachout.com/wp-content/uploads/2015/05/Young_and_Well_CRC_IM_PD_Guide.pdf) [accessed 2020-09-28]
27. Wiljer D, Abi-Jaoude A, Johnson A, Ferguson G, Sanches M, Levinson A, et al. Enhancing self-efficacy for help-seeking among transition-aged youth in postsecondary settings with mental health and/or substance use concerns, using crowd-sourced online and mobile technologies: the thought spot protocol. *JMIR Res Protoc* 2016 Nov 04;5(4):e201 [FREE Full text] [doi: [10.2196/resprot.6446](https://doi.org/10.2196/resprot.6446)] [Medline: [27815232](https://pubmed.ncbi.nlm.nih.gov/27815232/)]
28. VanHeerwaarden N, Ferguson G, Abi-Jaoude A, Johnson A, Hollenberg E, Chaim G, et al. The optimization of an eHealth solution (thought spot) with transition-aged youth in postsecondary settings: participatory design research. *J Med Internet Res* 2018 Mar 06;20(3):e79 [FREE Full text] [doi: [10.2196/jmir.8102](https://doi.org/10.2196/jmir.8102)] [Medline: [29510970](https://pubmed.ncbi.nlm.nih.gov/29510970/)]
29. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009 Apr;42(2):377-381 [FREE Full text] [doi: [10.1016/j.jbi.2008.08.010](https://doi.org/10.1016/j.jbi.2008.08.010)] [Medline: [18929686](https://pubmed.ncbi.nlm.nih.gov/18929686/)]
30. Gao M, Kortum P, Oswald F. Psychometric evaluation of the use (usefulness, satisfaction, and ease of use) questionnaire for reliability and validity. *Proc Hum Fact Ergonom Soc* 2018 Sep 27;62(1):1414-1418. [doi: [10.1177/1541931218621322](https://doi.org/10.1177/1541931218621322)]

31. Rickwood D, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. *Adv Mental Health* 2005 Dec;4(3):218-251. [doi: [10.5172/jamh.4.3.218](https://doi.org/10.5172/jamh.4.3.218)]
32. Fischer E, Farina A. Attitudes toward seeking professional psychological help: a shortened form and considerations for research. *J Coll Stud Dev* 1995;36(4):368-373.
33. Vogel DL, Wade NG, Haake S. Measuring the self-stigma associated with seeking psychological help. *J Couns Psychol* 2006;53(3):325-337. [doi: [10.1037/0022-0167.53.3.325](https://doi.org/10.1037/0022-0167.53.3.325)]
34. Walker J, Powers L. Introduction to the youth self-efficacy scale/mental health and the youth participation in planning scale. Research and Training Centre on Family Support and Children's Mental Health. 2007. URL: <https://www.pathwaysrtc.pdx.edu/pdf/pbCompleteSurveyPacket.pdf> [accessed 2020-09-28]
35. Smith CL, Shochet IM. The impact of mental health literacy on help-seeking intentions: results of a pilot study with first year psychology students. *Int J Mental Health Promot* 2011 Jan;13(2):14-20. [doi: [10.1080/14623730.2011.9715652](https://doi.org/10.1080/14623730.2011.9715652)]
36. Costin DL, Mackinnon AJ, Griffiths KM, Batterham PJ, Bennett AJ, Bennett K, et al. Health e-cards as a means of encouraging help seeking for depression among young adults: randomized controlled trial. *J Med Internet Res* 2009;11(4):e42 [FREE Full text] [doi: [10.2196/jmir.1294](https://doi.org/10.2196/jmir.1294)] [Medline: [19850549](https://pubmed.ncbi.nlm.nih.gov/19850549/)]
37. Geert M, Geert V. Models for Discrete Longitudinal Data. New York: Springer; 2005.
38. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition. Hillsdale: Lawrence Erlbaum; 1988.
39. Lüdtke D. sjstats: statistical functions for regression models version 0.17.6. 2018 Nov 18. URL: <https://zenodo.org/record/1489175> [accessed 2020-02-14]
40. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods* 2002 Mar;7(1):105-125. [Medline: [11928886](https://pubmed.ncbi.nlm.nih.gov/11928886/)]
41. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. *J Med Internet Res* 2014;16(3):e66 [FREE Full text] [doi: [10.2196/jmir.3103](https://doi.org/10.2196/jmir.3103)] [Medline: [24594922](https://pubmed.ncbi.nlm.nih.gov/24594922/)]
42. Xu Z, Huang F, Kösters M, Staiger T, Becker T, Thornicroft G, et al. Effectiveness of interventions to promote help-seeking for mental health problems: systematic review and meta-analysis. *Psychol Med* 2018 Dec;48(16):2658-2667. [doi: [10.1017/S0033291718001265](https://doi.org/10.1017/S0033291718001265)] [Medline: [29852885](https://pubmed.ncbi.nlm.nih.gov/29852885/)]
43. Levin ME, Stocke K, Pierce B, Levin C. Do college students use online self-help? A survey of intentions and use of mental health resources. *J Coll Student Psychother* 2017 Oct 09;32(3):181-198. [doi: [10.1080/87568225.2017.1366283](https://doi.org/10.1080/87568225.2017.1366283)]
44. Collin PJ, Metcalf AT, Stephens-Reicher JC, Blanchard ME, Herrman HE, Rahilly K, et al. ReachOut.com: the role of an online service for promoting help-seeking in young people. *Adv Mental Health* 2014 Dec 17;10(1):39-51. [doi: [10.5172/jamh.2011.10.1.39](https://doi.org/10.5172/jamh.2011.10.1.39)]
45. Shandley K, Austin D, Klein B, Kyrios M. An evaluation of "Reach Out Central": an online gaming program for supporting the mental health of young people. *Health Educ Res* 2010 Aug;25(4):563-574. [doi: [10.1093/her/cyq002](https://doi.org/10.1093/her/cyq002)] [Medline: [20150530](https://pubmed.ncbi.nlm.nih.gov/20150530/)]
46. Bradley KL, Robinson LM, Brannen CL. Adolescent help-seeking for psychological distress, depression, and anxiety using an Internet program. *Int J Mental Health Promotion* 2012 Feb;14(1):23-34. [doi: [10.1080/14623730.2012.665337](https://doi.org/10.1080/14623730.2012.665337)]
47. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The internet as a setting for mental health service utilisation by young people. *Med J Aust* 2010 Jun 7;192(11 Suppl):S22-S26. [Medline: [20528703](https://pubmed.ncbi.nlm.nih.gov/20528703/)]
48. Kummervold PE, Gammon D, Bergvik S, Johnsen JK, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in Norway. *Nord J Psychiatry* 2002;56(1):59-65. [doi: [10.1080/08039480252803945](https://doi.org/10.1080/08039480252803945)] [Medline: [11869468](https://pubmed.ncbi.nlm.nih.gov/11869468/)]
49. Horgan A, Sweeney J. Young students' use of the Internet for mental health information and support. *J Psychiatr Ment Health Nurs* 2010 Mar;17(2):117-123. [doi: [10.1111/j.1365-2850.2009.01497.x](https://doi.org/10.1111/j.1365-2850.2009.01497.x)] [Medline: [20465756](https://pubmed.ncbi.nlm.nih.gov/20465756/)]
50. Findlay L, Sunderland A. Professional and informal mental health support reported by Canadians aged 15 to 24. *Health Reports* 2014;1. [Medline: [255179351](https://pubmed.ncbi.nlm.nih.gov/255179351/)]
51. Huckvale K, Nicholas J, Torous J, Larsen ME. Smartphone apps for the treatment of mental health conditions: status and considerations. *Curr Opin Psychol* 2020 Dec;36:65-70. [doi: [10.1016/j.copsyc.2020.04.008](https://doi.org/10.1016/j.copsyc.2020.04.008)]
52. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. *Psychiatr Serv* 2018 Feb 15;69(5):498-500. [doi: [10.1176/appi.ps.201700423](https://doi.org/10.1176/appi.ps.201700423)] [Medline: [29446337](https://pubmed.ncbi.nlm.nih.gov/29446337/)]
53. Leask CF, Sandlund M, Skelton DA, Altenburg TM, Cardon G, Chinapaw MJM, GrandStand, Safe StepTeenage Girls on the Move Research Groups. Framework, principles and recommendations for utilising participatory methodologies in the co-creation and evaluation of public health interventions. *Res Involv Engagem* 2019;5:2 [FREE Full text] [doi: [10.1186/s40900-018-0136-9](https://doi.org/10.1186/s40900-018-0136-9)] [Medline: [30652027](https://pubmed.ncbi.nlm.nih.gov/30652027/)]
54. Orłowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-being interventions. *JMIR Hum Factors* 2015;2(2):e12 [FREE Full text] [doi: [10.2196/humanfactors.4361](https://doi.org/10.2196/humanfactors.4361)] [Medline: [27025279](https://pubmed.ncbi.nlm.nih.gov/27025279/)]
55. Anderson JE, Lowen CA. Connecting youth with health services: systematic review. *Can Fam Physician* 2010 Aug;56(8):778-784. [Medline: [20705886](https://pubmed.ncbi.nlm.nih.gov/20705886/)]

56. O'Donoghue JL, Kirshner B, McLaughlin M. Introduction: moving youth participation forward. *New Dir Youth Dev* 2002(96):15-26. [doi: [10.1002/yd.24](https://doi.org/10.1002/yd.24)] [Medline: [12630271](https://pubmed.ncbi.nlm.nih.gov/12630271/)]
57. Curtin SC, Heron M. Death rates due to suicide and homicide among persons aged 10-24: United States, 2000-2017. *NCHS Data Brief* 2019 Oct(352):1-8. [Medline: [31751202](https://pubmed.ncbi.nlm.nih.gov/31751202/)]
58. Jaworska N, De Somma E, Fonseka B, Heck E, MacQueen GM. Mental health services for students at postsecondary institutions: a national survey. *Can J Psychiatry* 2016 Dec;61(12):766-775 [FREE Full text] [doi: [10.1177/0706743716640752](https://doi.org/10.1177/0706743716640752)] [Medline: [27310230](https://pubmed.ncbi.nlm.nih.gov/27310230/)]
59. Mori SC. Addressing the mental health concerns of international students. *J Couns Develop* 2000;78(2):137-144. [doi: [10.1002/j.1556-6676.2000.tb02571.x](https://doi.org/10.1002/j.1556-6676.2000.tb02571.x)]
60. Kaur A, Isaranuwachai W, Jaffer A, Ferguson G, Abi-Jaoude A, Johnson A, et al. A web- and mobile-based map of mental health resources for postsecondary students (thought spot): protocol for an economic evaluation. *JMIR Res Protoc* 2018 Mar 29;7(3):e83 [FREE Full text] [doi: [10.2196/resprot.9478](https://doi.org/10.2196/resprot.9478)] [Medline: [29599106](https://pubmed.ncbi.nlm.nih.gov/29599106/)]

## Abbreviations

**AHSQ:** Actual Help-Seeking Questionnaire

**ATSPPH-SF:** Attitudes Toward Seeking Professional Psychological Help Scale–Short Form

**CAMH:** Centre for Addiction and Mental Health

**GHSQ:** General Help-Seeking Questionnaire

**mHealth:** mobile health

**OR:** odds ratio

**RCT:** randomized controlled trial

**REDCap:** Research Electronic Data Capture

**SSOSH:** Self-Stigma of Seeking Help Scale

**YES-MH:** Youth Efficacy/Empowerment Scale–Mental Health

*Edited by G Eysenbach; submitted 17.06.20; peer-reviewed by N Wang, AM Duncan; comments to author 19.07.20; revised version received 10.08.20; accepted 16.08.20; published 30.10.20.*

*Please cite as:*

Wiljer D, Shi J, Lo B, Sanches M, Hollenberg E, Johnson A, Abi-Jaoudé A, Chaim G, Cleverley K, Henderson J, Isaranuwachai W, Levinson A, Robb J, Wong HW, Voineskos A

*Effects of a Mobile and Web App (Thought Spot) on Mental Health Help-Seeking Among College and University Students: Randomized Controlled Trial*

*J Med Internet Res* 2020;22(10):e20790

URL: <http://www.jmir.org/2020/10/e20790/>

doi:[10.2196/20790](https://doi.org/10.2196/20790)

PMID:[33124984](https://pubmed.ncbi.nlm.nih.gov/33124984/)

©David Wiljer, Jenny Shi, Brian Lo, Marcos Sanches, Elisa Hollenberg, Andrew Johnson, Alexxa Abi-Jaoudé, Gloria Chaim, Kristin Cleverley, Joanna Henderson, Wanrudee Isaranuwachai, Andrea Levinson, Janine Robb, Howard W Wong, Aristotle Voineskos. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# The Association Between Electronic Device Use During Family Time and Family Well-Being: Population-Based Cross-Sectional Study

Sheng Zhi Zhao<sup>1</sup>, MPH; Ningyuan Guo<sup>1</sup>, BSc; Man Ping Wang<sup>1</sup>, PhD; Daniel Yee Tak Fong<sup>1</sup>, PhD; Agnes Yuen Kwan Lai<sup>1</sup>, PhD; Sophia Siu-Chee Chan<sup>1</sup>, PhD; Tai Hing Lam<sup>2</sup>, PhD; Daniel Sai Yin Ho<sup>2</sup>, PhD

<sup>1</sup>School of Nursing, University of Hong Kong, Hong Kong, China (Hong Kong)

<sup>2</sup>School of Public Health, University of Hong Kong, Hong Kong, China (Hong Kong)

**Corresponding Author:**

Man Ping Wang, PhD  
School of Nursing  
University of Hong Kong  
21 Sassoon Road  
Pokfulam, HK  
Hong Kong, 000000  
China (Hong Kong)  
Phone: 852 39176636  
Email: [mpwang@hku.hk](mailto:mpwang@hku.hk)

## Abstract

**Background:** Electronic devices (eDevices) may have positive or negative influences on family communication and well-being depending on how they are used.

**Objective:** We examined eDevice use during family time and its association with the quality of family communication and well-being in Hong Kong Chinese adults.

**Methods:** In 2017, a probability-based 2-stage random sampling landline telephone survey collected data on eDevice use in daily life and during family time (eg, family dinner) and the presence of rules banning eDevice use during family dinner. Family communication quality was rated from 0 to 10 with higher scores being favorable. Family well-being was calculated as a composite mean score of 3 items each using the same scale from 0 to 10. The associations of family communication quality and well-being with eDevice use in daily life and during family time were estimated using beta-coefficient ( $\beta$ ) adjusting for sociodemographics. The mediating role of family communication quality in the association between eDevice use and family well-being was analyzed.

**Results:** Of the 2064 respondents (mean age 56.4 [SD 19.2] years, 1269/2064 [61.48%] female), 1579/2059 (76.69%) used an eDevice daily for a mean of 3.6 hours (SD 0.1) and 257/686 (37.5%) used it for 30+ minutes before sleep. As much as 794/2046 (38.81%) often or sometimes used an eDevice during family time including dinner (311/2017, 15.42%); 713/2012 (35.44%) reported use of an eDevice by family members during dinner. Lower family communication quality was associated with hours of eDevice use before sleep (adjusted  $\beta=-.25$ ; 95% CI  $-0.44$  to  $-0.05$ ), and often use (vs never use) of eDevice during family dinner by oneself (adjusted  $\beta=-.51$ ; 95% CI  $-0.91$  to  $-0.10$ ) and family members (adjusted  $\beta=-.54$ ; 95% CI  $-0.79$  to  $-0.29$ ). Similarly, lower family well-being was associated with eDevice use before sleep (adjusted  $\beta=-.26$ ; 95% CI  $-0.42$  to  $-0.09$ ), and often use during family dinner by oneself (adjusted  $\beta=-.48$ ; 95% CI  $-0.83$  to  $-0.12$ ) and family members (adjusted  $\beta=-.50$ ; 95% CI  $-0.72$  to  $-0.28$ ). Total ban of eDevice use during family dinner was negatively associated with often use by oneself (adjusted odds ratio 0.49; 95% CI 0.29 to 0.85) and family members (adjusted odds ratio 0.41; 95% CI 0.28, 0.60) but not with family communication and well-being. Lower family communication quality substantially mediated the total effect of the association of eDevice use time before sleep (61.2%) and often use at family dinner by oneself (87.0%) and by family members (67.8%) with family well-being.

**Conclusions:** eDevice use before sleep and during family dinner was associated with lower family well-being, and the association was substantially mediated by family communication quality. Our results suggest that interventions on smart use of eDevice may improve family communication and well-being.

(*J Med Internet Res* 2020;22(10):e20529) doi:[10.2196/20529](https://doi.org/10.2196/20529)



**KEYWORDS**

eDevice; smartphone; mobile phone; well-being; family dinner; family communication; public health

**Introduction**

Family well-being (also known as family life satisfaction, family welfare, and family functioning), which emphasizes the importance of health and satisfaction of family relationships, is a hallmark of individual happiness and social cohesion [1,2]. While family relationships and the sense of well-being may differ widely by family structure, good family verbal and nonverbal communication is essential regardless [3]. Family well-being in Chinese culture highlights family health, harmony, and happiness (3Hs) with family communication being a core component [4]. Socioeconomic (SES) status has been a robust predictor of family well-being, but recent studies showed that higher income only contributed to its dynamic fluctuation but not to a long-lasting increase due to hedonic adaptation and social comparison [5,6]. Family life and communication are robust predictors of family well-being and our previous trial showed that improved family interactions and communication quality enhanced family well-being [7]. However, in busy modern societies, time for family life and communication is limited.

Information communication technology has transformed interpersonal communication and lifestyle. Internet-based electronic devices (eDevices) including personal computers, tablets, and smartphones may improve well-being through increasing work efficiency and capability, easy access to information, and convenient social interactions using communication tools [8-10]. However, inappropriate use of eDevice is a significant public health issue. Addiction-like symptoms and feelings of dependence, dangerous use (eg, distracted driving), and loss of control were common symptoms of problematic eDevice use associated with various health problems [11]. Obsessive eDevice use was linked to poor sleep quality, sedentary lifestyle, physical inactivity, and obesity [12,13]. Psychological symptoms such as depression, social isolation, low self-esteem, and anxiety have also been associated with problematic eDevice use [14].

Time spent on eDevice reduces leisure and family gathering hours. There is a growing concern that eDevice may detract from, rather than complement, social interactions [15]. The act of ignoring others in favor of smartphone use at a social setting, also called phone snubbing (phubbing), has become increasingly common and is associated with poorer relationship satisfaction [15-17]. Smartphone use during social gatherings may hamper conversation and interaction [18]. As the major domain of family life in which family members express love, care, and support; share value and happiness; and resolve problems, quality family communication is vital for maintaining family well-being [19]. eDevice use, especially smartphone use, is invading traditional family functions, such as family gatherings and dinners, which threatens communication quality and hence family well-being [20,21]. Experiencing family phubbing was associated with lower levels of interpersonal relationship and perceived well-being [20]. We also found that problematic smartphone

use is also associated with poorer family well-being via deprived family communication time and quality [22].

We found no reports on the relationships of eDevice use with perceived family communication quality and family well-being. eDevice use during family time (eg, family dinner) is rarely reported despite its potential link to poor well-being [15]. Some families may adopt rules banning eDevice use during family dinner, but its effect on family well-being is unclear. Hong Kong, as one of the most developed non-Western urbanized cities in China with a high internet penetration rate and a dense family living environment, provided an appropriate platform to observe and understand the eDevice use behavior in family life. We examined eDevice use in daily life, especially during family time, and its associations with family communication quality and well-being in Chinese adults in Hong Kong. The potential mediation effect of communication quality on the association between eDevice use and family well-being was analyzed. Moreover, the prevalence and effect of the rules banning eDevice use during family dinner were assessed.

**Methods****Sampling**

The Hong Kong Family and Health Information Trends Survey (FHinTs) under the FAMILY project was a probability-based cross-sectional telephone survey conducted in 2017 to monitor family health, information use, and communication among Cantonese speaking Hong Kong adults (aged 18+). All interviews were conducted by trained interviewers of the Public Opinion Programme of the University of Hong Kong. Details of the methods were reported elsewhere [10,23,24]. Briefly, adopting a 2-stage random sampling strategy, residential telephone directories that covered 76% of Hong Kong residents were used to generate population-representative telephone numbers by random. Invalid household numbers and nonresponses (after 5 calls at different times and days of the week) were excluded. The second stage was to select an eligible individual in each household with the soonest next birthday. Each interview took 25-30 minutes. Of 5449 eligible households, 4054 adults were successfully interviewed (response rate 74.40%). Of these, 2064 (50.91%) respondents were randomly selected to answer a subset of questions related to the use of computer, smartphone, or tablet (eDevice); family communication; and family well-being. The Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster granted ethics approval. Verbal informed consent was obtained.

**Measurements**

Family included biological, marital, or cohabited relationships. eDevices included computers, smartphones, and tablets. Duration of self-reported eDevice use daily was categorized into "No use," "1 hour or less," "1-2 hours," "2-3 hours," "3-5 hours," and "more than 5 hours." eDevice use before sleep was categorized as "No use," "30 minutes or less," "31-60 minutes",

“1-2 hours,” and “more than 2 hours”. Family time referred to the time spent with family members. The frequency of eDevice use during family time was categorized as “Never,” “Seldom,” “Sometimes,” and “Often.” The same response options were used for the frequency of eDevice use during family dinner by oneself and by family members. Rules banning eDevice use during family dinner were categorized as “No ban,” “Partial ban (limited time or frequency),” and “Total ban.”

Perceived communication quality between family members was rated on a scale of 0 (very poor) to 10 (very good), which positively correlated with the Subjective Happiness Scale, Short War-wick-Edinburgh Mental Well-being Scale, and Family Functioning and Family Communication Scale (Pearson correlation coefficients [ $r$ ] range: .37-.60, all  $P<.001$ ) [22]. Family well-being was calculated based on the composite score of family health, harmony, and happiness (3Hs) using 3 separate questions on the same scale of 0-10, with a higher score indicating better family well-being. The Chinese version of Family Happiness was found valid and reliable among Hong Kong adults [25]. The internal consistency of family well-being in this study was excellent (Cronbach  $\alpha=.89$ ).

Sex, age, and marital status (never married, married or cohabitated, divorced or separated, and widowed) of the respondents were recorded. Education attainment was categorized as primary or below, secondary, and tertiary. Monthly household income was categorized as HK \$9999 or less (US \$1290 or less), HK \$10,000-19,999 (US \$1290-US \$2580), HK \$20,000-29,999 (US \$2580-US \$3870), HK \$30,000-39,999 (US \$3870-US \$5161), HK \$40,000 or more (US \$5161 or more), and unstable [10,23].

### Data Analysis

To improve the representativeness of the sample, all descriptive data were weighted according to the sex–age distribution of the Hong Kong population in the year end of 2015 and the education

attainment distribution in the 2011 census. We used chi-square tests to compare the prevalence of eDevice use by sociodemographics. We also used paired  $t$  test and analysis of variance to compare the average duration of daily eDevice use, family communication, and family well-being across sociodemographic status. Multivariable linear regression was conducted to calculate the coefficient of family communication quality and family well-being in relation to eDevice use during family time, adjusted for potential sociodemographic confounders including sex, age, marital status, education attainment, and income. We used the Baron and Kenny's approach [26] to examine the mediating effect of family communication quality on the association between eDevice use and family well-being. Sobel–Goodman tests were adopted to decompose the total effects of eDevice use on family well-being. Bootstrapping with 1000 replications was used to calculate the 95% CI of the indirect effect. Logistic regression was conducted to estimate the odds ratios of sometimes or often use of eDevice during family dinner in relation to rules banning its use with the same adjustment model. A 2-sided  $P$ -value  $<.05$  was considered statistically significant. All analyses were conducted using STATA 15.0 (StataCorp).

## Results

Table 1 shows that among the 2064 respondents, 1269 were female (61.48%), 1035 aged 25 to 64 years (50.15%), and 1306 married or cohabitated (63.28%). A total of 1582 respondents had secondary or higher educational attainment (76.65%) and 1032 had a monthly household income of more than HK \$20,000 (US \$2580; 50.00%). Respondents who were male, younger, never married, and had higher education attainment and household income used eDevice more (all  $P<.01$ ). Family communication quality and well-being scores were higher among older adults and those who had higher household income (all  $P<.01$ ; Multimedia Appendix 1).

**Table 1.** Sociodemographic characteristics and hours of eDevice<sup>a</sup> use daily (N=2064).

Sociodemographic characteristic	Values		Prevalence of eDevice use		eDevice usage daily (N=1579 hours)	
	Crude n (%)	Weighted % <sup>b</sup>	Crude n (Weighted %)	P value	Mean (SD)	P value
<b>Sex</b>				.007		.004
Male	795 (38.52)	45.03	634 (84.64)		3.7 (3.3)	
Female	1269 (61.48)	54.97	945 (81.82)		3.4 (3.2)	
<b>Age, years</b>				<.001		<.001
18-24	213 (10.32)	9.12	211 (97.69)		5.9 (3.4)	
25-44	297 (14.39)	35.42	290 (96.99)		5.0 (3.6)	
45-64	738 (35.76)	36.96	655 (85.51)		3.1 (2.9)	
≥65	816 (39.53)	18.50	423 (44.11)		1.9 (2.1)	
<b>Marital status</b>				<.001		<.001
Never married	445 (21.56)	29.30	423 (94.09)		5.6 (3.7)	
Married/Cohabitated	1306 (63.28)	60.47	1010 (83.40)		2.9 (2.7)	
Divorced/Separated	79 (3.83)	3.59	52 (69.23)		3.1 (3.2)	
Widowed	234 (11.34)	6.64	94 (38.58)		1.9 (2.0)	
<b>Education attainment</b>				<.001		<.001
≤Primary	482 (23.35)	23.66	193 (50.79)		1.6 (1.6)	
Secondary	864 (41.86)	48.09	700 (89.74)		3.0 (3.0)	
Tertiary	718 (34.79)	28.25	686 (98.71)		4.6 (3.4)	
<b>Monthly household income (HK \$ [US \$])<sup>c</sup></b>				<.001		<.001
≤9999 (≤1290)	478 (23.16)	16.04	227 (47.69)		2.3 (2.5)	
10,000-19,999 (1290-2580)	294 (14.24)	17.00	223 (82.50)		3.1 (3.0)	
20,000-29,999 (2580-3870)	326 (15.79)	19.72	284 (90.32)		3.6 (3.4)	
30,000-39,999 (3870-5161)	222 (10.76)	13.01	205 (96.18)		3.8 (3.2)	
≥40,000 (≥5161)	484 (23.45)	24.00	456 (97.22)		4.3 (3.3)	
Unstable	260 (12.60)	10.23	184 (75.01)		3.2 (3.2)	

<sup>a</sup>eDevice: electronic device.<sup>b</sup>Weighted by sex, age and educational attainment distribution of Hong Kong census.<sup>c</sup>US \$1 = HK \$7.8.

**Table 2** shows that 1579/2059 (76.69%) respondents used eDevice daily for 3.6 hours on average; 257/686 (37.5%) used eDevice for more than 30 minutes before sleep and 794/2046 (38.81%) sometimes or often use eDevice during family time; 311/2017 (15.42%) and 713/2012 (35.44%) reported oneself and family members sometimes or often use eDevice during family dinner, respectively. Only 376/2045 (18.39%) reported a partial or total ban of eDevice use during family dinner.

**Table 3** shows that per-hour increase in using eDevice daily was not associated with family communication quality (adjusted  $\beta=-.02$ , 95% CI  $-0.04$  to  $0.01$ ) and family well-being (adjusted  $\beta=-.01$ , 95% CI  $-0.03$  to  $0.01$ ) after controlling for sociodemographic characteristics. Per-hour increase in eDevice

use before sleep, however, was associated with lower family communication quality and family well-being (all  $P$  for trend  $<.01$ ). Specifically, more than 2 hours of use before sleep was significantly associated with lower family communication quality (adjusted  $\beta=-1.05$ , 95% CI  $-1.86$  to  $-0.24$ ;  $P=.002$ ) and family well-being (adjusted  $\beta=-.97$ , 95% CI  $-1.66$  to  $-0.28$ ;  $P<.001$ ). Often use of eDevice during family dinner both by oneself and by family members was associated with lower family communication quality (adjusted  $\beta=-.54$  to  $-0.51$ ) and family well-being (adjusted  $\beta=-.50$  to  $-0.48$ ). However, overall use of eDevice during family time was not associated with family communication quality (adjusted  $\beta=-.14$ , 95% CI  $-0.41$  to  $0.13$ ) and family well-being (adjusted  $\beta=-.06$ , 95% CI  $-0.29$  to  $0.18$ ).

**Table 2.** Frequency of eDevice<sup>a</sup> use and prevalence of rules banning eDevice use during family dinner.

Content	Crude n (%) <sup>b</sup>	Weighted % <sup>c</sup>
<b>Time using eDevice daily (hours) (N=2059)</b>		
No use	480 (23.31)	16.98
≤1	424 (20.59)	18.74
1-2	291 (14.13)	15.15
2-3	237 (11.51)	11.93
3-5	258 (12.53)	14.70
>5	369 (17.92)	22.51
Mean (SD) in hours among users	3.5 (3.2)	3.6 (0.1)
<b>Time using eDevice before sleep (N=686)</b>		
No use	329 (47.96)	38.43
≤30 minutes	100 (14.58)	16.17
31-60 minutes	179 (26.09)	30.90
1-2 hours	52 (7.58)	9.87
>2 hours	26 (3.79)	4.63
Mean (SD) in hours among users	0.94 (0.9)	0.98 (0.1)
<b>eDevice use during family time (N=2046)</b>		
Never	867 (42.38)	33.85
Seldom	385 (18.82)	19.48
Sometimes	503 (24.58)	29.34
Often	291 (14.22)	17.32
<b>eDevice use during family dinner by oneself (N=2017)</b>		
Never	1329 (65.89)	58.26
Seldom	377 (18.69)	21.34
Sometimes	217 (10.76)	14.56
Often	94 (4.66)	5.84
<b>eDevice use during family dinner by family members (N=2012)</b>		
Never	888 (44.14)	39.46
Seldom	411 (20.43)	22.31
Sometimes	424 (21.07)	23.14
Often	289 (14.36)	15.09
<b>Rules banning eDevice use during family dinner (N=2045)</b>		
No ban	1669 (81.61)	78.34
Partial ban <sup>d</sup>	201 (9.83)	11.21
Total ban	175 (8.56)	10.46

<sup>a</sup>eDevice: electronic device.<sup>b</sup>Sample sizes varied because of missing values on some variables.<sup>c</sup>Weighted to sex, age, and educational attainment distribution of Hong Kong census.<sup>d</sup>Partial ban: Limited time or frequency of use.

**Table 3.** Daily eDevice<sup>a</sup> use in relation to family communication quality and family well-being.

eDevice use	Family communication quality (0-10)		Family well-being (0-10)	
	Crude $\beta$ (95% CI)	Adjusted $\beta$ (95% CI) <sup>b</sup>	Crude $\beta$ (95% CI)	Adjusted $\beta$ (95% CI) <sup>b</sup>
<b>Time using eDevice daily (hours)</b>				
Not used	Reference	Reference	Reference	Reference
≤1	-.14 (-0.32 to 0.04)	-.16 (-0.35 to 0.03)	-.07 (-0.23 to 0.08)	-.12 (-0.28 to 0.05)
1-2	-.16 (-0.36 to 0.03)	-.15 (-0.38 to 0.07)	-.03 (-0.20 to 0.14)	-.07 (-0.26 to 0.13)
2-3	-.29 (-0.51 to -0.08) <sup>c</sup>	-.26 (-0.50 to -0.01) <sup>d</sup>	-.11 (-0.29 to 0.08)	-.12 (-0.33 to 0.09)
3-5	-.26 (-0.46 to -0.05) <sup>d</sup>	-.18 (-0.42 to 0.07)	-.01 (-0.20 to 0.17)	.03 (-0.19 to 0.24)
>5	-.39 (-0.58 to -0.20) <sup>e</sup>	-.24 (-0.49 to 0.01)	-.26 (-0.42 to -0.09) <sup>c</sup>	-.16 (-0.37 to 0.06)
Per-hour increase	-.04 (-0.06 to -0.02) <sup>e</sup>	-.02 (-0.04 to 0.01)	-.02 (-0.04 to -0.01) <sup>c</sup>	-.01 (-0.03 to 0.01)
<b>Time using eDevice before sleep</b>				
Not used	Reference	Reference	Reference	Reference
<30 minutes	-.26 (-0.68 to 0.16)	-.14 (-0.60 to 0.31)	-.08 (-0.44 to 0.28)	-.11 (-0.49 to 0.28)
30-60 minutes	-.40 (-0.74 to -0.05) <sup>d</sup>	-.14 (-0.53 to 0.26)	-.15 (-0.45 to 0.15)	-.04 (-0.37 to 0.30)
61-120 minutes	-.46 (-1.01 to 0.09)	-.13 (-0.72 to 0.46)	-.36 (-0.83 to 0.11)	-.19 (-0.69 to 0.31)
>120 minutes	-1.57 (-2.33 to -0.80) <sup>e</sup>	-1.05 (-1.86 to -0.24) <sup>c</sup>	-1.32 (-1.98 to -0.67) <sup>e</sup>	-.97 (-1.66 to -0.28) <sup>c</sup>
Per-hour increase	-.39 (-0.57 to -0.22) <sup>e</sup>	-.25 (-0.44 to -0.05) <sup>c</sup>	-.34 (-0.49 to -0.19) <sup>e</sup>	-.26 (-0.42 to -0.09) <sup>c</sup>
<b>eDevice use during family time</b>				
Never	Reference	Reference	Reference	Reference
Seldom	.01 (-0.22 to 0.24)	.05 (-0.19 to 0.29)	.07 (-0.13 to 0.27)	.10 (-0.11 to 0.31)
Sometimes	-.08 (-0.29 to 0.13)	.02 (-0.21 to 0.25)	-.06 (-0.25 to 0.12)	.02 (-0.18 to 0.22)
Often	-.25 (-0.50 to 0.09)	-.14 (-0.41 to 0.13)	-.15 (-0.37 to 0.08)	-.06 (-0.29 to 0.18)
<b>eDevice use during family dinner by oneself</b>				
Never	Reference	Reference	Reference	Reference
Seldom	-.03 (-0.25 to 0.19)	.06 (-0.17 to 0.29)	-.02 (-0.21 to 0.18)	.04 (-0.15 to 0.24)
Sometimes	-.19 (-0.47 to 0.08)	-.08 (-0.36 to 0.20)	-.12 (-0.36 to 0.12)	-.03 (-0.28 to 0.22)
Often	-.61 (-1.01 to -0.22) <sup>c</sup>	-.51 (-0.91 to -0.10) <sup>d</sup>	-.58 (-0.93 to -0.23) <sup>c</sup>	-.48 (-0.83 to -0.12) <sup>c</sup>
<b>Use of eDevice during family dinner by family members</b>				
Never	Reference	Reference	Reference	Reference
Seldom	-.09 (-0.31 to 0.13)	-.01 (-0.23 to 0.21)	.00 (-0.19 to 0.19)	.06 (-0.13 to 0.26)
Sometimes	-.29 (-0.51 to -0.08) <sup>c</sup>	-.24 (-0.46 to -0.02) <sup>d</sup>	-.29 (-0.49 to -0.10) <sup>c</sup>	-.25 (-0.45 to -0.06) <sup>c</sup>
Often	-.55 (-0.80 to -0.30) <sup>e</sup>	-.54 (-0.79 to -0.29) <sup>e</sup>	-.51 (-0.73 to -0.28) <sup>e</sup>	-.50 (-0.72 to -0.28) <sup>e</sup>
<b>Rules banning eDevice use during family dinner</b>				
No ban	Reference	Reference	Reference	Reference
Partial ban <sup>f</sup>	-.02 (-0.30 to 0.26)	-.04 (-0.33 to 0.24)	-.11 (-0.36 to 0.14)	-.12 (-0.37 to 0.12)
Total ban	.27 (-0.03 to 0.57)	.21 (-0.08 to 0.37)	.23 (-0.03 to 0.50)	.18 (-0.08 to 0.44)

<sup>a</sup>eDevice: electronic device.<sup>b</sup>Adjusted for sex, age, marital status, education attainment, and family income.<sup>c</sup> $P < .01$ .<sup>d</sup> $P < .05$ .<sup>e</sup> $P < .001$ .



<sup>f</sup>Partial ban: Limited time or frequency of use.

Table 4 shows that family communication quality mediated the association of per-hour increase in eDevice use before sleep and often use of eDevice during family dinner with family well-being by meeting the Baron and Kenny's mediation assumptions [26]. Sobel–Goodman tests showed that family communication quality mediated 61.2%, 87.0%, and 67.8% of the total effect of hourly increase in eDevice use before sleep and eDevice use by oneself and by family members during family dinner on family well-being, respectively. The total ban

was not associated with family communication quality (adjusted  $\beta=-.21$ , 95% CI  $-0.08$  to  $0.37$ ) and family well-being (adjusted  $\beta=-.18$ , 95% CI  $-0.08$  to  $0.44$ ). Table 5 shows that the total ban was associated with decreased eDevice use during family dinner (adjusted odds ratio 0.49, 95% CI 0.29–0.85). Family members were also less likely to use eDevice during dinner with the total ban (adjusted odds ratio 0.41, 95% CI 0.28–0.60). Partial ban showed no effect on eDevice use or well-being.

**Table 4.** Mediation of the association between eDevice<sup>a</sup> use and family by perceived family communication quality.

Communication quality	Family well-being	
	$\beta^b$	95% CI <sup>c</sup>
<b>Hourly increase in eDevice use before sleep<sup>d</sup></b>		
Total effect	–.27	–0.43 to –0.10 <sup>e</sup>
Indirect effect (through mediator)	–.16	–0.30 to –0.02 <sup>f</sup>
Direct effect (without mediator)	–.10	–0.26 to 0.05
<b>Often use of eDevice during family dinner by oneself<sup>g</sup></b>		
Total effect	–.42	–0.77 to –0.07 <sup>e</sup>
Indirect effect	–.28	–0.54 to –0.02 <sup>f</sup>
Direct effect	–.14	–0.40 to 0.12
<b>Often use of eDevice during family dinner by family members<sup>h</sup></b>		
Total effect	–.16	–0.22 to –0.09 <sup>e</sup>
Indirect effect	–.11	–0.16 to –0.06 <sup>e</sup>
Direct effect	–.05	–0.10 to –0.00 <sup>i</sup>

<sup>a</sup>eDevice: electronic device.

<sup>b</sup>Adjusted for sex, age, marital status, education attainment and family income.

<sup>c</sup>Bootstrapping with 1000 replications for bias-corrected 95% confidence interval.

<sup>d</sup>Proportion of total effect mediated: 61.2%.

<sup>e</sup> $P<.001$ .

<sup>f</sup> $P<.01$ .

<sup>g</sup>Proportion of total effect mediated: 87.0%.

<sup>h</sup>Proportion of total effect mediated: 67.8%.

<sup>i</sup> $P<.05$ .

**Table 5.** Association between eDevice<sup>a</sup> use during dinner and rules banning such use.

Family rules on eDevice use	Sometimes or often use eDevice during family dinner by oneself	Adjusted odds ratio (95% CI) <sup>b</sup>		P value
Rules banning eDevice use during family dinner				
No ban	Reference		Reference	
Partial ban <sup>c</sup>	1.02 (0.69 to 1.52)		1.33 (0.98 to 1.81)	.07
Total ban	0.49 (0.29 to 0.85)		0.41 (0.28 to 0.60)	<.001

<sup>a</sup>eDevice: electronic device.<sup>b</sup>Adjusted for sex, age, marital status, education attainment, and family income.<sup>c</sup>Partial ban: Limited time or frequency of use.

## Discussion

### Principal Findings

This study provided the first evidence that eDevice use during family dinner was associated with lower family communication and well-being in a population-representative adult sample. Male sex, younger age, and being single were associated with longer time spent on eDevice [27]. Respondents with higher SES status may use eDevice more because of longer work-related use. In our previous study, respondents with higher SES status used eDevice more frequently for health and family information seeking and sharing [24], and communicated with family members more via phone calls, instant messaging (IM), and video calls [9,10]. Higher SES status respondents had higher access to eDevice, which may also lead to greater use [24]. As eDevices become cheaper and easier to use, such SES status difference of eDevice use may reduce. The associations between eDevice use and psychological well-being were mixed and inconclusive [28]. eDevice use has facilitated interpersonal communication through different social platforms but may intervene in face-to-face interactions at social settings. Our results were consistent with large-scale surveys and systematic reviews that higher daily use of eDevice was generally not associated with poorer well-being [29-31]. Moderate daily eDevice use may benefit well-being through more efficient social interactions and greater perceived social support [32]. Moreover, we found that excessive eDevice use before sleep was associated with lower family well-being and the association was partially mediated by a perceived lower family communication quality. Studies have reported that sleep disturbance also partially mediated the relationship between eDevice use before sleep and depression symptoms in adolescents [13].

Excessive smartphone use reduces the interaction among family members and leads to lower levels of relationship satisfaction and family cohesion and induces family conflicts [17,33,34]. However, eDevice use during dinner has become a social norm [15]. We found that often or sometimes use of eDevice was very common during family time (794/2046, 38.81%) and family dinner (311/2017, 15.42%). Family gatherings and quality communication are vital for family well-being and dinner time is the most important part of family life. Our results indicate a potential risk to fundamental social needs resulting from eDevice

use during family dinner. Previous studies found that individuals who phubbed others or experienced phubbing in social settings may experience lower relationship satisfaction than those who received direct eye contact [35]. Our results showed that poor family communication quality substantially mediated the associations between eDevice use at dinner and family well-being. A perceived interruption of interaction, including verbal and nonverbal signals such as gestures and attitude, displays most common features of social exclusion, which can lead to detrimental effects on psychological social needs including belongingness, self-esteem, meaningful existence, and control [36]. In addition, negative perceived interaction quality and negative relationship satisfaction were also associated with phubbing others and may reinforce the eDevice use under the family environment [15,20]. Nevertheless, the temporal association could not be inferred in this cross-sectional study. According to the Compensatory Internet Use theory [37], people with poor family support may use eDevice to compensate for their social needs, avoid conflict, and alleviate negative emotions. The perceived lower family well-being can be supported by other psychological effects in relation to the increased use of eDevice such as feeling lonely, having poor social support, and having higher likelihood of developing depression or other mental distress [14].

Communication is the keystone of harmony and well-being of family relationship as well as the determinant of other satisfactory social connections such as friendship and partnership [3]. Our results, consistent with the literature, indicate that eDevice use at family settings may harm relationships [20]. However, we have reported that using eDevice for family communication such as family chat groups and video chat was associated with higher family communication quality and family well-being [10]. An increasing number of people are using IM for family communication. Future studies may investigate the context of eDevice use during family time. Involving and encouraging family members to use eDevice instead of excluding them from social networking could be practical for maintaining a harmonious family relationship. Seldom or sometimes use of eDevice during family time is generally not associated with lower family communication and well-being. Considering a total ban might not be feasible for every family in modern societies; instead, avoiding excessive use of eDevice would be more feasible and less likely to induce confrontations among family members. Research on community interventions

to improve the awareness of the family members to use eDevice smartly for a healthy and happy family environment is warranted.

We found that 18.39% (376/2045) of Hong Kong households limited eDevice use at family dinner and a total ban was significantly associated with less use. Regulated eDevice use during family activities could provide an interactive venue to foster social awareness and improve self-regulation. A previous trial [38] has found that a family colearning program of eDevice use behavior could improve the mutual understanding of usage behavior, improve use-limiting, appreciably changed smartphone-mediated parent-child interaction, and decreased the total smartphone usage. Only 175/2045 (8.56%) of Hong Kong families had a total ban of eDevice use during dinner. Future studies might consider adopting community-based interventions to promote and facilitate family rules for greater family health and well-being. eDevice companies also recognized that being overoccupied with notifications and attached to internet-based activities could be problematic. A “do not disturb” feature has thus been developed in some smartphones (eg, iPhone; Apple) to release people from vibrations, lights, or rings for a nondisrupted meeting, conversation, dinner, and sleep. However, evidence to guide policies on appropriate eDevice exposure is limited, and mostly designed for children and adolescents. Proper family guidance is needed for a better family communication and well-being.

This study had several limitations. First, unmeasured residual confounding cannot be excluded in any observational study.

Dispositional factors such as personality trait, including impulsivity and neuroticism, were predictors of problematic internet use and well-being. Major life events (eg, disability, divorce), family relationships, and social capital were also vital to subjective and family well-being but were not adjusted for [1,39,40]. Second, the cross-sectional design has restricted causal interpretation. Third, we used landline telephone directories for household sampling, and thus excluded families with only mobile phones. Mobile sample might include younger, higher educated, and high-income households [22]. However, the internet penetration in Hong Kong has reached 98% in 2017 [41], suggesting a very limited difference for eDevice access in this population. Recall bias in reporting the eDevice use cannot be eliminated. We did not separate eDevice use into computer, smartphone, or tablet use. The associations with well-being might not be significantly different between different eDevices used as they shared most commonly used features. Finally, perceived family communication quality and family well-being were assessed by questionnaires developed under Chinese social context. The generalizability of the findings with family communication and well-being warrants further study in other settings.

## Conclusion

eDevice use before sleep and during family dinner was associated with lower family well-being, and the association was substantially mediated by family communication quality. Our results suggest that interventions on smart use of eDevice may improve family communication and well-being.

## Acknowledgments

This work was supported by the Hong Kong Jockey Club Charities Trust.

## Authors' Contributions

MW, SC, and TL conceived the study. SZ and MW analyzed the data. SZ, NG, MW, and TL interpreted the data. SZ wrote the first draft of the manuscript. All authors critically revised and approved the final version of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary Table 1.

[DOCX File, 17 KB - [jmir\\_v22i10e20529\\_app1.docx](#)]

## References

1. Thomas PA, Liu H, Umberson D. Family Relationships and Well-Being. *Innov Aging* 2017 Nov;1(3):igx025 [FREE Full text] [doi: [10.1093/geroni/igx025](#)] [Medline: [29795792](#)]
2. Kryś K, Capaldi CA, Zelenski JM, Park J, Nader M, Kocimska-Zych A, et al. Family well-being is valued more than personal well-being: A four-country study. *Curr Psychol* 2019 Apr 27;1-12. [doi: [10.1007/s12144-019-00249-2](#)]
3. Vangelisti AL. *The Routledge Handbook of Family Communication*. Oxford, UK: Routledge; 2012.
4. Lam WWT, Fielding R, McDowell I, Johnston J, Chan S, Leung GM, et al. Perspectives on family health, happiness and harmony (3H) among Hong Kong Chinese people: a qualitative study. *Health Educ Res* 2012 Oct;27(5):767-779. [doi: [10.1093/her/cys087](#)] [Medline: [22907531](#)]
5. Frey BS. *Happiness: A Revolution in Economics*. Cambridge, MA: MIT Press; 2008.
6. Easterlin RA. Explaining happiness. *Proc Natl Acad Sci U S A* 2003 Sep 16;100(19):11176-11183. [doi: [10.1073/pnas.1633144100](#)] [Medline: [12958207](#)]

7. Shen C, Wan A, Kwok LT, Pang S, Wang X, Stewart SM, et al. A Community-Based Intervention Program to Enhance Family Communication and Family Well-being: The Learning Families Project in Hong Kong. *Front Public Health* 2017;5:257 [FREE Full text] [doi: [10.3389/fpubh.2017.00257](https://doi.org/10.3389/fpubh.2017.00257)] [Medline: [29085815](https://pubmed.ncbi.nlm.nih.gov/29085815/)]
8. Castellacci F, Tveito V. Internet use and well-being: A survey and a theoretical framework. *Research Policy* 2018 Feb;47(1):308-325. [doi: [10.1016/j.respol.2017.11.007](https://doi.org/10.1016/j.respol.2017.11.007)]
9. Shen C, Wang MP, Chu JT, Wan A, Viswanath K, Chan SSC, et al. Sharing Family Life Information Through Video Calls and Other Information and Communication Technologies and the Association With Family Well-Being: Population-Based Survey. *JMIR Ment Health* 2017 Nov 23;4(4):e57 [FREE Full text] [doi: [10.2196/mental.8139](https://doi.org/10.2196/mental.8139)] [Medline: [29170145](https://pubmed.ncbi.nlm.nih.gov/29170145/)]
10. Wang MP, Chu JTW, Viswanath K, Wan A, Lam TH, Chan SS. Using Information and Communication Technologies for Family Communication and Its Association With Family Well-Being in Hong Kong: FAMILY Project. *J Med Internet Res* 2015 Aug 24;17(8):e207 [FREE Full text] [doi: [10.2196/jmir.4722](https://doi.org/10.2196/jmir.4722)] [Medline: [26303434](https://pubmed.ncbi.nlm.nih.gov/26303434/)]
11. Chóliz M. Mobile phone addiction: a point of issue. *Addiction* 2010 Feb;105(2):373-374. [doi: [10.1111/j.1360-0443.2009.02854.x](https://doi.org/10.1111/j.1360-0443.2009.02854.x)] [Medline: [20078493](https://pubmed.ncbi.nlm.nih.gov/20078493/)]
12. Kenney EL, Gortmaker SL. United States Adolescents' Television, Computer, Videogame, Smartphone, and Tablet Use: Associations with Sugary Drinks, Sleep, Physical Activity, and Obesity. *J Pediatr* 2017 Mar;182:144-149. [doi: [10.1016/j.jpeds.2016.11.015](https://doi.org/10.1016/j.jpeds.2016.11.015)] [Medline: [27988020](https://pubmed.ncbi.nlm.nih.gov/27988020/)]
13. Dewi RK, Efendi F, Has EMM, Gunawan J. Adolescents' smartphone use at night, sleep disturbance and depressive symptoms. *Int J Adolesc Med Health* 2018 Nov 17:e1 (forthcoming). [doi: [10.1515/ijamh-2018-0095](https://doi.org/10.1515/ijamh-2018-0095)] [Medline: [30447141](https://pubmed.ncbi.nlm.nih.gov/30447141/)]
14. Horwood S, Anglim J. Problematic smartphone usage and subjective and psychological well-being. *Comput Hum Behav* 2019 Aug;97:44-50. [doi: [10.1016/j.chb.2019.02.028](https://doi.org/10.1016/j.chb.2019.02.028)]
15. Chotpitayasunondh V, Douglas KM. How “phubbing” becomes the norm: The antecedents and consequences of snubbing via smartphone. *Comput Hum Behav* 2016 Oct;63:9-18. [doi: [10.1016/j.chb.2016.05.018](https://doi.org/10.1016/j.chb.2016.05.018)]
16. Rotondi V, Stanca L, Tomasuolo M. Connecting Alone: Smartphone Use, Quality of Social Interactions and Well-Being. *SSRN J* 2017;63:17-26. [doi: [10.2139/ssrn.2893027](https://doi.org/10.2139/ssrn.2893027)]
17. Lapiere MA, Lewis MN. Should it stay or should it go now? Smartphones and relational health. *Psychol Pop Media Cult* 2018 Jul;7(3):384-398. [doi: [10.1037/ppm0000119](https://doi.org/10.1037/ppm0000119)]
18. Pew Research Center. Numbers, Facts, and Trends Shaping Your World. Washington, DC: Pew Research Center; 2018. URL: <https://www.loc.gov/item/lcwaN0002821/> [accessed 2020-09-26]
19. Olson DH, Russell CS, Sprenkle DH. Circumplex model of marital and family systems: VI. Theoretical update. *Fam Process* 1983 Mar;22(1):69-83. [doi: [10.1111/j.1545-5300.1983.00069.x](https://doi.org/10.1111/j.1545-5300.1983.00069.x)] [Medline: [6840263](https://pubmed.ncbi.nlm.nih.gov/6840263/)]
20. Kadylak T. An investigation of perceived family phubbing expectancy violations and well-being among U.S. older adults. *Mob Media Commun* 2019 Sep 28;8(2):247-267. [doi: [10.1177/2050157919872238](https://doi.org/10.1177/2050157919872238)]
21. Hawi NS, Samaha M. Relationships among smartphone addiction, anxiety, and family relations. *Behav Inf Technol* 2017 Jun 07;36(10):1046-1052. [doi: [10.1080/0144929x.2017.1336254](https://doi.org/10.1080/0144929x.2017.1336254)]
22. Guo N, Wang MP, Luk TT, Ho SY, Fong DYT, Chan SS, et al. The association of problematic smartphone use with family well-being mediated by family communication in Chinese adults: A population-based study. *J Behav Addict* 2019 Sep 01;8(3):412-419 [FREE Full text] [doi: [10.1556/2006.8.2019.39](https://doi.org/10.1556/2006.8.2019.39)] [Medline: [31394912](https://pubmed.ncbi.nlm.nih.gov/31394912/)]
23. Wang MP, Viswanath K, Lam TH, Wang X, Chan SS. Social determinants of health information seeking among Chinese adults in Hong Kong. *PLoS One* 2013;8(8):e73049 [FREE Full text] [doi: [10.1371/journal.pone.0073049](https://doi.org/10.1371/journal.pone.0073049)] [Medline: [24009729](https://pubmed.ncbi.nlm.nih.gov/24009729/)]
24. Wang MP, Wang X, Viswanath K, Wan A, Lam TH, Chan SS. Digital inequalities of family life information seeking and family well-being among Chinese adults in Hong Kong: a population survey. *J Med Internet Res* 2014 Oct 03;16(10):e227 [FREE Full text] [doi: [10.2196/jmir.3386](https://doi.org/10.2196/jmir.3386)] [Medline: [25281831](https://pubmed.ncbi.nlm.nih.gov/25281831/)]
25. Shen C, Wang MP, Ho HCY, Wan A, Stewart SM, Viswanath K, et al. Test-retest reliability and validity of a single-item Self-reported Family Happiness Scale in Hong Kong Chinese: findings from Hong Kong Jockey Club FAMILY Project. *Qual Life Res* 2019 Feb;28(2):535-543. [doi: [10.1007/s11136-018-2019-9](https://doi.org/10.1007/s11136-018-2019-9)] [Medline: [30284182](https://pubmed.ncbi.nlm.nih.gov/30284182/)]
26. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986 Dec;51(6):1173-1182. [Medline: [3806354](https://pubmed.ncbi.nlm.nih.gov/3806354/)]
27. Christensen MA, Bettencourt L, Kaye L, Moturu ST, Nguyen KT, Olgin JE, et al. Direct Measurements of Smartphone Screen-Time: Relationships with Demographics and Sleep. *PLoS One* 2016 Nov 9;11(11):e0165331 [FREE Full text] [doi: [10.1371/journal.pone.0165331](https://doi.org/10.1371/journal.pone.0165331)] [Medline: [27829040](https://pubmed.ncbi.nlm.nih.gov/27829040/)]
28. Twenge JM, Martin GN, Campbell WK. Decreases in psychological well-being among American adolescents after 2012 and links to screen time during the rise of smartphone technology. *Emotion* 2018 Sep;18(6):765-780. [doi: [10.1037/emo0000403](https://doi.org/10.1037/emo0000403)] [Medline: [29355336](https://pubmed.ncbi.nlm.nih.gov/29355336/)]
29. Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review* 2014 Jun;41:27-36. [doi: [10.1016/j.childyouth.2014.03.001](https://doi.org/10.1016/j.childyouth.2014.03.001)]
30. Shakya HB, Christakis NA. Association of Facebook Use With Compromised Well-Being: A Longitudinal Study. *Am J Epidemiol* 2017 Feb 01;185(3):203-211. [doi: [10.1093/aje/kww189](https://doi.org/10.1093/aje/kww189)] [Medline: [28093386](https://pubmed.ncbi.nlm.nih.gov/28093386/)]



31. Przybylski AK, Weinstein N. A Large-Scale Test of the Goldilocks Hypothesis. *Psychol Sci* 2017 Feb 13;28(2):204-215. [doi: [10.1177/0956797616678438](https://doi.org/10.1177/0956797616678438)] [Medline: [28085574](https://pubmed.ncbi.nlm.nih.gov/28085574/)]
32. Stiglic N, Viner RM. Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews. *BMJ Open* 2019 Jan 03;9(1):e023191. [doi: [10.1136/bmjopen-2018-023191](https://doi.org/10.1136/bmjopen-2018-023191)] [Medline: [30606703](https://pubmed.ncbi.nlm.nih.gov/30606703/)]
33. Carvalho J, Francisco R, Relvas AP. Family functioning and information and communication technologies: How do they relate? A literature review. *Comput Hum Behav* 2015 Apr;45:99-108. [doi: [10.1016/j.chb.2014.11.037](https://doi.org/10.1016/j.chb.2014.11.037)]
34. Lenhart A, Rainie L, Lewis O. Teenage Life Online: The Rise of the Instant-Message Generation and the Internet's Impact on Friendships and Family Relationships. Washington, DC: Pew Research Center; Jun 20, 2001:1-46.
35. Wirth JH, Sacco DF, Hugenberg K, Williams KD. Eye gaze as relational evaluation: averted eye gaze leads to feelings of ostracism and relational devaluation. *Pers Soc Psychol Bull* 2010 Jul;36(7):869-882. [doi: [10.1177/0146167210370032](https://doi.org/10.1177/0146167210370032)] [Medline: [20505162](https://pubmed.ncbi.nlm.nih.gov/20505162/)]
36. Gerber J, Wheeler L. On Being Rejected: A Meta-Analysis of Experimental Research on Rejection. *Perspect Psychol Sci* 2009 Sep;4(5):468-488. [doi: [10.1111/j.1745-6924.2009.01158.x](https://doi.org/10.1111/j.1745-6924.2009.01158.x)] [Medline: [26162220](https://pubmed.ncbi.nlm.nih.gov/26162220/)]
37. Kardefelt-Winther D. A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use. *Comput Hum Behav* 2014 Feb;31:351-354. [doi: [10.1016/j.chb.2013.10.059](https://doi.org/10.1016/j.chb.2013.10.059)]
38. Ko M, Choi S, Yang S, Lee J, Lee U. FamiLync: facilitating participatory parental mediation of adolescents' smartphone use. New York: ACM Press; 2015 Presented at: UbiComp 2015: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (2015); September 2015; Osaka, Japan p. 867-878 URL: <https://dl.acm.org/doi/10.1145/2750858.2804283> [doi: [10.1145/2750858.2804283](https://doi.org/10.1145/2750858.2804283)]
39. Cleland C, Kearns A, Tannahill C, Ellaway A. The impact of life events on adult physical and mental health and well-being: longitudinal analysis using the GoWell health and well-being survey. *BMC Res Notes* 2016 Oct 18;9(1):470 [FREE Full text] [doi: [10.1186/s13104-016-2278-x](https://doi.org/10.1186/s13104-016-2278-x)] [Medline: [27760568](https://pubmed.ncbi.nlm.nih.gov/27760568/)]
40. Lee S, Chung JE, Park N. Network Environments and Well-Being: An Examination of Personal Network Structure, Social Capital, and Perceived Social Support. *Health Commun* 2018 Jan;33(1):22-31. [doi: [10.1080/10410236.2016.1242032](https://doi.org/10.1080/10410236.2016.1242032)] [Medline: [27858469](https://pubmed.ncbi.nlm.nih.gov/27858469/)]
41. Census Statistics Department. Hong Kong as an Information Society. Hong Kong: Census Statistics Department (HK); 2017. URL: <https://www.statistics.gov.hk/pub/B11100062017AN17B0100.pdf> [accessed 2020-03-15]

## Abbreviations

**eDevice:** electronic device

**SES:** socioeconomic

*Edited by G Eysenbach; submitted 21.05.20; peer-reviewed by S Durmaz, G Costa; comments to author 02.07.20; revised version received 10.07.20; accepted 26.07.20; published 14.10.20.*

*Please cite as:*

Zhao SZ, Guo N, Wang MP, Fong DYT, Lai AYK, Chan SSC, Lam TH, Ho DSY

*The Association Between Electronic Device Use During Family Time and Family Well-Being: Population-Based Cross-Sectional Study*

*J Med Internet Res* 2020;22(10):e20529

URL: <http://www.jmir.org/2020/10/e20529/>

doi: [10.2196/20529](https://doi.org/10.2196/20529)

PMID: [33052120](https://pubmed.ncbi.nlm.nih.gov/33052120/)

©Sheng Zhi Zhao, Ningyuan Guo, Man Ping Wang, Daniel Yee Tak Fong, Agnes Yuen Kwan Lai, Sophia Siu-Chee Chan, Tai Hing Lam, Daniel Sai Yin Ho. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org/>), 14.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Patterns of Use and Key Predictors for the Use of Wearable Health Care Devices by US Adults: Insights from a National Survey

Ranganathan Chandrasekaran<sup>1</sup>, PhD; Vipanchi Katthula<sup>1</sup>, MSc; Evangelos Moustakas<sup>2</sup>, PhD

<sup>1</sup>Department of Information & Decision Sciences, University of Illinois at Chicago, Chicago, IL, United States

<sup>2</sup>Middlesex University Dubai, Dubai, United Arab Emirates

**Corresponding Author:**

Ranganathan Chandrasekaran, PhD

Department of Information & Decision Sciences

University of Illinois at Chicago

601 S Morgan St

Chicago, IL, 60607

United States

Phone: 1 3129962847

Email: [ranga@uic.edu](mailto:ranga@uic.edu)

## Abstract

**Background:** Despite the growing popularity of wearable health care devices (from fitness trackers such as Fitbit to smartwatches such as Apple Watch and more sophisticated devices that can collect information on metrics such as blood pressure, glucose levels, and oxygen levels), we have a limited understanding about the actual use and key factors affecting the use of these devices by US adults.

**Objective:** The main objective of this study was to examine the use of wearable health care devices and the key predictors of wearable use by US adults.

**Methods:** Using a national survey of 4551 respondents, we examined the usage patterns of wearable health care devices (use of wearables, frequency of their use, and willingness to share health data from a wearable with a provider) and a set of predictors that pertain to personal demographics (age, gender, race, education, marital status, and household income), individual health (general health, presence of chronic conditions, weight perceptions, frequency of provider visits, and attitude towards exercise), and technology self-efficacy using logistic regression analysis.

**Results:** About 30% (1266/4551) of US adults use wearable health care devices. Among the users, nearly half (47.33%) use the devices every day, with a majority (82.38% weighted) willing to share the health data from wearables with their care providers. Women (16.25%), White individuals (19.74%), adults aged 18-50 years (19.52%), those with some level of college education or college graduates (25.60%), and those with annual household incomes greater than US \$75,000 (17.66%) were most likely to report using wearable health care devices. We found that the use of wearables declines with age: Adults aged >50 years were less likely to use wearables compared to those aged 18-34 years (odds ratios [OR] 0.46-0.57). Women (OR 1.26, 95% CI 0.96-1.65), White individuals (OR 1.65, 95% CI 0.97-2.79), college graduates (OR 1.05, 95% CI 0.31-3.51), and those with annual household incomes greater than US \$75,000 (OR 2.6, 95% CI 1.39-4.86) were more likely to use wearables. US adults who reported feeling healthier (OR 1.17, 95% CI 0.98-1.39), were overweight (OR 1.16, 95% CI 1.06-1.27), enjoyed exercise (OR 1.23, 95% CI 1.06-1.43), and reported higher levels of technology self-efficacy (OR 1.33, 95% CI 1.21-1.46) were more likely to adopt and use wearables for tracking or monitoring their health.

**Conclusions:** The potential of wearable health care devices is under-realized, with less than one-third of US adults actively using these devices. With only younger, healthier, wealthier, more educated, technoliterate adults using wearables, other groups have been left behind. More concentrated efforts by clinicians, device makers, and health care policy makers are needed to bridge this divide and improve the use of wearable devices among larger sections of American society.

(*J Med Internet Res* 2020;22(10):e22443) doi:[10.2196/22443](https://doi.org/10.2196/22443)

**KEYWORDS**

wearable healthcare devices; mobile health; HINTS; health technology adoption and use; smart wearables

## Introduction

### Background and Motivation

Advances in wireless sensors and digital technologies have led to a proliferation of wearable health care devices with which users can examine, monitor, and track their physiological conditions. Wearable health care devices are autonomous, noninvasive, wearable equipment with embedded sensors to collect a variety of physiological health information [1]. These devices range from the popular fitness trackers (eg, Fitbit, AppleWatch, Samsung, Galaxy Fit) that collect data on physical activities such as number of steps taken, calories burned, sleep duration, and heart rate to more sophisticated devices that can collect information on blood pressure, glucose levels, and oxygen levels. The collected health data can be transmitted to smartphones or other computer-aided applications to store and analyze to provide appropriate health interventions. Fueled by increased popularity, the use of wearables has significantly increased in recent years. According to estimates, the market for wearable health care devices in 2018 was US \$24.57 billion and was slated to grow 24.7% annually to US \$139.35 billion by 2026 [2].

Wearable health care devices offer several potential benefits to users. First, they offer a convenient way to monitor, store, and share health information in real-time. Second, wearables provide feedback to users to make appropriate changes to their daily routines or behavior [3,4]. Third, wearables can facilitate remote patient monitoring and provide proactive and faster data access to physicians, resulting in improved health outcomes [4,5] and reduced number of physician visits [6]. Fourth, these devices can be particularly useful for patients with chronic conditions [7], patients with cardiovascular risks [8], and elderly populations [9]. Therefore, the use of wearable devices has the potential to significantly improve health care delivery and reduce costs.

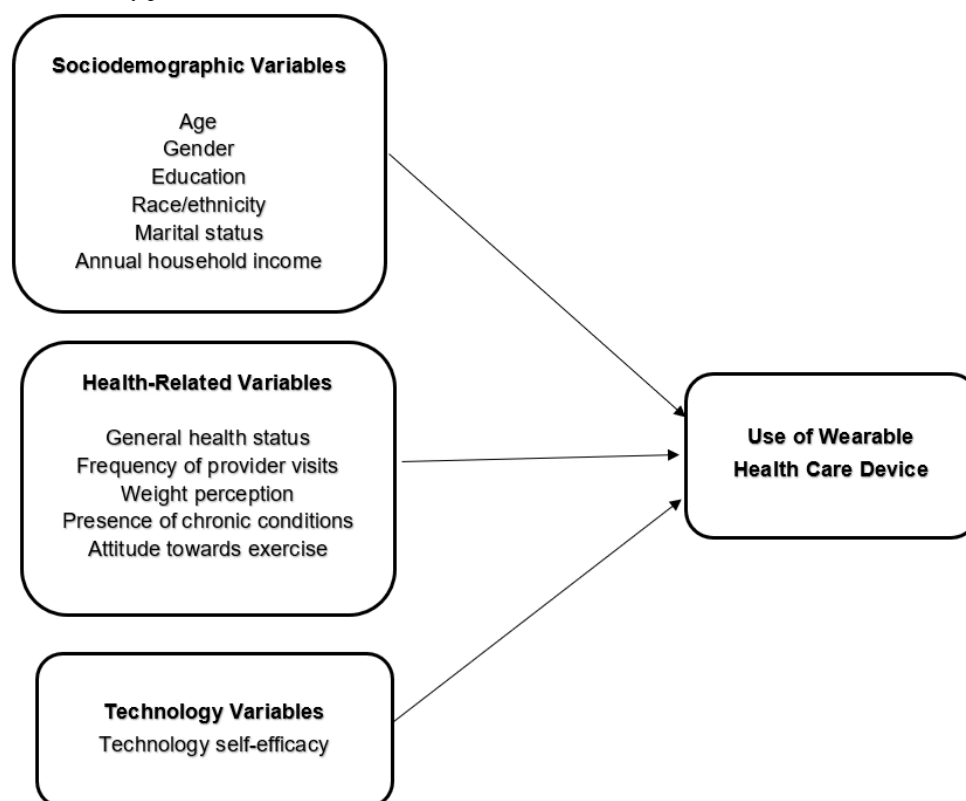
Despite potential benefits, significant challenges remain to the widespread adoption and use of wearable health care devices [10-12]. The ability of these devices to track, store, and transmit patients' health information raises questions about data security and privacy [13-15]. In addition, the design, accuracy, and

reliability of wearables have also been a major concern [16,17]. Concerns have also been raised about the accuracy of data gathered by wearables in people of color [18]. Technology acceptance of new wearable devices remains another significant barrier [19]. Despite the forecasted growth, use of these devices has reportedly slowed [20]. Market studies have pointed out a gradual decline in the use of these devices as well as abandonment within a few months of purchase [21,22].

Extant research on the use of wearables has found these devices to increase physical activity, increase energy expenditure, and reduce sedentary behavior [23]. Wearables offer an easy way to obtain large amounts of real-time health data that can be useful for clinicians [24]. Health data obtained from wearables, when combined with sophisticated machine learning algorithms, have helped develop predictive models that can greatly improve health delivery [25-27]. If effectively used, wearables can greatly help in the management of several conditions including rheumatic and musculoskeletal diseases [28,29], chronic pain management [30], and cardiovascular problems [31]. To effectively realize the potential benefits of wearable health care devices, a solid understanding of the factors related to their adoption and use is warranted. Most extant studies have largely examined the intention to adopt wearable devices [32-36] rather than the actual usage. Further, these studies have focused on a limited set of antecedents driven by the technology adoption lens. Hence, we know very little about US adults' use of wearables and key influential factors affecting usage. Addressing this gap, we build on an emergent stream of studies to examine the key factors affecting the use of wearable health care devices by US adults.

### Objectives

The main objective of this study was to examine the use of wearable health care devices and the key predictors of wearable use by US adults. We examine predictors related to individual health [37], technology self-efficacy [36,38], personal demographics [32], and attitudes towards fitness or exercise [35] as well as their associations with the use of wearable health care devices. Our research model, with all the predictor variables, is shown in Figure 1.

**Figure 1.** Research model with key predictors of wearable health care device use.

## Methods

### Data

The dataset for this study comes from the National Cancer Institute's Health Information National Trends Survey (HINTS)-5, Cycle 3, with data collected from January 2019 to April 2019 through self-administered mailed questionnaires and a web-based pilot. HINTS is a nationally representative survey that includes US adults  $\geq 18$  years of age in civilian, noninstitutionalized settings. This survey collects data on US adults' need, access, and use of information related to health and health care and the related behaviors, perceptions, and knowledge. It uses a stratified sampling method defined by areas with high concentrations of minorities and areas with low concentrations of minorities. Survey invitees for both mailed questionnaires and the web pilot involved an initial mailing of the questionnaire, followed by a reminder postcard and up to two additional mailings of the questionnaire as needed for nonrespondents. More details on the survey and survey methodology can be found elsewhere [39].

Since HINTS used probability sampling to improve representation of specific groups, our analysis applied weights to calculate US population estimates and standard errors. These sampling weights were needed to ensure unbiased estimations. Sampling weights are inverse probabilities of selection but are modified based on census or population information and further adjusted for nonresponses, so that the weighted totals of poststratification variables match the population totals. Weight adjustment accounted for nonresponse and known population totals based on the 2017 American Community Survey of the US Census Bureau on age, gender, education, marital status,

race, ethnicity, and census region. We used the jackknife approach to compute replication weights [40]. Jackknife is a popular resampling approach that creates many replicate samples taken from the original sample. Each replicate sample provides an estimate of the variable of interest, and the variability among the estimates from multiple replicate samples are used to estimate the variance of the variable of interest [41].

For data analysis, we included all respondents who responded to the question about their use of an electronic health care device to track or monitor their health or physical activity in the past 12 months. Of the 5380 overall respondents, 4551 had indicated if they used (yes/no) a wearable health care device and were included in our analysis. STATA 16.1 software was used to perform the statistical analyses. We compared the demographics of all respondents with our sample that had answered the question about the use of wearable devices and did not detect any significant differences [42].

### Variables

The main variable of interest was the use of a wearable health care device. This was assessed as a binary variable (yes/no) that asked respondents to indicate their use of an electronic wearable device to monitor or track health or activity in the past 12 months. In addition, we also explored the responses about frequency of wearable health care device use in the past month (daily, almost daily, 1-2 times a week,  $<1$  time a week, never used) and the respondents' willingness to share the data from wearables with a provider (yes/no). Data from several survey items were included to capture survey respondents' self-reported characteristics: sociodemographic (age, gender, race/ethnicity, education, marital status, annual household income), health-related (general health, presence of any chronic

conditions, frequency of provider visits, self-perception about weight), and technology-related (technology self-efficacy) variables. General health was assessed using a question for which respondents rated their health on a Likert scale (rated as 1-5) as being poor, fair, good, very good, or excellent. Presence of chronic conditions was coded as a binary variable (0/1) if the respondents indicated they had at least one of the following conditions: diabetes or high blood sugar; high blood pressure or hypertension; heart condition such as heart attack, angina, or congestive heart failure; or a chronic lung disease such as asthma, emphysema, or chronic bronchitis. Frequency of provider visits was captured using a Likert scale based on the number of times the respondents visited a health provider in the past 12 months. Self-perception about weight was assessed on a Likert scale (rated as 1-5) as underweight, slightly underweight, just about the right weight, slightly overweight, or overweight.

As people interested in physical fitness and exercise are likely to be drawn towards a wearable health care device, another study variable included attitude towards exercise (to what extent do you enjoy exercising?). This was captured using a Likert scale (rated as 1-4), with the following options: not at all, a little, some, or a lot. Technology self-efficacy was captured using an additive score (ranging from 0 to 6 points) that was captured using questions that asked participants if they used a computer, smartphone, or electronic means for performing 6 tasks: (1) look for health information, (2) purchase medicines or vitamins online, (3) communicate with a provider using email or the internet, (4) track health care charges, (5) look up medical test results, and (6) make appointments with a provider.

## Analyses

We first conducted a descriptive analysis on our data sample. To assess the relationship between the use of wearable health care devices and sociodemographic or health-related categorical predictor variables, crosstab tables were generated and tested using the Wald chi-square test. All results were weighted to give US population level inferences using a standard weighting approach that was specifically developed for the HINTS dataset.

Given that our main variable of interest (ie, use of a wearable health care device) was a binary variable and our predictor variables were a mix of categorical and continuous variables, logistic regression analysis was chosen to assess the influence of sociodemographic variables, health-related predictors, and technology self-efficacy on the use of a wearable health care device. We computed adjusted odds ratios (ORs) and 95% CI estimates for the predictors and used a cutoff of  $P < .05$  to determine the statistical significance of all our analyses. In accordance with the complex survey design, weights with jackknife replications were applied in our analyses.

## Results

All percentages reported in this section are the weighted values. Of the 4551 respondents who responded to the question on wearable devices, 1266 (29.95%) indicated using the device in the past 12 months. Of the adopters, 47.33% used it every day, and an additional 24.85% indicated using the device “almost every day.” Of the adopters, a majority (82.38%) reported that they were willing to share the data from wearable devices with their health care provider (Figure 2).

**Figure 2.** Patterns of wearable health device use by US adults.

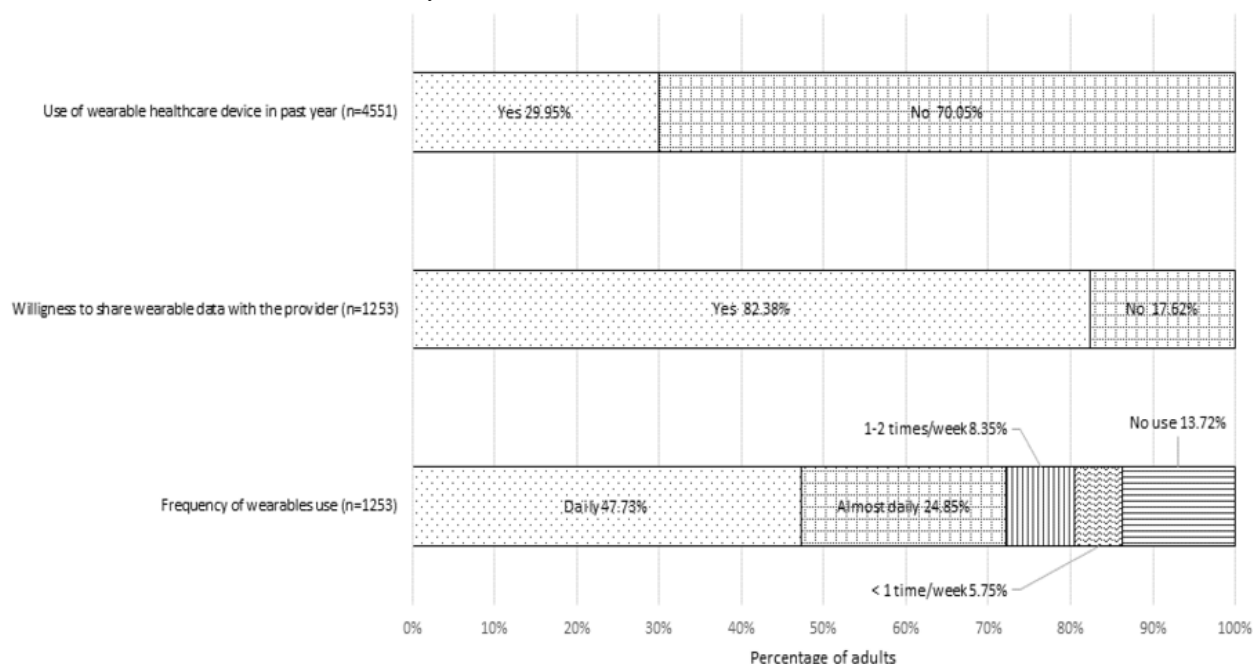


Table 1 presents the demographic characteristics of the respondents. Significant differences were observed between users and nonusers of wearable health care devices across different age groups, gender, education levels, and household income. Women (16.41%), White individuals (19.74%), adults

aged 18-50 years (19.52%), those with some level of college education or who were college graduates (25.60%), and those with annual household incomes greater than US \$75,000 (17.66%) were most likely to report using wearable health care devices. There were no significant differences based on marital

status or race/ethnicity. We did not detect any significant differences in the willingness to share data from wearable devices with a provider or frequency of use of wearable devices across any of the demographic variables. Based on these results,

we performed logistic regression analysis to examine how the use or nonuse of wearable devices was associated with demographic variables, health-related variables, technology self-efficacy, and attitude towards exercise.

**Table 1.** Weighted results of sociodemographic characteristics of US respondents, by use of wearable health care devices and willingness to share wearable data with a health care provider.<sup>a</sup>

Respondent characteristics	Use of wearable health care device in past 12 months (n=4551)				Willingness to share wearable data with health care provider (n=1253)			
	Total, %	Yes, %	No, %	P value <sup>b</sup>	Total, %	Yes, %	No, %	P value
Total sample, n (%)	N/A <sup>c</sup>	1266 (27.82)	3285 (72.18)	N/A	N/A	1013 (82.38)	240 (17.62)	N/A
<b>Age (years)</b>				<.001				.22
18-34	26.78	10.20	16.58		34.18	30.01	4.17	
35-49	26.50	9.32	17.18		30.85	24.71	6.14	
50-64	30.77	7.93	22.84		25.77	20.39	5.37	
65-74	10.47	2.00	8.47		6.65	5.12	1.53	
≥75	5.48	0.80	4.68		2.55	2.11	0.44	
<b>Gender</b>								
Male	49.34	13.54	35.80	.04	44.98	36.57	8.41	.55
Female	50.66	16.41	34.25		55.02	45.81	9.21	
<b>Education</b>								
Less than high school	4.92	0.78	4.14	<.001	2.63	2.24	0.39	.20
High school	21.73	3.49	18.24		11.48	8.10	3.38	
Some college	41.29	12.75	28.54		42.49	35.77	6.72	
At least a college graduate	32.06	12.85	19.21		43.40	36.40	7.00	
<b>Marital status</b>								
Married	69.27	20.59	48.68	.85	68.39	54.78	13.61	.03
Other	30.73	9.32	21.41		31.61	27.86	3.75	
<b>Race</b>								
Non-Hispanic Asian	5.84	1.76	4.09	.65	5.83	4.18	1.65	.39
African American	10.58	2.68	7.90		8.94	7.65	1.29	
Hispanic	16.69	5.29	11.40		17.15	13.37	3.78	
White	63.63	19.74	43.89		64.91	55.15	9.76	
Other	3.26	0.95	2.31		3.17	2.70	0.47	
<b>Household income (US \$)</b>								
<20,000	16.13	2.37	13.76	<.001	7.71	6.66	1.05	.73
20,000 to <35,000	9.61	1.44	8.17		4.83	3.89	0.94	
35,000 to <50,000	13.18	3.89	9.29		12.90	9.90	3.00	
50,000 to <75,000	18.14	4.92	13.22		16.25	13.11	3.14	
≥75,000	42.94	17.66	25.27		58.31	49.14	9.17	

<sup>a</sup>The frequency of wearable usage was omitted because no significant differences were observed based on sociodemographic characteristics.

<sup>b</sup>Wald chi-square test.

<sup>c</sup>N/A: not applicable.

The logistic regression results revealed significant associations between demographic characteristics and the use of wearable health care devices (Table 2). Individuals aged 50-64 years (OR

0.57), 65-74 years (OR 0.46), or ≥75 years (OR 0.47) were less likely to use wearable health care devices than those aged 18-34 years. Women were 1.26 times more likely to use wearable



health care devices than men (OR 1.26, 95% CI 0.96-1.65). As compared to non-Hispanic Asian individuals, White individuals (OR 1.65, 95% CI 0.97-2.79) were more likely to use wearables. Individuals with higher levels of education (ie, some college education; OR 1.06, 95% CI 0.30-3.69) or who were at least

college graduates (OR 1.05, 95% CI 0.31-3.51) exhibited greater likelihood to use wearables. Individuals whose annual household income was at least US \$75,000 were 2.6 times more likely to use wearables (OR 2.6, 95% CI 1.39-4.86) than those with incomes less than US \$20,000.

**Table 2.** The respondent characteristics that had the greatest influence in predicting the use of wearable health care devices.

Predictors	Prediction of the use of a wearable health care device in the last 12 months		
	Adjusted odds ratio <sup>a</sup>	95% CI	P value
<b>Age (years)<sup>b</sup></b>			
35-49	0.79	0.54-1.16	.22
50-64	0.57	0.37-0.87	<.001
65-74	0.46	0.28-0.76	<.001
≥75	0.47	0.24-0.89	.01
Gender <sup>c</sup> : Female	1.26	0.96-1.65	.01
<b>Education<sup>d</sup></b>			
High school graduate	0.48	0.14-1.62	.14
Some college	1.06	0.30-3.69	.04
At least a college graduate	1.04	0.31-3.51	.05
<b>Race/ethnicity<sup>e</sup></b>			
African American	1.48	0.89-3.81	.09
Hispanic	1.24	0.88-3.06	.12
White	1.65	0.97-2.79	.05
Other	1.29	0.42-4.01	.65
Marital status <sup>f</sup> : married	1.02	0.68-1.54	.91
<b>Household income (\$ US)<sup>g</sup></b>			
20,000 to <35,000	0.80	0.41-1.57	.51
35,000 to <50,000	1.82	0.84-3.97	.12
50,000 to <75,000	1.49	0.82-2.68	.18
≥75,000	2.60	1.39-4.86	<.001
General health	1.17	0.98-1.39	.01
Frequency of provider visits	0.97	0.89-1.05	.38
Weight perception	1.16	1.06-1.27	<.001
Presence of chronic conditions	0.91	0.63-1.31	.61
Attitude towards exercise	1.23	1.06-1.43	<.001
Technology self-efficacy	1.33	1.21-1.46	<.001

<sup>a</sup>Adjusted odds ratios and 95% CIs generated from multivariate logistic regression. Model accounts for replicate weights.

<sup>b</sup>Reference: 18-34 years.

<sup>c</sup>Reference: male.

<sup>d</sup>Reference: less than high school.

<sup>e</sup>Reference: Asian.

<sup>f</sup>Reference: nonmarried.

<sup>g</sup>Reference: <20,000.

Individuals who felt their general health was better (OR 1.17, 95% CI 0.98-1.39) and those who perceived themselves to be overweight (OR 1.16, 95% CI 1.06-1.27) were more likely to

adopt wearable devices. No significant differences were found based on the frequency of provider visits and for those with ≥1 chronic condition. Those who enjoyed physical activity and

exercise were more likely to use health care wearables (OR 1.23, 95% CI 1.06-1.43), and individuals with higher levels of technology self-efficacy were more likely to adopt and use wearables to track or monitor their health (OR 1.33, 95% CI 1.21-1.46).

The logistic regression analysis to determine the associations between the willingness to share wearable device data with providers and all the other predictors found only two variables — race and marital status — to be significant. White adults were 3 times more likely (OR 2.87, 95% CI 1.18-7.01), and married individuals were less likely (OR 0.46, 95% CI 0.22-0.96) to share their wearable data with providers. Our examination of associations between frequency of wearable usage and the predictors did not yield any significant insights.

## Discussion

This study found wearable health care devices to be used by roughly 30% of US adults. Women, individuals with higher levels of household income, and those with higher levels of education were more likely to be wearable users. The tendency to use wearable devices seems to decline with age. US adults who consider themselves to be healthier are likely to use wearables. Other studies have also reported that healthier individuals have greater intention to adopt wearables [35]. Although individuals with chronic conditions or a higher number of provider visits have greater potential to benefit from wearables, this study did not find them to be actively using these devices. Individuals who enjoy exercise and fitness and those who are more comfortable using electronic devices exhibit a greater propensity to use wearable health care devices.

Our results indicate that only younger, healthier, wealthier, more educated, and technoliterate adults are more likely to use wearables. Younger adults are likely to be attracted to the design and aesthetic appeal of the wearable device [43], consider it “cool” [44], use it to signal their fitness intentions to peers, and also like the gamification elements such as rewards, points, and recognition of achievements awarded by the wearable devices and associated mobile apps [45-47]. Elderly adults could find wearables to be more intrusive and less comfortable [34]. Elderly adults may also not appreciate the design elements that are more geared towards younger users and would often need health professionals’ regular support and feedback to sustain the use of wearable devices [48]. Our findings imply that the design and gamification techniques in wearables that are primarily appealing to younger adults need to be customized to attract other strata of users [47]. Wearable device makers can enhance the marketability of their devices and broaden the reach of their products by coming up with devices tailored to different age groups. Our findings regarding wealthier and more educated adults to be active users of wearable technology is similar to those from other studies [37]. Educated adults are likely to be more familiar with the value they can derive from wearables. Also, given the high costs of some of the wearable devices in the market, the affordability might be better for high-income individuals [35].

Our results pertaining to the use of wearables by healthier or more health-conscious users are consistent with studies that

point to the health-related motivations for using wearables such as to monitor physical activities, improve fitness, and sustain general health [49]. Individuals who set health-related goals such as weight loss and increased physical activity are likely to find wearables to be useful to monitor their progress in achieving those goals. Studies have found individuals with wellness-oriented lifestyles to engage more in preventive health behaviors like exercising regularly, and these individuals are more likely to use wearables [43]. Among other health variables, we did not find any significant association between chronic conditions and use of wearable devices. Few studies examining the use of wearable health care devices among patients with chronic conditions have reported very limited impact of these devices on disease outcomes [4]. Perhaps, such lack of tangible value for specific health conditions might discourage adults from using these devices for chronic care management purposes.

Our findings regarding the importance of technology self-efficacy validate findings from several prior studies that have examined behavioral intentions to adopt wearables from a technology acceptance perspective [35,36,50]. Individuals’ familiarity and experience with technologies such as smartphones, tablets, and other digital devices for health-related purposes would likely make them more open to using wearable health care devices. Higher technology self-efficacy can help individuals foster more positive attitudes towards smart wearables that can further augment use of these devices.

This study notes that wearable devices have not yet found widespread use among those groups that can benefit the most from them. Individuals with poor health, with chronic conditions, and who are aged can greatly benefit from tracking their physical activity and by letting their providers access the health data captured through wearables. Although our findings are consistent with other studies that have found fewer adopters of wearable devices among the senior population [51-53], they indicate a critical need to identify key barriers for low adoption of wearables among elderly US adults and work towards addressing those barriers. Factors like anxiety towards new technologies and resistance to change could impede adoption and use of wearables among elderly adults [54]. Our results have important implications for health care providers and wearable device manufacturers to educate individuals on the potential benefits of using wearables to improve health outcomes. Use cases and testimonials of wearable use by patients and elderly need to be actively publicized and promoted.

Our findings also indicate a digital divide in the US that only those with higher levels of education, household income, and technology proficiency are actively using these devices [55]. To increase patient use and engagement with digital technologies such as wearable devices, narrowing the digital divide becomes increasingly important. Making digital devices more affordable and easier to use can greatly aid in widening access to wearables and effectively using them to self-track and monitor health. This study also points to the importance of digital skills development to enable US adults to utilize wearables more effectively.

## Limitations

Little is known about the adoption and use of wearable health care devices and the predictors of adoption and use among US

adults. Our work includes a broad US population and moves beyond reported intentions to use wearables to actual use of these devices. Future work should attempt to explore more predictors of the frequency of wearable usage and the actual sharing of wearable data with providers. Our study has many limitations. Although the HINTS survey used a national sample and involved stratified selection to improve the responses of subgroups, the data had responses from both mailed questionnaires and a web-based pilot, which included an incentive bonus. Web pilot responses are likely to be from adults who are more comfortable using technology and the internet, and this could possibly introduce bias. Moreover, the

self-perception questions are subjective and could be interpreted differently by participants.

## Conclusions

The potential of wearable health care devices is under-realized, with less than one-third of US adults actively using these devices. With only younger, healthier, wealthier, more educated, technoliterate adults using wearables, other groups have been left behind. To effectively capitalize on the growing popularity of wearable health care devices to improve health care delivery and outcomes, more concentrated efforts by clinicians, device makers, and health care leaders are needed.

## Conflicts of Interest

None declared.

## References

1. Fotiadis D, Glaros C, Likas A. Wearable Medical Devices. In: Akay M, editor. Wiley Encyclopedia of Biomedical Engineering. Hoboken, NJ: John Wiley & Sons, Inc; Apr 14, 2006.
2. Wearable Medical Devices Market Size, Share and Industry Analysis by Product (Diagnostic & Patient Monitoring, Therapeutics), by Application (Remote Patient Monitoring and Home Healthcare, Sports and Fitness), by Distribution Channel (Retail Pharmacies, Online Pharmacies, Hypermarkets) and Regional Forecast 2019-2026. Fortune Business Insights. 2019 Aug. URL: <https://www.fortunebusinessinsights.com/industry-reports/wearable-medical-devices-market-101070> [accessed 2020-08-22]
3. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EEK, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. *The Lancet Diabetes & Endocrinology* 2016 Dec;4(12):983-995. [doi: [10.1016/S2213-8587\(16\)30284-4](https://doi.org/10.1016/S2213-8587(16)30284-4)] [Medline: [27717766](https://pubmed.ncbi.nlm.nih.gov/27717766/)]
4. Jo A, Coronel BD, Coakes CE, Mainous AG. Is There a Benefit to Patients Using Wearable Devices Such as Fitbit or Health Apps on Mobiles? A Systematic Review. *Am J Med* 2019 Dec;132(12):1394-1400.e1. [doi: [10.1016/j.amjmed.2019.06.018](https://doi.org/10.1016/j.amjmed.2019.06.018)] [Medline: [31302077](https://pubmed.ncbi.nlm.nih.gov/31302077/)]
5. Burnham JP, Lu C, Yaeger LH, Bailey TC, Kollef MH. Using wearable technology to predict health outcomes: a literature review. *J Am Med Inform Assoc* 2018 Sep 01;25(9):1221-1227 [FREE Full text] [doi: [10.1093/jamia/ocy082](https://doi.org/10.1093/jamia/ocy082)] [Medline: [29982520](https://pubmed.ncbi.nlm.nih.gov/29982520/)]
6. Landi H. Patients will use health wearables to reduce trips to the doctor: survey. Fierce Healthcare. 2019 Jun 25. URL: <https://www.fiercehealthcare.com/tech/patients-will-use-health-wearables-to-reduce-trips-to-doctor-survey> [accessed 2020-08-22]
7. Wulfovich S, Fiordelli M, Rivas H, Concepcion W, Wac K. Design Implications for Mobile Apps and Wearables Contributing to Self-Efficacy of Patients With Chronic Conditions. *Front Psychol* 2019;10:2388 [FREE Full text] [doi: [10.3389/fpsyg.2019.02388](https://doi.org/10.3389/fpsyg.2019.02388)] [Medline: [31749733](https://pubmed.ncbi.nlm.nih.gov/31749733/)]
8. Kańtoch E. Recognition of Sedentary Behavior by Machine Learning Analysis of Wearable Sensors during Activities of Daily Living for Telemedical Assessment of Cardiovascular Risk. *Sensors (Basel)* 2018 Sep 24;18(10):3219 [FREE Full text] [doi: [10.3390/s18103219](https://doi.org/10.3390/s18103219)] [Medline: [30249987](https://pubmed.ncbi.nlm.nih.gov/30249987/)]
9. Stavropoulos TG, Papastergiou A, Mpaltadoros L, Nikolopoulos S, Kompatsiaris I. IoT Wearable Sensors and Devices in Elderly Care: A Literature Review. *Sensors (Basel)* 2020 May 16;20(10):2826 [FREE Full text] [doi: [10.3390/s20102826](https://doi.org/10.3390/s20102826)] [Medline: [32429331](https://pubmed.ncbi.nlm.nih.gov/32429331/)]
10. Patel M, Asch D, Volpp K. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA* 2015 Feb 03;313(5):459-460 [FREE Full text] [doi: [10.1001/jama.2014.14781](https://doi.org/10.1001/jama.2014.14781)] [Medline: [25569175](https://pubmed.ncbi.nlm.nih.gov/25569175/)]
11. Xue Y. A review on intelligent wearables: Uses and risks. *Human Behav and Emerg Tech* 2019 Sep 16;1(4):287-294 [FREE Full text] [doi: [10.1002/hbe2.173](https://doi.org/10.1002/hbe2.173)]
12. Piwek L, Ellis DA, Andrews S, Joinson A. The Rise of Consumer Health Wearables: Promises and Barriers. *PLoS Med* 2016 Feb;13(2):e1001953 [FREE Full text] [doi: [10.1371/journal.pmed.1001953](https://doi.org/10.1371/journal.pmed.1001953)] [Medline: [26836780](https://pubmed.ncbi.nlm.nih.gov/26836780/)]
13. Amyx S. Privacy Dangers of Wearables and the Internet of Things. In: Identity Theft: Breakthroughs in Research and Practice. Hershey, PA: IGI Global; 2017:379-402.
14. Perez A, Zeadally S. Privacy Issues and Solutions for Consumer Wearables. *IT Prof* 2018 Jul;20(4):46-56 [FREE Full text] [doi: [10.1109/mitp.2017.265105905](https://doi.org/10.1109/mitp.2017.265105905)]
15. Li H, Wu J, Gao Y, Shi Y. Examining individuals' adoption of healthcare wearable devices: An empirical study from privacy calculus perspective. *Int J Med Inform* 2016 Apr;88:8-17. [doi: [10.1016/j.ijmedinf.2015.12.010](https://doi.org/10.1016/j.ijmedinf.2015.12.010)] [Medline: [26878757](https://pubmed.ncbi.nlm.nih.gov/26878757/)]

16. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovic V. Literature on Wearable Technology for Connected Health: Scoping Review of Research Trends, Advances, and Barriers. *J Med Internet Res* 2019 Sep 05;21(9):e14017 [[FREE Full text](#)] [doi: [10.2196/14017](https://doi.org/10.2196/14017)] [Medline: [31489843](https://pubmed.ncbi.nlm.nih.gov/31489843/)]
17. Chuah S, Rauschnabel P, Krey N, Nguyen B, Ramayah T, Lade S. Wearable technologies: The role of usefulness and visibility in smartwatch adoption. *Computers in Human Behavior* 2016 Dec;65:276-284 [[FREE Full text](#)] [doi: [10.1016/j.chb.2016.07.047](https://doi.org/10.1016/j.chb.2016.07.047)]
18. Hailu R. Fitbits and other wearables may not accurately track heart rates in people of color. *STAT Internet*. 2019 Jul 24. URL: <https://www.statnews.com/2019/07/24/fitbit-accuracy-dark-skin/> [accessed 2020-08-22]
19. Sergueeva K, Shaw N, Lee SH. Understanding the barriers and factors associated with consumer adoption of wearable technology devices in managing personal health. *Can J Adm Sci* 2019 Sep 02;37(1):45-60 [[FREE Full text](#)] [doi: [10.1002/cjas.1547](https://doi.org/10.1002/cjas.1547)]
20. Nieroda M, Mrad M, Solomon M. How do consumers think about hybrid products? Computer wearables have an identity problem. *Journal of Business Research* 2018 Aug;89:159-170 [[FREE Full text](#)] [doi: [10.1016/j.jbusres.2018.04.024](https://doi.org/10.1016/j.jbusres.2018.04.024)]
21. Gartner survey shows wearable devices need to be more useful. Gartner. 2016 Dec 07. URL: <https://www.gartner.com/en/newsroom/press-releases/2016-12-07-gartner-survey-shows-wearable-devices-need-to-be-more-useful> [accessed 2020-08-22]
22. Ledger D, McCaffrey D. Inside wearables: How the science of human behavior change offers the secret to long-term engagement. Endeavour Partners. 2014 Jan. URL: <https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3> [accessed 2020-08-22]
23. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. *JMIR Mhealth Uhealth* 2019 Apr 12;7(4):e11819 [[FREE Full text](#)] [doi: [10.2196/11819](https://doi.org/10.2196/11819)] [Medline: [30977740](https://pubmed.ncbi.nlm.nih.gov/30977740/)]
24. Wu J, Li H, Cheng S, Lin Z. The promising future of healthcare services: When big data analytics meets wearable technology. *Information & Management* 2016 Dec;53(8):1020-1033 [[FREE Full text](#)] [doi: [10.1016/j.im.2016.07.003](https://doi.org/10.1016/j.im.2016.07.003)]
25. Bini S, Shah R, Bendich I, Patterson J, Hwang K, Zaid M. Machine Learning Algorithms Can Use Wearable Sensor Data to Accurately Predict Six-Week Patient-Reported Outcome Scores Following Joint Replacement in a Prospective Trial. *J Arthroplasty* 2019 Oct 1;34(10):2242-2247 [[FREE Full text](#)] [doi: [10.1016/j.arth.2019.07.024](https://doi.org/10.1016/j.arth.2019.07.024)] [Medline: [31439405](https://pubmed.ncbi.nlm.nih.gov/31439405/)]
26. Nweke H, Teh Y, Al-garadi M, Alo U. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications* 2018 Sep;105:233-261 [[FREE Full text](#)] [doi: [10.1016/j.eswa.2018.03.056](https://doi.org/10.1016/j.eswa.2018.03.056)]
27. Meng Y, Speier W, Shufelt C, Joung S, E Van Eyk J, Bairey Merz CN, et al. A Machine Learning Approach to Classifying Self-Reported Health Status in a Cohort of Patients With Heart Disease Using Activity Tracker Data. *IEEE J. Biomed. Health Inform* 2020 Mar;24(3):878-884 [[FREE Full text](#)] [doi: [10.1109/jbhi.2019.2922178](https://doi.org/10.1109/jbhi.2019.2922178)]
28. Davergne T, Pallot A, Dechartres A, Fautrel B, Gossec L. Use of Wearable Activity Trackers to Improve Physical Activity Behavior in Patients With Rheumatic and Musculoskeletal Diseases: A Systematic Review and Meta-Analysis. *Arthritis Care Res (Hoboken)* 2019 Jun;71(6):758-767 [[FREE Full text](#)] [doi: [10.1002/acr.23752](https://doi.org/10.1002/acr.23752)] [Medline: [30221489](https://pubmed.ncbi.nlm.nih.gov/30221489/)]
29. Leese J, Macdonald G, Tran B, Wong R, Backman C, Townsend A, et al. Using Physical Activity Trackers in Arthritis Self-Management: A Qualitative Study of Patient and Rehabilitation Professional Perspectives. *Arthritis Care Res (Hoboken)* 2019 Feb;71(2):227-236 [[FREE Full text](#)] [doi: [10.1002/acr.23780](https://doi.org/10.1002/acr.23780)] [Medline: [30295430](https://pubmed.ncbi.nlm.nih.gov/30295430/)]
30. Janevic MR, Shute V, Murphy SL, Piette JD. Acceptability and Effects of Commercially Available Activity Trackers for Chronic Pain Management Among Older African American Adults. *Pain Med* 2020 Feb 01;21(2):e68-e78. [doi: [10.1093/pm/pnz215](https://doi.org/10.1093/pm/pnz215)] [Medline: [31509196](https://pubmed.ncbi.nlm.nih.gov/31509196/)]
31. Leong I. Exercise: Wearable activity trackers in cardiovascular research. *Nat Rev Endocrinol* 2018 May;14(5):253 [[FREE Full text](#)] [doi: [10.1038/nrendo.2018.33](https://doi.org/10.1038/nrendo.2018.33)] [Medline: [29546876](https://pubmed.ncbi.nlm.nih.gov/29546876/)]
32. Zhang M, Luo M, Nie R, Zhang Y. Technical attributes, health attribute, consumer attributes and their roles in adoption intention of healthcare wearable technology. *Int J Med Inform* 2017 Dec;108:97-109. [doi: [10.1016/j.ijmedinf.2017.09.016](https://doi.org/10.1016/j.ijmedinf.2017.09.016)] [Medline: [29132639](https://pubmed.ncbi.nlm.nih.gov/29132639/)]
33. Park E. User acceptance of smart wearable devices: An expectation-confirmation model approach. *Telematics and Informatics* 2020 Apr;47:101318 [[FREE Full text](#)] [doi: [10.1016/j.tele.2019.101318](https://doi.org/10.1016/j.tele.2019.101318)]
34. Papa A, Mital M, Pisano P, Del Giudice M. E-health and wellbeing monitoring using smart healthcare devices: An empirical investigation. *Technological Forecasting and Social Change* 2020 Apr;153:119226 [[FREE Full text](#)] [doi: [10.1016/j.techfore.2018.02.018](https://doi.org/10.1016/j.techfore.2018.02.018)]
35. Lee S, Lee K. Factors that influence an individual's intention to adopt a wearable healthcare device: The case of a wearable fitness tracker. *Technological Forecasting and Social Change* 2018 Apr;129:154-163 [[FREE Full text](#)] [doi: [10.1016/j.techfore.2018.01.002](https://doi.org/10.1016/j.techfore.2018.01.002)]
36. Wang H, Tao D, Yu N, Qu X. Understanding consumer acceptance of healthcare wearable devices: An integrated model of UTAUT and TTF. *Int J Med Inform* 2020 Jul;139:104156. [doi: [10.1016/j.ijmedinf.2020.104156](https://doi.org/10.1016/j.ijmedinf.2020.104156)] [Medline: [32387819](https://pubmed.ncbi.nlm.nih.gov/32387819/)]
37. Li J, Ma Q, Chan AH, Man SS. Health monitoring through wearable technologies for older adults: Smart wearables acceptance model. *Appl Ergon* 2019 Feb;75:162-169. [doi: [10.1016/j.apergo.2018.10.006](https://doi.org/10.1016/j.apergo.2018.10.006)] [Medline: [30509522](https://pubmed.ncbi.nlm.nih.gov/30509522/)]



38. Abouzahra M, Ghasemaghaei M. The antecedents and results of seniors' use of activity tracking wearable devices. *Health Policy and Technology* 2020 Jun;9(2):213-217 [FREE Full text] [doi: [10.1016/j.hlpt.2019.11.002](https://doi.org/10.1016/j.hlpt.2019.11.002)]
39. Methodological Reports of HINTS: Health Information National Trends Survey. National Cancer Institute. URL: <https://hints.cancer.gov/data/methodology-reports.aspx> [accessed 2020-08-22]
40. Kolenikov S. Resampling Variance Estimation for Complex Survey Data. *The Stata Journal* 2018 Nov 19;10(2):165-199 [FREE Full text] [doi: [10.1177/1536867x1001000201](https://doi.org/10.1177/1536867x1001000201)]
41. Lavrakas PJ, editor. Jackknife Variance Estimation. In: *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage Publications, Inc; 2008.
42. HINTS 5 Cycle 3 Public Codebook. Bethesda, MD: National Cancer Institute; Jan 2020.
43. Dehghani M, Kim K, Dangelico R. Will smartwatches last? factors contributing to intention to keep using smart wearable technology. *Telematics and Informatics* 2018 May 1;35(2):480-490 [FREE Full text] [doi: [10.1016/j.tele.2018.01.007](https://doi.org/10.1016/j.tele.2018.01.007)]
44. Kim J, Park E. Beyond coolness: Predicting the technology adoption of interactive wearable devices. *Journal of Retailing and Consumer Services* 2019 Jul;49:114-119 [FREE Full text] [doi: [10.1016/j.jretconser.2019.03.013](https://doi.org/10.1016/j.jretconser.2019.03.013)]
45. Ridgers N, McNarry M, Mackintosh K. Feasibility and Effectiveness of Using Wearable Activity Trackers in Youth: A Systematic Review. *JMIR Mhealth Uhealth* 2016 Nov 23;4(4):e129 [FREE Full text] [doi: [10.2196/mhealth.6540](https://doi.org/10.2196/mhealth.6540)] [Medline: [27881359](https://pubmed.ncbi.nlm.nih.gov/27881359/)]
46. Schaffarczyk L, Ilhan A. Healthier Life and More Fun? Users' Motivations to Apply Activity Tracking Technology and the Impact of Gamification. 2019 Presented at: International Conference on Human-Computer Interaction; July 26-31, 2019; Orlando, FL p. 124-136 URL: [https://link.springer.com/chapter/10.1007/978-3-030-21905-5\\_10](https://link.springer.com/chapter/10.1007/978-3-030-21905-5_10) [doi: [10.1007/978-3-030-21905-5\\_10](https://doi.org/10.1007/978-3-030-21905-5_10)]
47. Spil T, Sunyaev A, Thiebes S, Van Baalen R. The Adoption of Wearables for a Healthy Lifestyle: Can Gamification Help? 2017 Presented at: Hawaii International Conference on System Sciences (HICSS-50); January 4-7, 2017; Waikoloa Village, HI URL: [https://aisel.aisnet.org/hicss-50/hc/apps\\_for\\_health\\_management/7/](https://aisel.aisnet.org/hicss-50/hc/apps_for_health_management/7/) [doi: [10.24251/hicss.2017.437](https://doi.org/10.24251/hicss.2017.437)]
48. Brickwood K, Williams AD, Watson G, O'Brien J. Older adults' experiences of using a wearable activity tracker with health professional feedback over a 12-month randomised controlled trial. *Digit Health* 2020;6:2055207620921678 [FREE Full text] [doi: [10.1177/2055207620921678](https://doi.org/10.1177/2055207620921678)] [Medline: [32426152](https://pubmed.ncbi.nlm.nih.gov/32426152/)]
49. Maher C, Ryan J, Ambrosi C, Edney S. Users' experiences of wearable activity trackers: a cross-sectional study. *BMC Public Health* 2017 Nov 15;17(1):880 [FREE Full text] [doi: [10.1186/s12889-017-4888-1](https://doi.org/10.1186/s12889-017-4888-1)] [Medline: [29141607](https://pubmed.ncbi.nlm.nih.gov/29141607/)]
50. Lunney A, Cunningham N, Eastin M. Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes. *Computers in Human Behavior* 2016 Dec 1;65:114-120 [FREE Full text] [doi: [10.1016/j.chb.2016.08.007](https://doi.org/10.1016/j.chb.2016.08.007)]
51. Kekade S, Hsieh C, Islam MM, Atique S, Mohammed Khalfan A, Li Y, et al. The usefulness and actual use of wearable devices among the elderly population. *Comput Methods Programs Biomed* 2018 Jan;153:137-159. [doi: [10.1016/j.cmpb.2017.10.008](https://doi.org/10.1016/j.cmpb.2017.10.008)] [Medline: [29157447](https://pubmed.ncbi.nlm.nih.gov/29157447/)]
52. Pobiruchin M, Suleder J, Zowalla R, Wiesner M. Accuracy and Adoption of Wearable Technology Used by Active Citizens: A Marathon Event Field Study. *JMIR Mhealth Uhealth* 2017 Feb 28;5(2):e24 [FREE Full text] [doi: [10.2196/mhealth.6395](https://doi.org/10.2196/mhealth.6395)] [Medline: [28246070](https://pubmed.ncbi.nlm.nih.gov/28246070/)]
53. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of Commercially Available Wearable Activity Trackers Among Adults Aged Over 50 and With Chronic Illness: A Mixed-Methods Evaluation. *JMIR Mhealth Uhealth* 2016 Jan 27;4(1):e7 [FREE Full text] [doi: [10.2196/mhealth.4225](https://doi.org/10.2196/mhealth.4225)] [Medline: [26818775](https://pubmed.ncbi.nlm.nih.gov/26818775/)]
54. Talukder M, Sorwar G, Bao Y, Ahmed J, Palash M. Predicting antecedents of wearable healthcare technology acceptance by elderly: A combined SEM-Neural Network approach. *Technological Forecasting and Social Change* 2020 Jan;150:119793 [FREE Full text] [doi: [10.1016/j.techfore.2019.119793](https://doi.org/10.1016/j.techfore.2019.119793)]
55. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. *J Med Internet Res* 2014 Jul 16;16(7):e172 [FREE Full text] [doi: [10.2196/jmir.3117](https://doi.org/10.2196/jmir.3117)] [Medline: [25048379](https://pubmed.ncbi.nlm.nih.gov/25048379/)]

## Abbreviations

**HINTS:** Health Information National Trends Survey  
**OR:** odds ratio



*Edited by G Eysenbach; submitted 12.07.20; peer-reviewed by KJ Brickwood; comments to author 22.07.20; revised version received 26.07.20; accepted 26.07.20; published 16.10.20.*

*Please cite as:*

Chandrasekaran R, Katthula V, Moustakas E

*Patterns of Use and Key Predictors for the Use of Wearable Health Care Devices by US Adults: Insights from a National Survey*

*J Med Internet Res* 2020;22(10):e22443

URL: <http://www.jmir.org/2020/10/e22443/>

doi: [10.2196/22443](https://doi.org/10.2196/22443)

PMID: [33064083](https://pubmed.ncbi.nlm.nih.gov/33064083/)

©Ranganathan Chandrasekaran, Vipanchi Katthula, Evangelos Moustakas. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 16.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Wearable Technology Acceptance in Health Care Based on National Culture Differences: Cross-Country Analysis Between Chinese and Swiss Consumers

Dong Yang Meier<sup>1\*</sup>, MSc; Petra Barthelmess<sup>1\*</sup>, Prof Dr; Wei Sun<sup>1\*</sup>, MSc; Florian Liberatore<sup>2\*</sup>, PhD

<sup>1</sup>International Management Institute, School of Management and Law, Zurich University of Applied Sciences, Winterthur, Switzerland

<sup>2</sup>Winterthur Institute of Health Economics, School of Management and Law, Zurich University of Applied Sciences, Winterthur, Switzerland

\* all authors contributed equally

**Corresponding Author:**

Dong Yang Meier, MSc

International Management Institute

School of Management and Law

Zurich University of Applied Sciences

Theaterstrasse 17

Winterthur, 8400

Switzerland

Phone: 41 792987655

Email: [yanm@zhaw.ch](mailto:yanm@zhaw.ch)

## Abstract

**Background:** The advancement of wearable devices and growing demand of consumers to monitor their own health have influenced the medical industry. Health care providers, insurers, and global technology companies intend to develop more wearable devices incorporating medical technology and to target consumers worldwide. However, acceptance of these devices varies considerably among consumers of different cultural backgrounds. Consumer willingness to use health care wearables is influenced by multiple factors that are of varying importance in various cultures. However, there is insufficient knowledge of the extent to which social and cultural factors affect wearable technology acceptance in health care.

**Objective:** The aims of this study were to examine the influential factors on the intention to adopt health care wearables, and the differences in the underlying motives and usage barriers between Chinese and Swiss consumers.

**Methods:** A new model for acceptance of health care wearables was conceptualized by incorporating predictors of different theories such as technology acceptance, health behavior, and privacy calculus based on an existing framework. To verify the model, a web-based survey in both the Chinese and German languages was conducted in China and Switzerland, resulting in 201 valid Chinese and 110 valid Swiss respondents. A multigroup partial least squares path analysis was applied to the survey data.

**Results:** Performance expectancy ( $\beta=.361, P<.001$ ), social influence ( $\beta=.475, P<.001$ ), and hedonic motivation ( $\beta=.111, P=.01$ ) all positively affected the behavioral intention of consumers to adopt wearables, whereas effort expectancy, functional congruence, health consciousness, and perceived privacy risk did not demonstrate a significant impact on behavioral intention. The group-specific path coefficients indicated health consciousness ( $\beta=.150, P=.01$ ) as a factor positively affecting only the behavior intention of the Chinese respondents, whereas the factors affecting only the behavioral intention of the Swiss respondents proved to be effort expectancy ( $\beta=.165, P=.02$ ) and hedonic motivation ( $\beta=.212, P=.02$ ). Performance expectancy asserted more of an influence on the behavioral intention of the Swiss ( $\beta=.426, P<.001$ ) than the Chinese ( $\beta=.271, P<.001$ ) respondents, whereas social influence had a greater influence on the behavioral intention of the Chinese ( $\beta=.321, P<.001$ ) than the Swiss ( $\beta=.217, P=.004$ ) respondents. Overall, the Chinese consumers displayed considerably higher behavioral intention ( $P<.001$ ) than the Swiss. These discrepancies are explained by differences in national culture.

**Conclusions:** This is one of the first studies to investigate consumers' intention to adopt wearables from a cross-cultural perspective. This provides a theoretical and methodological foundation for future research, as well as practical implications for global vendors and insurers developing and promoting health care wearables with appropriate features in different countries. The testimonials and support by physicians, evidence of measurement accuracy, and easy handling of health care wearables would be useful in promoting the acceptance of wearables in Switzerland. The opinions of in-group members, involvement of employers,

and multifunctional apps providing credible health care advice and solutions in cooperation with health care institutions would increase acceptance among the Chinese.

(*J Med Internet Res* 2020;22(10):e18801) doi:[10.2196/18801](https://doi.org/10.2196/18801)

## KEYWORDS

wearables; health care wearables; wearables acceptance; cross culture; national culture; Chinese; Swiss; moderator; digital health; health technology acceptance; smartwatch

## Introduction

### Background

The global market for wearable devices is growing at a remarkable rate [1]. The terms “wearable technology,” “wearable devices,” and “wearables” all refer to electronic technologies or computers incorporated into items of accessories and clothing, which can comfortably be worn on the body and enable users to collect and self-monitor their health vitals [2]. In this paper, we use the word “wearables” to represent the terms “wearable technology” and “wearable devices.” As the number of potential users continues to grow, wearables will have increasing sociological and cultural impacts [2]. Numerous global companies developing wearables in health care aim to target consumers in many countries. Nevertheless, the intention to accept and adopt wearables varies tremendously among consumers of different cultural backgrounds. It is evident that countries that differ greatly regarding their technological development, social structure, and usage habits have different levels of technology acceptance [3].

Some literature on the influential factors of consumers’ acceptance and adoption of health care wearables is available [4-12]. There are also some studies that focused on the differences in these influential factors in various national cultures with respect to technology acceptance [3,13,14]. However, there is a notable research gap concerning variation in the influential factors on consumers’ acceptance of health care wearables from distinct national cultures [5,6,11]. Based on this, the following research questions were formulated: What are the influential factors on consumers’ behavioral intention to adopt health care wearables? How different are the drivers of the behavioral intention to adopt health care wearables of Chinese and Swiss consumers?

The objective of this study was to investigate certain patterns of influential factors on usage intention of health care wearables by means of comparison of essential acceptance motives and usage barriers of Chinese and Swiss consumers. The different perceptions between Chinese and Swiss users were compared and are discussed in view of the differing national cultures of these two countries. Results of this study will have implications for global digital technology providers to develop and market wearables successfully across borders, as well as possible incentives that the insurers might offer with respect to lifestyle changes to enhance people’s health conditions effectively.

## Prior Research

### Theories and Models of Wearable Technology Acceptance

Technology acceptance is widely recognized as an aspect of understanding the approval, favorable reception, and continued use of newly introduced devices and systems [3]. Davis [15] developed the first technology acceptance model (TAM), in which perceived usefulness and perceived ease of use were shown to be two main factors affecting the attitude of a user toward new technologies. The TAM was subsequently expanded to include more factors influencing users’ acceptance of computer-related technologies. Recent studies have investigated technology acceptance on the part of consumers, particularly in the area of information technology. This is often performed in the context of the model of unified theory of acceptance and use of technology 2 (UTAUT2) introduced by Venkatesh et al [16]. In the UTAUT2 model, intrinsic motivation such as hedonic motivation and two aspects of consumer behavior, namely price value and habit, were added to the previous construct of UTAUT. The acceptance of medical technology intersects with intimate and personal aspects, and is therefore a highly sensitive topic that sets itself apart from the acceptance of information technology in general [3].

By examining the factors influencing mobile health technology acceptance, Rogers’ Protection Motivation Theory (PMT) [17] has been integrated into the health TAM [4]. Since health care wearables collect users’ personal health information on an ongoing basis, concern about data privacy risk increases. An individual’s decision to adopt health care wearable devices would involve an obvious privacy calculus, in which users may consider the tradeoff between perceived benefit and perceived privacy risk [5]. Based on this theory, various researchers have added and abandoned variables according to the characteristics of technology and the targeted user groups to predict a user’s intention to adopt health care wearables. Gao et al [6] developed an integrated framework comprehensively examining wearable technology acceptance in health care (WTAH) by combining the theories described above [18,19]. Gao et al [6] tested this framework through an empirical survey conducted in China with 462 qualified responses (users of health care wearables) and confirmed that the 8 predictors in their WTAH model, namely performance expectancy, hedonic motivation, effort expectancy, functional congruence, self-efficacy, social influence, perceived vulnerability, and perceived severity, positively influence an individual’s intention to adopt health care wearables; perceived privacy risk negatively affected an individual’s intention to adopt health care wearables. Among all factors, social influence and perceived privacy risk were the most significant predictors [6]. However, this survey was only

conducted in China, which did not consider the cultural differences between different countries [6]. Hence, testing whether the approved relationships are still held in other countries is necessary.

### ***Influence of National Culture on Technology Acceptance***

A country's cultural values influence technology acceptance [13]. Although the TAM has been used extensively when studying information technology adoption in Western countries, researchers noted that the TAM was not valid when applied to other cultures [13,20]. The UTAUT model was tested in non-Western cultures such as in Saudi Arabia [21], India [22], and China as compared to the United States [23]. These studies provided evidence that there is an interaction between the two phenomena of technology acceptance and national culture.

Several sets of dimensions have been developed to characterize the concept of national culture [24]. At present, at least 6 models of national cultures are widely cited and utilized in the management research literature [14]. These 6 culture models attempt to provide a well-reasoned set of dimensions to facilitate comparison of differing cultures. Among these models, Hofstede and Minkov [25] provided detailed guidelines to explain and measure cultural value differences applied to national culture. To date, these cultural dimensions have been the most commonly used variables for examining models cross-culturally [13,20,24,26]. McCoy et al [27] conducted a simple analysis of variance for each of Hofstede and Minkov's [25] cultural dimensions measured at the individual level across 8 countries. All of the F-scores obtained were significant at  $P < .001$ , which provides clear evidence of the existence of national culture (ie, the variance between groups is larger than the variance within groups) [27]. For this reason, this study primarily adopted the Hofstede and Minkov dimensions (hereafter referred to as "Hofstede's cultural dimensions") to describe, explain, and to a degree measure the difference in the national cultures of Chinese and Swiss consumers.

Hofstede's cultural dimensions [28] utilized a national level of analysis, whereas most TAMs were developed for an individual level of analysis. Ford et al [18] stated that it would be beneficial to consider national culture as a moderating variable, since this

might play an important role in comparing different populations. Alshare et al [13] confirmed this statement in their empirical study with samples from the United States, Chile, and the United Arab Emirates, showing that national culture dimensions represented by masculinity, power distance, individualism, and uncertainty avoidance moderate four relationships of an extended TAM. McCoy et al [29] showed that high power distance, high masculinity, low uncertainty avoidance, and high collectivism seem to nullify the effects of perceived ease of use or perceived usefulness in the TAM. Taken together, these studies suggested that national culture moderates relationships in the extended TAM, UTAUT, and other related models [13]. Following a similar reasoning, this study examined the moderating role of the national cultures of China and Switzerland with an adapted conceptual model of WTAH.

### ***Differences in National Culture Between the Chinese and Swiss***

The differences in the national cultures of China and Switzerland according to Hofstede's country comparison tool [30] are summarized in Table 1. According to Hofstede's [31] cultural dimension, Switzerland holds the cultural values of high individualism, moderately high uncertainty avoidance, moderately high indulgence, and low power distance. These contrast with the cultural values of the Chinese of low individualism (high collectivism), low indulgence (high restraint), moderately low uncertainty avoidance, and high power distance. Both countries have similarly high values of masculinity and long-term orientation. Multilingualism is an essential part of Switzerland's identity, with more than 64% of the population speaking German, and approximately 20% speaking French, 8% Italian, and 1% Romansh. Hofstede and Minkov [30] indicated that the German-speaking and French-speaking parts of Switzerland exhibit slightly different cultural values in the aspects of power distance and uncertainty avoidance, but are otherwise very similar with respect to the values of individualism, masculinity, long-term orientation, and indulgence. Therefore, only the cultural values of the German-speaking part of Switzerland were considered for this study.

**Table 1.** Cultural scores of China and Switzerland according to Hofstede and Minkov [30].

Cultural dimensions	China	Switzerland
Power distance	80	34
Individualism	20	68
Masculinity	66	70
Uncertainty avoidance	30	58
Long-term orientation	87	74
Indulgence	24	66

The cultural differences between the Chinese and the Swiss are additionally presented in Table 2, as mapped onto the "Big Five" dimensions in Nardon and Steer's [14] comparative study of all cultural models. Among these five dimensions,

"relationship with the environment" and "time orientation" are additional dimensions that were not explicitly mentioned by Hofstede.

**Table 2.** Country ratings of China and Switzerland in line with “Big Five” dimensions [14].<sup>a</sup>

Cultural dimensions	China	Switzerland
Relationship with the environment	Harmony	Mastery
Social organization	Collectivist+	Individualist
Power distribution	Hierarchical	Egalitarian
Rule orientation	Relationship-based	Rule-based+
Time orientation	Polychronic	Monochronic+

<sup>a</sup>All ratings are comparative, with a “+” sign indicating a stronger tendency toward a particular dimension.

## Methods

### Conceptual Model

To analyze influential factors for the intention to adopt health care wearables, and especially to distinguish the different perspectives between Chinese and Swiss consumers, a conceptual model was developed, which is illustrated in [Figure 1](#). This model was adapted from Gao et al’s [6] WTAH model, since it is one of the most comprehensive models that incorporates consumers’ behavioral intention on technology acceptance (UTAUT2), health behavior (PMT), and privacy calculus theories. The variables such as performance expectancy, hedonic motivation, effort expectancy, functional congruence, social influence, and perceived privacy risk are factors that influence the behavioral intention of consumers using health care wearables, which are adopted from the WTAH model. However, three other variables (self-efficacy, perceived vulnerability, and perceived severity) of the WTAH model based on the PMT were replaced by a single variable, “health consciousness,” in this study, for the following reasons.

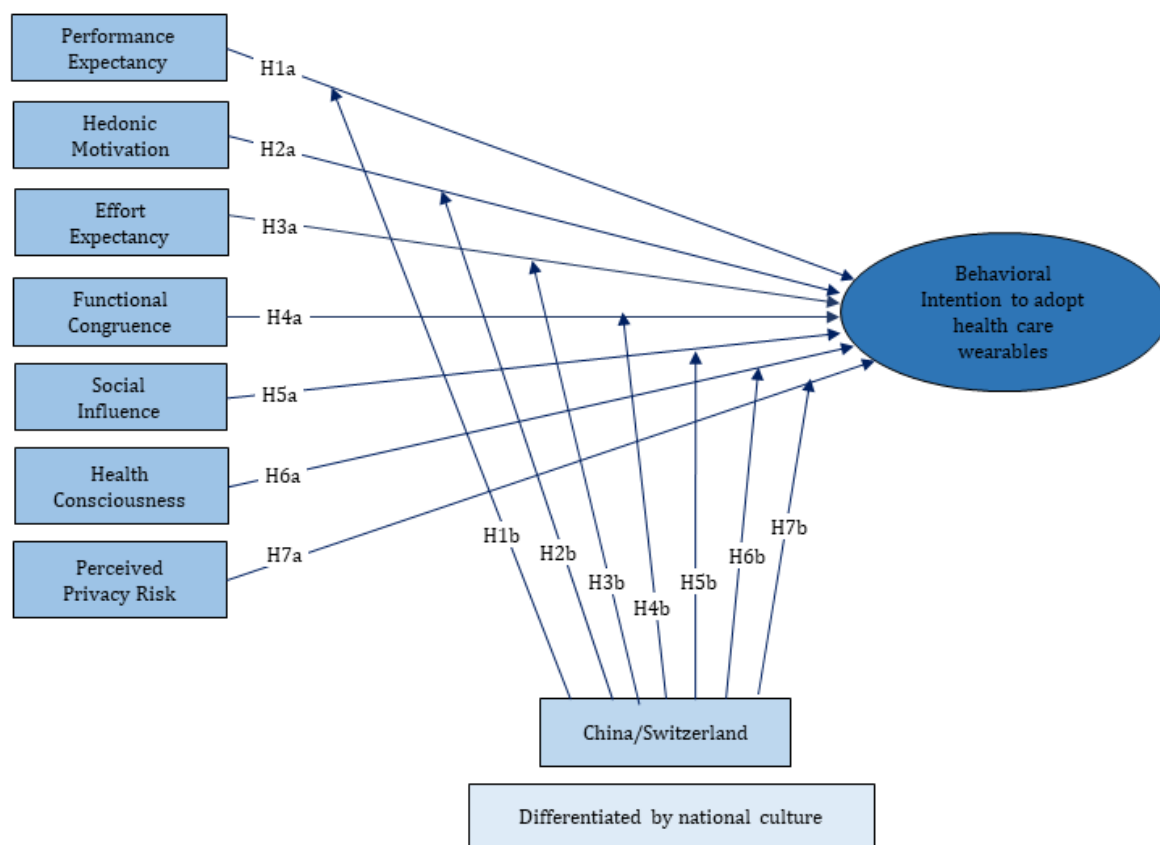
Perceived vulnerability refers to the possibility that one will experience a threat of certain diseases, whereas perceived severity represents the extent of the threat of certain diseases [4]. Sun et al [4] argued that the factors relevant to threat appraisal of PMT have only relatively weak (perceived vulnerability) or no (perceived severity) effects on behavioral intention to accept mobile technologies of health services. This is consistent with the meta-analysis results of Floyd et al [4,32]. Part of the health care wearables of focus in this study are consumer-grade devices that are typically used by relatively healthy members of the population who are interested in fitness/wellness. Therefore, the factors in the PMT model would not be suitable for measuring the acceptance of fitness/wellness

wearables, which constitute a significant portion of health care wearables. This was especially confirmed during the first round of the pilot study with the first version of the questionnaire, which included the variables of PMT. Some Swiss participants of the pilot study could not answer the related questions, and particularly could not distinguish the question with five different levels of a Likert scale, even when prompted with the situation that they would suffer from a certain disease and have poor knowledge about self-care regarding that disease. Self-efficacy is the belief in one’s ability to use health care wearables to monitor and improve their health condition [6]. The aspects related to this variable are partly covered by the variable effort expectancy, which was remarked as high by the test respondents in the first-round pilot study.

Health consciousness is conceptualized as the extent to which individuals have interest in and are aware of their own health conditions and well-being, and the extent to which a person maintains their own health [7]. A survey by the consulting firm MarketsandMarkets [33] described that the increasing health consciousness among people drives the growth of the wearable technology market. Cho et al [7] as well as Chen and Lin [8] confirmed that health consciousness has a significant direct effect on the perceived ease of use and usefulness of dietary and fitness apps. People’s behavior to maintain their health in the aspect of “health consciousness” covers the aspect of “self-efficacy.” Consequently, this study proposes that health consciousness represents people’s general health concern, awareness, and behavior instead of factors in the PMT.

The definitions of all factors in the proposed model are listed in [Table 3](#), which are adapted from previous published studies [6,7] with minor modifications in wording to fit into the health care wearables context.



**Figure 1.** Conceptual model.**Table 3.** Definitions of factors in the conceptual model.

Construct	Definition/Explanation
Performance expectancy	Degree to which adopting health care wearables will bring effectiveness to users in improving their health condition, which includes monitoring daily physical conditions, making personal health care plans, and reducing health-related threats.
Hedonic motivation	Pleasure or enjoyment derived from adopting and using health care wearables, such as enjoying the technical functions of the devices, sharing data with peers, and feeling of accomplishment after reaching the training goals.
Effort expectancy	Degree of perceived ease of using health care wearables, which includes wearing the device easily on the body, using other devices such as a smartphone to analyze the data, and understanding the data.
Functional congruence	Perceived suitability of health care wearables to fulfill the functional and basic product-related needs such as price reasonability, esthetics, and ergonomic design.
Social influence	Extent to which a user's decision-making is influenced by others' perceptions. These "others" include close relationships such as family members and close friends, important people such as employers or peers, professionals such as physicians, and technical specialists.
Health consciousness	Extent to which individuals have interest in and are aware of their own health condition and degree to which health concerns are integrated into their daily activities.
Perceived privacy risk	Perceived risk of reputation damage or other disadvantages by disclosing personal health data to people/organizations unwittingly.
Country China/Switzerland	Country dichotomy of China versus Switzerland distinguished by different national cultural values.
Behavioral intention	Users' formulation of conscious use or increasing use of health care wearables.

### Hypotheses Related to Influential Factors on Behavioral Intention

The effect of all of the independent variables (except for health consciousness) in this conceptual model were confirmed by

Gao et al [6] in their empirical study with Chinese respondents. As both Swiss and Chinese consumers were involved in this study, all of the influential factors in the conceptual model were tested with the entire group of valid respondents to evaluate

whether the relationships in this adapted model hold for both groups.

Based on this, the following hypotheses were drawn:

*H1a: Performance expectancy is positively correlated with an individual's intention to adopt health care wearables.*

*H2a: Hedonic motivation is positively correlated with an individual's intention to adopt health care wearables.*

*H3a: Effort expectancy is positively correlated with an individual's intention to adopt health care wearables.*

*H4a: Functional congruence is positively correlated with an individual's intention to adopt health care wearables.*

*H5a: Social influence is positively correlated with an individual's intention to adopt health care wearables.*

*H6a: Health consciousness is positively correlated with an individual's intention to adopt health care wearables.*

*H7a: Perceived privacy risk is negatively correlated with an individual's intention to adopt health care wearables.*

### **Hypotheses Related to Differences Between the Chinese and Swiss**

In the conceptual model, the country (China, Switzerland) distinguished by national culture acts as a moderating variable, which affects the influential degree of the above-mentioned independent variables and the individual's behavioral intention to adopt the wearables differently.

People in individualist cultures such as the Swiss culture pursue independence and freedom, and therefore advocate self-responsibility and self-reliance [31]. The opinions of close peers would not have much bearing on their decision to adopt wearables. By contrast, wearables enable them to live a more autonomous and freer lifestyle through self-monitoring of their health conditions (eg, not restricted by appointments with a physician, which was confirmed by Swiss interviewees). Therefore, performance expectancy of the Swiss on health care wearables might lead to higher intention to adopt health care wearables. Thus, we hypothesized:

*H1b: Performance expectancy has a greater impact on the intention to adopt wearable devices for Swiss consumers than for Chinese consumers.*

Swiss respondents exhibited a moderately high value of "indulgence," which means that they generally place a higher degree of importance on leisure time and having fun [31]. Pleasure or enjoyment derived from using health care wearables, such as enjoying the technical functions of the devices and the feeling of accomplishment after reaching the training goals, might cause the Swiss to have a higher intention to adopt health care variables. Therefore, we hypothesized:

*H2b: Hedonic motivation has greater impact on the intention to adopt wearable devices for Swiss consumers than for Chinese consumers.*

Ease of use of technology not only influences the user's motivation but also makes the technology more adaptive in the organization [20]. The individual Swiss consumer tends to solve technological problems by themselves much more than the Chinese consumer, who tends to live in a close community, taking support from the community for granted according to cultural value dimensions [14,20,31]. In addition, the high uncertainty avoidance value causes the Swiss to perceive new technology as more difficult, because the functions and consequences are uncertain [31]. Accordingly, perceived ease of handling health care wearables might have a greater impact on the behavioral intention of Swiss consumers. Therefore, we hypothesized:

*H3b: Effort expectancy has greater impact on the intention to adopt wearable devices for Swiss consumers than for Chinese consumers.*

The general income in China is dramatically lower than that in Switzerland, and this was clearly reflected in the sampling of this study. Thus, functional and basic product-related needs such as price reasonability, esthetics, and ergonomic design would have more impact on a Chinese consumer's intention to use health care wearables than a Swiss consumer's. Therefore, we hypothesized:

*H4b: Functional congruence has greater impact on the intention to adopt wearable devices of Chinese consumers than for Swiss consumers.*

With their generally collectivist values, Chinese consumers are more concerned about the maintenance of group cohesion, put more weight on the opinions of in-group members, and tend to assimilate their opinions or behaviors in their close community [31]. Researchers found that in collectivist countries, the positive effect of social influence on technology acceptance is stronger than that in individualist countries [34]. People in collectivist countries (ie, China) tend to seek out new information from their peers who have already adopted the technology, in contrast to people in individualist countries (ie, Switzerland), who tend to seek information on their own from formal/external sources [20].

The high power distance of Chinese consumers leads to the strong influence of superiors, employers, or authority on adopting wearables for health care or other purposes. For example, some Chinese companies distribute locally produced smartwatches to all of their employees as a kind of fringe benefit for health care. Through the encouragement, support, and influence of the local environment, Chinese consumers find it easy to be in a group of wearable users, which further fosters their intention to adopt the wearables. Therefore, we hypothesized:

*H5b: Social influence has greater impact on the intention to adopt wearable devices on the part of Chinese consumers than Swiss consumers.*

It is well known that the health care and insurance system in China is far from developed; health care providers are usually

overburdened with large volumes of patients and expenses for individuals are relatively high. The health-conscious Chinese might choose to pursue a healthier lifestyle through using wearables to track their health condition and avoid disease rather than going for treatment at a hospital after getting ill. This would save time and expenditures on health care. Therefore, we hypothesized:

*H6b: Health consciousness has greater impact on the intention to adopt wearable devices on the part of Chinese consumers than Swiss consumers.*

The high uncertainty avoidance of Swiss consumers might make them reluctant to engage with new technology and devices, and to exercise more discretion with personal information. They would perceive the privacy risk much higher than Chinese consumers, and demand clear regulations before adopting digital health care appliances, which might exert more of a negative influence on their intention to use wearables. Therefore, we hypothesized:

*H7b: Perceived privacy risk has greater impact on the intention to adopt wearable devices on the part of Swiss consumers than Chinese consumers.*

## Construction of Questionnaire

To validate the conceptual model and hypotheses, a quantitative research approach was employed by developing a written questionnaire. Several interviews were first held in Switzerland and China with current users of fitness and medical wearables to ensure the relevance and objectivity of the questions. The measurement items (see [Multimedia Appendix 1](#)) for the independent variables performance expectancy, hedonic motivation, effort expectancy, functional congruence, social influence, and perceived privacy risk, and the dependent variable behavioral intention in the questionnaire were adapted from Gao et al [6]; those for health consciousness were adapted from Michaelidou and Hassan [35]. A 5-point Likert scale was employed to measure the items. The questions regarding the cultural value dimension “individualism” were adopted from Hofstede and Minskova’s “Values Survey Module 2013 Questionnaire” [25].

Both variables of “nationality at birth” and “country of residence” were initially assessed to particularly ensure that Chinese living in Switzerland and Swiss living in China were not included in the valid samples, so that the country variable represents a distinct national culture to meet the requirement of reliability. Users and nonusers of health care wearables were included in the collected samples so that the different perceptions and attitudes of both groups (users with high propensity of intention and nonusers with low propensity of intention) were considered.

The questionnaire was translated from English into German and simplified Chinese. The translated versions were corrected by more than two native speakers in each language and have been back-translated. The German language was used for the Swiss questionnaire, as it covers more than 64% of the Swiss population.

Three rounds of pilot studies were conducted; since the measured items are all nonobservable variables, different understandings occur due to language barriers (English, German, and simplified Chinese) and divergent cultural backgrounds. The first pilot study round was conducted initially with 5 Swiss natives representing different age groups. Consequently, the variables of PMT in the WTAH model were replaced by health consciousness. The second and third rounds of pilot studies were conducted with 20 Chinese and 9 Swiss participants, through which the questions were adjusted in both languages a few times to ascertain that all items were formulated clearly, in logical sequences, relevant to the everyday lives of the respondents, and relatively easy to answer.

Smartwatches were selected as the representative health care wearables in this study because they combine the features of consumer and medical-grade devices [9,36]. Most smartwatches can monitor some human physiological signals and biomechanics, and thus act as fitness tracking devices that help users record their daily activities such as automatically recording workout times, tracking heart rates, step counts, and calories burnt [10]. With added apps and sensors, the new generation of smartwatches can further measure health vitals such as electrocardiography, glucose level, and blood pressure, as well as detect certain diseases such as arrhythmia and seizure [9,37]. Smartwatches are the most frequently purchased wearable devices worldwide currently and will continue to be in the near term [38].

## Data Collection

Both finalized questionnaires were distributed to the Chinese and Swiss populations using web-based survey tools and a snowball sampling method as a convenience sampling technique [39]. The German version, compiled in Survey Monkey, was distributed to Swiss consumers by email, with the request to forward the survey further to their colleagues and friends. The questionnaire was distributed in diverse industry and service companies, fitness centers, leisure and sport clubs, as well as in local communities. The Chinese version, designed in “Wen Juan Xing,” was distributed to Chinese consumers in mainland China by email and the social media platform WeChat with a requirement to share the survey. The collected samples were the responses to the survey from April 2 to April 24, 2019. Until the evening of April 24, 2019, 153 samples were collected from Switzerland and 203 samples were collected from 23 of 32 provinces and municipalities of mainland China (excluding Hong Kong and Macau).

## Data Analysis

First, a descriptive analysis was conducted related to the sample characteristics and the cultural values (individualism/collectivism) of the Chinese and Swiss. After validity assessments, *t* tests were conducted to compare mean differences in the constructs between the Chinese and Swiss samples. In the last step, the research model was tested using a multigroup partial least squares path analysis method.

## Results

### Sample Description

Among the 356 received responses, 13 incomplete samples that ended in the middle of the questionnaire were deleted, and 32 respondents with other nationalities (2 living in China and 30 in Switzerland) were excluded in the data analysis, being that

the moderating factor is a distinct culture of China and Switzerland. This resulted in a complete valid sample of 311 respondents with 201 Chinese respondents living in China and 110 Swiss respondents living in Switzerland. The sample characteristics are displayed in [Table 4](#), showing that the Chinese and Swiss samples represent the population of each country respectively against the current societal and economic backgrounds.

**Table 4.** Sample characteristics (N=311).

Variable	Chinese sample (n=201), n (%)	Swiss sample (n=110 <sup>a</sup> ), n (%)
<b>Gender</b>		
Male	89 (44.3)	52 (47.3)
Female	112 (55.7)	58 (52.7)
<b>Age (years)</b>		
16-25	8 (4.0)	10 (9.1)
26-40	72 (35.8)	33 (30.0)
41-55	56 (27.9)	33 (30.0)
56-70	38 (18.9)	29 (26.4)
>70	27 (13.4)	4 (3.6)
<b>Monthly income (US \$)</b>		
<500	16 (8.0)	2 (1.8)
501-1500	100 (49.8)	4 (3.6)
1501-3000	36 (17.9)	7 (6.4)
3001-5000	20 (10.0)	19 (17.3)
>5001	7 (3.5)	60 (54.5)
No information	22 (10.9)	17 (15.5)
<b>Highest education level</b>		
Apprenticeship	10 (5.0)	30 (27.3)
Senior high school	12 (6.0)	5 (4.5)
College	27 (13.4)	25 (22.7)
University <sup>b</sup> and above	143 (71.1)	43 (39.1)
No information	9 (4.5)	6 (5.5)

<sup>a</sup>One respondent did not answer the questions related to age, monthly income, and education, respectively, in the Swiss sample.

<sup>b</sup>Including universities of applied sciences.

### Difference in Cultural Values of the Chinese and Swiss

Following the 6 cultural dimensions of Hofstede [28], the biggest differentiation between the Chinese and the Swiss exists in the dimension of individualism versus collectivism (opposite of individualism), which was empirically examined in this study. Individualism is the degree to which people in a society are integrated into groups [31]. In a culture with individualistic values (like Switzerland), the ties between individuals are loose; that is, everyone is expected to look after themselves and their immediate family [31]. In cultures with a collectivistic value (like China), people are integrated from birth onward into strong, cohesive in-groups, often extended families that continue protecting them in exchange for unquestioning loyalty [31]. The issue addressed by this dimension is an extremely

fundamental one, relevant to all societies in the world [28]. In this study, the question items and calculation methods were based on Hofstede and Minkov's index formula [25]. The mean value of individualism for the entire Chinese population was 4.00 as compared to that of the Swiss at 46.24. Adding a constant of 20 to the mathematical means results in a value of 24.0 for the Chinese and 66.24 for the Swiss. This conforms almost exactly to Hofstede's original individualism values (20 vs 68) in [Table 1](#), which empirically confirms the distinguished cultural differences between China (low individualism) and Switzerland (high individualism).

### Validity Assessment

[Table 5](#) depicts the mean values and standard deviations of the constructs shown in [Figure 1](#). This table also includes the

reliability and validity statistics. The loadings of all reflective indicators were above .70 and significant, confirming item reliability. In line with this, the values of the composite reliability estimates revealed the internal consistency of the measurement instruments. The reliability estimates for behavioral intention to adopt health care wearables can also be regarded as satisfactory, because Hair et al [40] indicated that the true internal consistency reliability values usually lie between Cronbach  $\alpha$  and the composite reliability. In addition,

according to the values of the average variance extracted (AVE) statistic, the measurement showed convergent validity.

Based on the quotient of the square root of AVE on the diagonal and the correlation coefficient between constructs in the lines below, according to the Fornell-Larcker criterion [41], the square roots of AVE are larger than the correlation coefficient between constructs; thus, the discriminant validity of the measurement can ultimately be confirmed.

**Table 5.** Descriptive statistics, reliability statistics, and validity statistics.

Variable	Mean (SD)	Cronbach $\alpha$	Composite reliability	AVE <sup>a</sup>
Performance Expectancy	3.573 (0.806)	.869	0.919	0.791
Hedonic Motivation	3.438 (0.763)	.862	0.915	0.782
Functional Congruence	3.445 (0.678)	.713	0.836	0.631
Effort Expectancy	3.766 (0.803)	.900	0.937	0.833
Social Influence	3.034 (0.990)	.910	0.943	0.848
Health Consciousness	4.028 (0.577)	.806	0.803	0.590
Perceived Privacy Risk	3.393 (0.891)	.848	0.886	0.724
Behavioral Intention	3.129 (1.086)	.927	0.954	0.873

<sup>a</sup>AVE: average variance extracted.

### Differences Between Swiss and Chinese Participants in the Constructs

A *t* test was conducted to compare mean differences between the Chinese and Swiss samples in the constructs of the conceptual model.

As indicated in Table 6, the different responses of Chinese and Swiss respondents toward the variables performance expectancy, functional congruence, effort expectancy, social influence, behavioral intention, and the cultural value of individualism were significant. This indicates that Chinese respondents have higher performance expectancy on the presented wearables than Swiss respondents and are influenced more by their social environment. The moderate significant difference on effort

expectancy and functional congruence confirms that the Chinese consider health care wearables easy to use, and they pay more attention to additional functions such as comfort, esthetics, and price value. Differences between Chinese consumers and Swiss consumers toward hedonistic motivation, health consciousness, and perceived privacy risk were not significant. Both groups have quite high health consciousness, perceived privacy risk, and relatively high hedonistic motivation. From the significant differences between Chinese consumers and Swiss consumers in behavioral intention, it can be concluded that the Chinese clearly have more intention to use health care wearables than the Swiss. All of these distinctions are related to the cultural differences between Swiss and Chinese consumers according to cultural value dimensions.

**Table 6.** Comparing perceptions of the Chinese and Swiss (*t* test).

Variable	Chinese sample, mean (SD)	Swiss sample, mean (SD)	Mean difference	<i>P</i> value
Performance Expectancy	3.74 (0.76)	3.26 (0.79)	0.484	<.001
Hedonic Motivation	3.47 (0.73)	3.38 (0.82)	0.097	.29
Functional Congruence	3.50 (0.70)	3.34 (0.62)	0.168	.03
Effort Expectancy	3.84 (0.77)	3.63 (0.84)	0.215	.03
Social Influence	3.44 (0.84)	2.30 (0.80)	1.136	<.001
Health Consciousness	3.99 (0.61)	4.10 (0.50)	−0.108	.10
Perceived Privacy Risk	3.40 (0.91)	3.38 (0.85)	0.018	.86
Behavioral Intention	3.59 (0.80)	2.29 (1.40)	1.301	<.001
Individualism	4.004 (0.54)	46.24 (61.53)	−42.234	<.001



## Multigroup Partial Least Squares Path Analysis

The group-specific  $R^2$  value in the Chinese sample was 0.580 and that in the Swiss sample was 0.558, which revealed good explanatory power of the delineated model regarding the behavioral intention to adopt health care wearables.

Table 7 depicts the path coefficients according to the conceptional model (Figure 1). In view of hypotheses H1a-H7a and H1b-H7b, the results showed significant effects of performance expectancy, hedonistic motivation, and social influence on the behavior intention of consumers of health care

wearables in general. Performance expectancy, social influence, and health consciousness influenced the behavior intention of the Chinese significantly, whereas performance expectancy, hedonistic motivation, effort expectancy, and social influence had a significant influence on the behavior intention of the Swiss. Group-specific path coefficients showed a stronger influence of performance expectancy on the behavior intention of the Swiss than the Chinese and a stronger influence of social influence on the behavior intention of the Chinese than the Swiss. However, the results of the multigroup analysis indicated that these differences in the path coefficients were not significant when assessed at a level of  $P < .05$ .

**Table 7.** Group-specific path coefficients for each variable's influence on behavioral intention and multigroup analysis.

Variable	Path coefficient ( $P$ value) <sup>a</sup>			$P$ value for group differences in path coefficients
	Total sample	Chinese sample	Swiss sample	
Performance Expectancy	0.361 (<.001)	0.271(<.001)	0.426 (<.001)	.17
Hedonic Motivation	0.111 (.01)	0.082	0.212 (.02)	.28
Functional Congruence	−0.062	0.142	−0.084	.09
Effort Expectancy	0.067	−0.003	0.165 (.02)	.08
Social Influence	0.475 (<.001)	0.321 (<.001)	0.217 (.004)	.31
Health Consciousness	0.005	0.150 (.01)	−0.042	.08
Perceived Privacy Risk	−0.042	−0.030	−0.015	.90

<sup>a</sup>Significance levels are based on a 5000 bootstrap run.

## Principal Results

The results above partially confirmed the proposed hypotheses. Among all of the predictors, performance expectancy, hedonistic motivation, and social influence affected behavioral intention positively, which support hypotheses H1a, H2a, and H5a, respectively. Effort expectancy, functional congruence, health consciousness, and perceived privacy risk did not affect behavior intention significantly, thereby rejecting hypotheses H3a, H4a, H6a, and H7a. Nevertheless, group-specific multigroup analysis with Chinese and Swiss samples indicated that hedonistic motivation and effort expectancy are significant predictors affecting the behavior intention of Swiss consumers positively, but do not affect that of Chinese consumers. By contrast, health consciousness is an important predictor affecting the behavior intention of Chinese consumers positively, but does not appear to have an effect on the behavior intention of Swiss consumers. Performance expectancy is a key factor affecting the behavior intention of both Chinese and Swiss consumers positively, but the degree of influence on Swiss consumers was higher than that on Chinese consumers. Social influence is another key factor affecting the behavior intention of both Chinese and Swiss consumers positively, but the degree of influence on Chinese consumers was higher than that on Swiss consumers. Country variable was not a moderator that differentiated the influence degree of functional congruence or perceived privacy risk toward behavior intention between Chinese and Swiss consumers. These results confirm hypotheses H1b, H2b, H3b, H5b, and H6b, but do not confirm hypotheses H4b and H7b.

## Discussion

### Main Findings

The  $R^2$  values in the total and subsample analyses indicate that the adapted WTAH model is a suitable conceptual model to assess behavioral intentions to use smartwatches as health care wearables. The multigroup analysis showed no significant differences between the Chinese and Swiss samples with respect to path coefficients. However, clear country-specific differences in the data were still found. First, the bootstrapping results within both samples clearly indicated that the relevance of the factors from the adapted WTAH model in the two sample groups differed. Performance expectancy and social influence appear to play a role in both sample groups to explain behavioral intention, whereas hedonistic motivation and effort expectancy were only key factors for the intention to use a smartwatch for the Swiss sample, and health consciousness only emerged as an important factor affecting behavioral intention for the Chinese sample. Functional congruence did not affect behavior intention in either sample. This could be explained by the fact that functional congruence comprises three items, wearing comfort, fashion, and price reasonability, which belong to three parallel aspects. Although intuitively all of these items relate closely with user intention, their consistency should be further checked.

Moreover, the mean values of perceived privacy risk were relatively high for both Chinese and Swiss consumers, but this was not validated as a significant predictor of the disinclination for consumers to use health care wearables, for consumers in general or for the specific group of Chinese or Swiss consumers.

One reason could be that smartwatches were used as the research object in this study, and consequently the related health data are not considered as critical and strictly confidential by users. Another reason could be that as a result of recent advancements in data privacy legislation, people are more familiar and confident with the data protection issue.

### Role of Cultural Values on Acceptability

These results can be explained by the differences in cultural values or social and health care systems between China and Switzerland, which are summarized in Table 8.

Chinese collectivist values and Swiss individualist values were confirmed empirically in this study. As collectivists, Chinese consumers search for information and support on wearables from people around them, attach more importance to others' opinions, and adapt to their peers. This explains the significantly higher values of effort expectancy, functional congruence, and social influence for the Chinese consumers in the *t* test. Chinese consumers, holding low uncertainty avoidance and "harmony" values toward their surroundings, normally embrace new technology and believe in its effectiveness, which can explain their higher performance expectancy value toward wearables than that of Swiss consumers.

**Table 8.** Influential cultural values on differences between China (CN) and Switzerland (CH).

Variables	Perceptions comparison ( <i>t</i> test)	Influence degree of country (group-specific path coefficients)	Explanatory cultural dimensions/social systems
Performance expectancy	CN > CH	CN < CH	CN: low IDV <sup>a</sup> ; low UAI <sup>b</sup> ; "Harmony" CH: high IDV; high UAI; "Mastery"
Hedonistic motivation	No difference	CN ns <sup>c</sup> CH sig <sup>d</sup>	CN: low IDV; low IND <sup>e</sup> CH: high IDV; high IND
Effort expectancy	CN > CH	CN ns CH sig	CN: low IDV; low UAI CH: high IDV; high UAI
Functional congruence	CN > CH	ns	CN: low IDV; low income CH: high IDV; high income
Social influence	CN > CH	CN > CH	CN: low IDV; low UAI; high PDI <sup>f</sup> CH: high IDV; high UAI; low PDI
Health consciousness	No difference	CN sig CH ns	CN: lack of developed health care and insurance system CH: importance on sport activities
Perceived privacy risk	No difference	CN ns CH ns	N/A <sup>g</sup>
Behavioral intention	CN > CH	N/A	CN: low IDV; low UAI; "Harmony" CH: high IDV; high UAI; "Mastery"

<sup>a</sup>IDV: individualism.

<sup>b</sup>UAI: uncertainty avoidance.

<sup>c</sup>ns: not significant.

<sup>d</sup>sig: significant.

<sup>e</sup>IND: indulgence.

<sup>f</sup>PDI: power distance.

<sup>g</sup>N/A: not applicable.

Swiss consumers showed significantly lower behavioral intention than Chinese consumers, which can be explained by their high value of uncertainty avoidance and "mastery" relationship with their surroundings (as compared to those of Chinese consumers). Because using wearables reduces social presence, this could accentuate the feeling of uncertainty. As a novel technology, the side effect, functionality, and measurement accuracy are quite uncertain. Swiss consumers in high uncertainty avoidance cultures will be less oriented toward using wearables than Chinese consumers in low uncertainty avoidance cultures [24]. Furthermore, the "mastery" value of Swiss consumers toward their environment/surroundings makes them inclined to stick with their habits and perceived correctness, which prevents them from trying new devices.

Although the mean values of health consciousness were high for both Chinese and Swiss respondents, it was not validated as a predictor that influences consumer intention to use wearables generally. Nevertheless, through group-specific analysis, health consciousness emerged as a significant predictor for behavior intention for the Chinese but not for the Swiss. The reason probably lies in the fact that Swiss consumers already spend more time on sport or wellness activities to enhance their health, and they generally have no further interest or time to invest in studying wearables for health purposes [7]. This is supported by Seiler and Hüttermann [11], who showed that 51% of nonusers (74% of the full sample) in their study expressed no need to adopt fitness wearables. By contrast, health-conscious Chinese consumers prefer to use wearables to help them track

their health, which is likely related to the underdeveloped health system in China with limited health care resources for such a large population; thus, Chinese consumers are open to more diverse and preventive possibilities.

### Limitations

Although some hypotheses were validated, and the criteria of objectivity, reliability, and validity were followed during the whole process of research, this study is subject to certain limitations, and the results should be interpreted with caution.

First, some variables of the conceptual model should be reorganized and reconsidered in future research. For example, health consciousness relates to many different concepts from health awareness, ranging from health concern to health activities. It influences the user's perception and intention of using wearables multilaterally. Health consciousness might be one of the predetermining factors for other variables such as performance expectancy, effort expectancy, and others, as investigated by some studies with controversial results [7,8]. Therefore, health consciousness should be examined in the future in different model structures. The country variable based on national culture was examined in this study as a moderating variable, which affects the relationship between predictors and outcomes. Cultural values influence many aspects (perception, attitude, intention) further to action. Thus, the country/cultural variables might also be predetermining factors for other predictors. In further research, the roles and distinction of cultural variables should be considered carefully and examined in different model structures as well.

Second, the control variables such as gender, age, education, and income might influence people's response patterns [42], which means that they can interfere with the predictors to affect a user's intention to adopt wearables. The control variables and user experiences were not considered in this study. The effect of control variables as moderators alone or as covariates with other moderators such as country/culture should be analyzed in the future.

Third, the external validity of the study is limited. A smartwatch was used as an example of health care wearables, which cannot be generalized to all types of medical wearable devices. In addition, the results cannot be generalized to other countries, as the survey was conducted in only China and Switzerland, and the cultural dimensions were only used to explain the difference between Chinese and Swiss consumers theoretically. It makes more sense to apply cultural values directly in the quantitative analysis in future research. For this purpose, effective methods of obtaining qualified scores of cultural values must be further explored.

Finally, a quantitative approach was applied to examine the conceptual model of this study, which could not provide specific information on certain types of users or devices. In future research, expert or focus group interviews could be conducted regarding a certain type of medical-grade wearable to gain more specific information.

### Comparison With Prior Work

This study is among the first to investigate the influential factors on intention to use health care wearables involving samples from two countries with quite different national cultures. Although most items of the conceptual model were adapted from the framework of Gao et al [6], in this study, the three predictors (self-efficacy, perceived vulnerability, and perceived severity) based on the PMT were replaced by a single variable: health consciousness. Thus, the findings cannot be compared with other studies directly.

Nevertheless, this study verifies that performance expectancy and social influence are the most influential factors on people's intention to accept health care wearables, which is in line with the results of Gao et al [6]. However, perceived privacy risk was not validated as a significant predictor influencing the behavior intention of consumers negatively, which contrasts with the results of Gao et al [6]. The main reason for this difference could be that more users adopting medical-grade devices were included in the samples of Gao et al's research because fitness/medical devices were examined as a moderator. This is not the case in our study regarding Chinese consumers, with smartwatches used as an example of health care wearables. Most smartwatch users might not be suffering from severe health problems at this stage. Hence, Chinese respondents might not consider revealing their health information as a privacy risk that would prevent them from using wearables in health care.

### Conclusions

This study examined the factors influencing people's behavioral intention to accept health care wearables, explored the different perceptions and using intention of Chinese and Swiss consumers, and explained the effect of national culture on these differences. These findings have practical implications for global wearables vendors and insurers to develop and promote health care wearables for consumers from various cultural backgrounds.

Performance expectancy and social influence were the most significant predictors that positively influenced consumer intention to adopt health care wearables. This result indicates that consumers are more affected by the perceived effectiveness of health care wearables and by other people's opinions. Thus, these factors should be given more attention when global companies develop and market health care wearables. For Switzerland, with cultural values of individualism and high uncertainty avoidance, the positive opinions of professionals such as physicians toward wearables and clearly demonstrated measurement accuracy would make Swiss consumers feel more confident with health care wearables. For China, with cultural values of collectivism and high power distance, the opinions of people in their surroundings (eg, peers, family, and friends) and the engagement of the Chinese working unit (eg, employers' social benefit) would increase the intention of Chinese consumers to adopt wearables.

Effort expectancy was shown to be an important driver for Swiss consumers to adopt health care wearables. Thus, the wearable devices should be easy to handle to attract Swiss consumers. This includes easy wearing of devices, along with easy analysis and interpretation of the data. Health consciousness emerged

as an important driver for Chinese consumers in adopting health care wearables. Thus, multifunctional apps providing feasible health care advice and solutions in cooperation with Chinese health care institutions are essential to attract Chinese consumers.

In addition, this study is one of the first to investigate intentions to adopt health care wearables from a cross-cultural perspective. It thus provides a theoretical foundation in terms of a conceptual model and survey methodology for future research in similar contexts with other countries and cultures.

## Acknowledgments

We would like to thank Yongsheng Yang, Beatrice Meyer, and Mischa Rupp for supplying personal information on using health care wearables in interviews, which served as the credible basis for questionnaire development. We would like to thank Altar Yilmazer, Elisabeth Schönwald, and Michaela Settelen for reviewing the language in the German questionnaire, and Zhiying Leng and Xu Yang for reviewing the language in the Chinese questionnaire. We would like to thank the many friends and colleagues who distributed the questionnaire in their social networks.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Measurement items.

[[XLSX File \(Microsoft Excel File\), 23 KB - jmir\\_v22i10e18801\\_app1.xlsx](#)]

## References

1. Framingham M. IDC Reports Strong Growth in the Worldwide Wearables Market, Led by Holiday Shipments of Smartwatches, Wrist Bands, and Ear-Worn Devices. IDC. 2019 Mar. URL: <https://www.idc.com/getdoc.jsp?containerId=prUS44901819> [accessed 2019-04-25]
2. Tuzovic K, Mathews M. Points for Fitness – How Smart Wearable Technology Transforms Loyalty Programs. In: Bruhn M, Hadwich K, editors. Dienstleistungen 4.0: Geschäftsmodelle—Wertschöpfung—Transformation Band 2. Forum Dienstleistungsmanagement (S. 445–463). Switzerland: Springer; Jul 2017:445-463.
3. Alagöz F, Ziefle M, Wilkowska W, Valdez A. Openness to accept medical technology - a cultural view. In: Information Quality in e-Health. Berlin, Heidelberg: Springer; 2011 Presented at: 7th Conference of the Workshop Human Computer Interaction; November 2011; Graz, Austria p. 151-170. [doi: [10.1007/978-3-642-25364-5\\_14](https://doi.org/10.1007/978-3-642-25364-5_14)]
4. Sun Y, Wang N, Guo X, Peng Z. Understanding the acceptance of mobile health services: a comparison and integration of alternative models. J Electron Commer Res 2012 Dec;14(2):183-200.
5. Li H, Wu J, Gao Y, Shi Y. Examining individuals' adoption of healthcare wearable devices: An empirical study from privacy calculus perspective. Int J Med Inform 2016 Apr;88:8-17. [doi: [10.1016/j.ijmedinf.2015.12.010](https://doi.org/10.1016/j.ijmedinf.2015.12.010)] [Medline: [26878757](https://pubmed.ncbi.nlm.nih.gov/26878757/)]
6. Gao Y, Li H, Luo Y. An empirical study of wearable technology acceptance in healthcare. Industr Mngmnt Data Syst 2015 Oct 19;115(9):1704-1723. [doi: [10.1108/IMDS-03-2015-0087](https://doi.org/10.1108/IMDS-03-2015-0087)]
7. Cho J, Quinlan MM, Park D, Noh G. Determinants of adoption of smartphone health apps among college students. Am J Health Behav 2014 Nov;38(6):860-870. [doi: [10.5993/AJHB.38.6.8](https://doi.org/10.5993/AJHB.38.6.8)] [Medline: [25207512](https://pubmed.ncbi.nlm.nih.gov/25207512/)]
8. Chen M, Lin N. Incorporation of health consciousness into the technology readiness and acceptance model to predict app download and usage intentions. Internet Res 2018 Apr 04;28(2):351-373. [doi: [10.1108/intr-03-2017-0099](https://doi.org/10.1108/intr-03-2017-0099)]
9. Reeder B, David A. Health at hand: A systematic review of smart watch uses for health and wellness. J Biomed Inform 2016 Oct;63:269-276 [FREE Full text] [doi: [10.1016/j.jbi.2016.09.001](https://doi.org/10.1016/j.jbi.2016.09.001)] [Medline: [27612974](https://pubmed.ncbi.nlm.nih.gov/27612974/)]
10. Seneviratne S, Hu Y, Nguyen T, Lan G, Khalifa S, Thilakarathna K, et al. A Survey of Wearable Devices and Challenges. IEEE Commun Surv Tutor 2017;19(4):2573-2620. [doi: [10.1109/comst.2017.2731979](https://doi.org/10.1109/comst.2017.2731979)]
11. Seiler R, Hüttermann M. E-Health, fitness trackers and wearables: use among Swiss students. 2015 Mar 27 Presented at: ABSRC Advances in Business-Related Scientific Research Conference; March 25-27, 2015; Venice, Italy.
12. Wen D, Zhang X, Lei J. Consumers' perceived attitudes to wearable devices in health monitoring in China: A survey study. Comput Methods Programs Biomed 2017 Mar;140:131-137. [doi: [10.1016/j.cmpb.2016.12.009](https://doi.org/10.1016/j.cmpb.2016.12.009)] [Medline: [28254069](https://pubmed.ncbi.nlm.nih.gov/28254069/)]
13. Alshare KA, Mesak HI, Grandon EE, Badri MA. Examining the Moderating Role of National Culture on an Extended Technology Acceptance Model. J Glob Inf Technol Manag 2014 Sep 09;14(3):27-53. [doi: [10.1080/1097198x.2011.10856542](https://doi.org/10.1080/1097198x.2011.10856542)]
14. Nardon L, Steers R. Navigating the culture theory jungle: Divergence and convergence in models of national culture. Gent: Vlerick Leuven Gent Management School; 2006. URL: [https://www.researchgate.net/publication/23646752\\_Navigating\\_the\\_culture\\_theory\\_jungle\\_divergence\\_and\\_convergence\\_in\\_models\\_of\\_national\\_culture](https://www.researchgate.net/publication/23646752_Navigating_the_culture_theory_jungle_divergence_and_convergence_in_models_of_national_culture) [accessed 2019-04-23]
15. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quart 1989 Sep;13(3):319-340. [doi: [10.2307/249008](https://doi.org/10.2307/249008)]



16. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quart* 2012;36(1):157-178. [doi: [10.2307/41410412](https://doi.org/10.2307/41410412)]
17. Rogers RW. A Protection Motivation Theory of Fear Appeals and Attitude Change. *J Psychol* 1975 Sep 02;91(1):93-114. [doi: [10.1080/00223980.1975.9915803](https://doi.org/10.1080/00223980.1975.9915803)] [Medline: [28136248](https://pubmed.ncbi.nlm.nih.gov/28136248/)]
18. Ford D, Connelly C, Meister D. Information systems research and hofstede's culture's consequences: an uneasy and incomplete partnership. *IEEE Trans Eng Manage* 2003 Feb;50(1):8-25. [doi: [10.1109/tem.2002.808265](https://doi.org/10.1109/tem.2002.808265)]
19. Awad NF, Krishnan MS. The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization. *MIS Quart* 2006;30(1):13. [doi: [10.2307/25148715](https://doi.org/10.2307/25148715)]
20. Uğur NG. Cultural Differences and Technology Acceptance: A Comparative Study. *J Media Crit* 2017 Sep 10;3(11):123-132. [doi: [10.17349/jmc117310](https://doi.org/10.17349/jmc117310)]
21. Al-Gahtani SS, Hubona GS, Wang J. Information technology (IT) in Saudi Arabia: Culture and the acceptance and use of IT. *Inf Manag* 2007 Dec;44(8):681-691. [doi: [10.1016/j.im.2007.09.002](https://doi.org/10.1016/j.im.2007.09.002)]
22. Bandyopadhyay K, Fraccastoro KA. The Effect of Culture on User Acceptance of Information Technology. *Commun Assoc Inf Syst* 2007;19:521-544. [doi: [10.17705/1cais.01923](https://doi.org/10.17705/1cais.01923)]
23. Venkatesh V, Zhang X. Unified Theory of Acceptance and Use of Technology: U.S. vs. China. *J Glob Inf Technol Manag* 2014 Sep 09;13(1):5-27. [doi: [10.1080/1097198X.2010.10856507](https://doi.org/10.1080/1097198X.2010.10856507)]
24. Zakour A. Cultural differences and information technology acceptance. *SAIS* 2004 Proc 2004 Jan 3;26:156-161 [FREE Full text]
25. Hofstede G, Minkov M. Values survey module 2013 manual. 2013 May. URL: <https://geerthofstede.com/wp-content/uploads/2016/07/Manual-VSM-2013.pdf> [accessed 2019-03-25]
26. Nakata C, Sivakumar K. Instituting the Marketing Concept in a Multinational Setting: The Role of National Culture. *J Acad Market Sci* 2001 Jul 01;29(3):255-276. [doi: [10.1177/03079459994623](https://doi.org/10.1177/03079459994623)]
27. McCoy S, Galletta DF, King WR. Integrating National Culture into IS Research: The Need for Current Individual Level Measures. *Commun Assoc Inf Syst* 2005;15:210-224. [doi: [10.17705/1cais.01512](https://doi.org/10.17705/1cais.01512)]
28. Hofstede G. Dimensionalizing Cultures: The Hofstede Model in Context. *Online Read Psychol Cult* 2011 Dec 01;2(1):8. [doi: [10.9707/2307-0919.1014](https://doi.org/10.9707/2307-0919.1014)]
29. McCoy S, Galletta DF, King WR. Applying TAM across cultures: the need for caution. *Eur J Inf Syst* 2017 Dec 19;16(1):81-90. [doi: [10.1057/palgrave.ejis.3000659](https://doi.org/10.1057/palgrave.ejis.3000659)]
30. Hofstede G. Country Comparison. Hofstede Insights. 2019. URL: <https://www.hofstede-insights.com/country-comparison/china,switzerland/> [accessed 2019-03-01]
31. Hofstede G, Hofstede GJ, Minkov M. *Cultures and Organizations - Software of the Mind: Intercultural Cooperation and Its Importance for Survival*. USA: McGraw-Hill Education Ltd; Jul 1, 2010.
32. Floyd DL, Prentice-Dunn S, Rogers RW. A Meta-Analysis of Research on Protection Motivation Theory. *J Appl Social Psychol* 2000 Feb;30(2):407-429. [doi: [10.1111/j.1559-1816.2000.tb02323.x](https://doi.org/10.1111/j.1559-1816.2000.tb02323.x)]
33. Wearable Technology Market Size, Growth, Trend and Forecast to 2022. *MarketsandMarkets*. URL: [https://www.marketsandmarkets.com/Market-Reports/wearable-electronics-market-983.html?gclid=Cj0KCQjwnKHBRLARIsAMtMHDEwcviaYpWp9HFoR-OZ\\_DHXPcTn7fYbzc7yhujCaOBfSfwdLa7iMaApXEEALw\\_wcB](https://www.marketsandmarkets.com/Market-Reports/wearable-electronics-market-983.html?gclid=Cj0KCQjwnKHBRLARIsAMtMHDEwcviaYpWp9HFoR-OZ_DHXPcTn7fYbzc7yhujCaOBfSfwdLa7iMaApXEEALw_wcB) [accessed 2019-04-09]
34. Choi J, Lee HJ, Sajjad F, Lee H. The influence of national culture on the attitude towards mobile recommender systems. *Technol Forecast Soc Change* 2014 Jul;86:65-79. [doi: [10.1016/j.techfore.2013.08.012](https://doi.org/10.1016/j.techfore.2013.08.012)]
35. Michaelidou N, Hassan LM. The role of health consciousness, food safety concern and ethical identity on attitudes and intentions towards organic food. *Int J Cons Stud* 2008 Mar;32(2):163-170. [doi: [10.1111/j.1470-6431.2007.00619.x](https://doi.org/10.1111/j.1470-6431.2007.00619.x)]
36. Nield D. 5 ways smartwatches could break out as medical devices. *Wearable*. URL: <https://www.wearable.com/smartwatches/ways-break-out-medical-devices-6463> [accessed 2019-03-09]
37. Picard R. An AI smartwatch that detects seizures. TED Talk. 2019. URL: [https://www.ted.com/talks/rosalind\\_picard\\_an\\_ai\\_smartwatch\\_that\\_detects\\_seizures?language=en](https://www.ted.com/talks/rosalind_picard_an_ai_smartwatch_that_detects_seizures?language=en) [accessed 2019-05-05]
38. Van der Meulen R, Forni A. Gartner Says Worldwide Wearable Device Sales to Grow 17 Percent in 2017. *Gartner*. 2017 Aug 24. URL: <https://www.gartner.com/en/newsroom/press-releases/2017-08-24-gartner-says-worldwide-wearable-device-sales-to-grow-17-percent-in-2017> [accessed 2019-04-14]
39. Tansey O. Process Tracing and Elite Interviewing: A Case for Non-probability Sampling. *PS Polit Sci Polit* 2007 Oct 3;40(04):765-772. [doi: [10.1017/s1049096507071211](https://doi.org/10.1017/s1049096507071211)]
40. Hair, Jr. JF, Sarstedt M, Matthews LM, Ringle CM. Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I – method. *Eur Bus Rev* 2016 Jan 11;28(1):63-76. [doi: [10.1108/ebv-09-2015-0094](https://doi.org/10.1108/ebv-09-2015-0094)]
41. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J Market Res* 1981 Feb;18(1):39. [doi: [10.2307/3151312](https://doi.org/10.2307/3151312)]
42. van Thiel S. *Research methods in public administration and public management*. London: Routledge; 2014.



## Abbreviations

**AVE:** average variance extracted

**PMT:** protection motivation theory

**TAM:** technology acceptance model

**UTAUT:** unified theory of acceptance and use of technology

**UTAUT2:** unified theory of acceptance and use of technology 2

**WTAH:** wearable technology acceptance in health care

*Edited by G Eysenbach; submitted 19.03.20; peer-reviewed by R Seiler, J Lei, Z Su; comments to author 12.06.20; revised version received 27.07.20; accepted 29.07.20; published 22.10.20.*

*Please cite as:*

*Yang Meier D, Barthelmess P, Sun W, Liberatore F*

*Wearable Technology Acceptance in Health Care Based on National Culture Differences: Cross-Country Analysis Between Chinese and Swiss Consumers*

*J Med Internet Res 2020;22(10):e18801*

URL: <http://www.jmir.org/2020/10/e18801/>

doi: [10.2196/18801](https://doi.org/10.2196/18801)

PMID: [33090108](https://pubmed.ncbi.nlm.nih.gov/33090108/)

©Dong Yang Meier, Petra Barthelmess, Wei Sun, Florian Liberatore. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Brain Tumor Discussions on Twitter (#BTSM): Social Network Analysis

Josemari T Feliciano<sup>1</sup>, MPH; Liz Salmi<sup>2</sup>, AS; Charlie Blotner<sup>3</sup>, MSW; Adam Hayden<sup>4</sup>, MA; Edjah K Nduom<sup>5</sup>, MD; Bethany M Kwan<sup>6</sup>, PhD, MSPH; Matthew S Katz<sup>7</sup>, MD; Elizabeth B Claus<sup>1,8</sup>, MD, PhD

<sup>1</sup>Department of Biostatistics, Yale University School of Public Health, New Haven, CT, United States

<sup>2</sup>OpenNotes, Beth Israel Deaconess Medical Center, Boston, MA, United States

<sup>3</sup>Social Work Hospice Care Team, Evergreen Health Medical Center, Kirkland, WA, United States

<sup>4</sup>Department of Philosophy, Indiana University - Purdue University Indianapolis, Indianapolis, IN, United States

<sup>5</sup>Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, United States

<sup>6</sup>Department of Family Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

<sup>7</sup>Department of Radiation Medicine, Lowell General Hospital, Lowell, MA, United States

<sup>8</sup>Department of Neurosurgery, Brigham and Women's Hospital, Boston, MA, United States

**Corresponding Author:**

Elizabeth B Claus, MD, PhD

Department of Biostatistics

Yale University School of Public Health

60 College St.

P.O. Box 208034

New Haven, CT, 06520

United States

Phone: 1 203 785 6415

Email: [elizabeth.claus@yale.edu](mailto:elizabeth.claus@yale.edu)

## Abstract

**Background:** The Brain Tumor Social Media (#BTSM) Twitter hashtag was founded in February 2012 as a disease-specific hashtag for patients with brain tumor.

**Objective:** To understand #BTSM's role as a patient support system, we describe user descriptors, growth, interaction, and content sharing.

**Methods:** We analyzed all tweets containing #BTSM from 2012 to 2018 using the Symplur Signals platform to obtain data and to describe Symplur-defined user categories, tweet content, and trends in use over time. We created a network plot with all publicly available retweets involving #BTSM in 2018 to visualize key stakeholders and their connections to other users.

**Results:** From 2012 to 2018, 59,764 unique users participated in #BTSM, amassing 298,904 tweets. The yearly volume of #BTSM tweets increased by 264.57% from 16,394 in 2012 to 43,373 in 2018 with #BTSM constantly trending in the top 15 list of disease hashtags, as well the top 15 list of tweet chats. Patient advocates generated the most #BTSM tweets (33.13%), while advocacy groups, caregivers, doctors, and researchers generated 7.01%, 4.63%, 3.86%, and 3.37%, respectively. Physician use, although still low, has increased over time. The 2018 network plot of retweets including #BTSM identifies a number of key stakeholders from the patient advocate, patient organization, and medical researcher domains and reveals the extent of their reach to other users.

**Conclusions:** From its start in 2012, #BTSM has grown exponentially over time. We believe its growth suggests its potential as a global source of brain tumor information on Twitter for patients, advocates, patient organizations as well as health care professionals and researchers.

(*J Med Internet Res* 2020;22(10):e22005) doi:[10.2196/22005](https://doi.org/10.2196/22005)

**KEYWORDS**

brain tumors; social media; health care; patient support; network analysis

## Introduction

Social media acts as one of the greatest facilitators of information, ideas, and discussions by creating a borderless global platform. As of April 2019, there are an estimated 3.5 billion active social media users who connect with people around the world at near-instantaneous speed [1]. Three of the largest global platforms are Facebook, WeChat, and Twitter with active user counts of 2.320 billion, 1.098 billion, and 330 million, respectively [2]. Given the ubiquity and integration of social media into everyday life, patients have taken to these platforms to share, connect, collaborate, communicate, and self-create online support communities, making it possible to circumvent traditional barriers to support groups (eg, geography, and lack of time, awareness, and transportation) [3,4].

Twitter makes it possible for unique Twitter users to follow each other, create unique posts, and interact through *tweets*—short messages limited to 280 characters [5]. Tweets can be accompanied by hashtags, which are words preceded by the octothorpe (#) symbol [6]. On Twitter, hashtags become a link that, when clicked, displays other tweets aggregated by the same topic [5,6]. Twitter describes hashtags as a way of categorizing tweets relevant to user's interest [6]. Recent studies have reported that patients increasingly use disease-specific hashtags as ad hoc support communities to share information and create meaningful connections [7,8]. In more recent years, disease-specific hashtags (including #BTSM [Brain Tumor Social Media hashtag]) have made it possible for clinicians and researchers to connect and communicate with patients and caregivers on areas of shared interest [8,9].

The #BTSM was founded in February 2011 by authors LS and CB [9]—2 brain tumor patients inspired by the success of other Twitter hashtag communities in oncology, notably Breast Cancer Social Media (#BCSM) [10]. LS and CB sought to make it easier for people in the brain tumor community to find each other and connect via Twitter, so they created #BTSM. In a blog post in 2013, LS described #BTSM as a *patient-run* Twitter community *not owned by any organization* [11].

LS and CB started a monthly tweet chat (an organized and moderated, live conversation on Twitter) operating under the Twitter account @BTSMchat in 2013. The #BTSM chat occurs on the first Sunday of every month and discusses broad topics of interest to the brain tumor community (eg, advance care planning, clinical trials, brain tumor representation in arts and entertainment) [11,12]. As of July 2020, @BTSMchat is formally organized by 5 patients, 1 caregiver, and 1 neuro-oncology clinician.

In this study, we examine #BTSM as a hashtag, and its role as an aggregator and technical facilitator of brain tumor conversations by identifying its key users and their interactions, and the type of information shared within this community since 2012. Furthermore, we report recent trends and developments within #BTSM.

## Methods

The data are all publicly available tweets with the #BTSM hashtag from January 1, 2012, to December 31, 2018, obtained via the Symplur Signals platform [13]. Information available includes user names and usage, tweet transcripts, and usage trends over time. Using R (version 3.5.3) [14], we generated descriptive statistics and visual output to provide insights into the overall use (growth) and demographic breakdown (eg, patient, doctor, researcher) of #BTSM users, along with trends and patterns about interactions between users of the #BTSM hashtag. We highlight #BTSM user engagement by describing the proportions of tweets that were retweets, replies, or mentions (ie, tweets that contain another Twitter user's name). To account for *passive consumption* of tweets featuring the #BTSM hashtag (ie, simply viewing a tweet without engaging with a retweet, reply, or mention), we report total impressions, which is the number of times a tweet has appeared in any user timeline or search result on Twitter [15]. Aggregate data on user engagement were collected to highlight the yearly share of tweet authorship among Symplur-defined user categories. To create a stack plot that focuses on proportions, the yearly raw count for user-specific tweets was transformed to represent yearly share. By representing count as a proportion, the users who generate the most tweets at any given year are presented.

To visualize the interactions between key #BTSM stakeholders, a network plot was created to visualize content sharing via retweets in 2018. Tweet transcripts from Symplur were first downloaded and merged to ensure that all nodes can be plotted and represented in the network. To create the retweet network plot, the open source software Gephi [16] (version 0.9.2) was used. To highlight the users with the most unique connections to other users based on having been retweeted, the color and size of their nodes were adjusted to visualize influential users by applying the outdegree setting on Gephi. Outdegree ranks nodes based on unique outward connections with other nodes [17], which in effect assigned bluer and larger nodes to users with more unique connections. Minor positioning adjustments were made to the network to ensure the readability of nodes and their labels. Additionally, we examined the impact of other hashtags on #BTSM activity in 2018.

## Results

### Content and User Descriptors

From 2012 to 2018, 59,764 unique Twitter users used the #BTSM hashtag in their tweets, amassing a total of 298,904 tweets (Table 1). Among these, 78.17% (n=233,658), 53.49% (n=159,884), and 5.19% (n=15,516) involved mentions, retweets, and replies, respectively. Many tweets included media, such as website links (42.00%, n=125,534) and media objects such as photos and videos (27.66%, n=82,679). The yearly volume of #BTSM tweets increased by 264.57% from 16,394 in 2012 to 43,373 in 2018, making #BTSM one of the most heavily trended medical hashtags in the Symplur platform. Those identified as patients generated the most #BTSM tweets (33.1%), followed by brain tumor advocacy organizations (28.8%), researchers (7.0%), caregivers (4.6%), and doctors

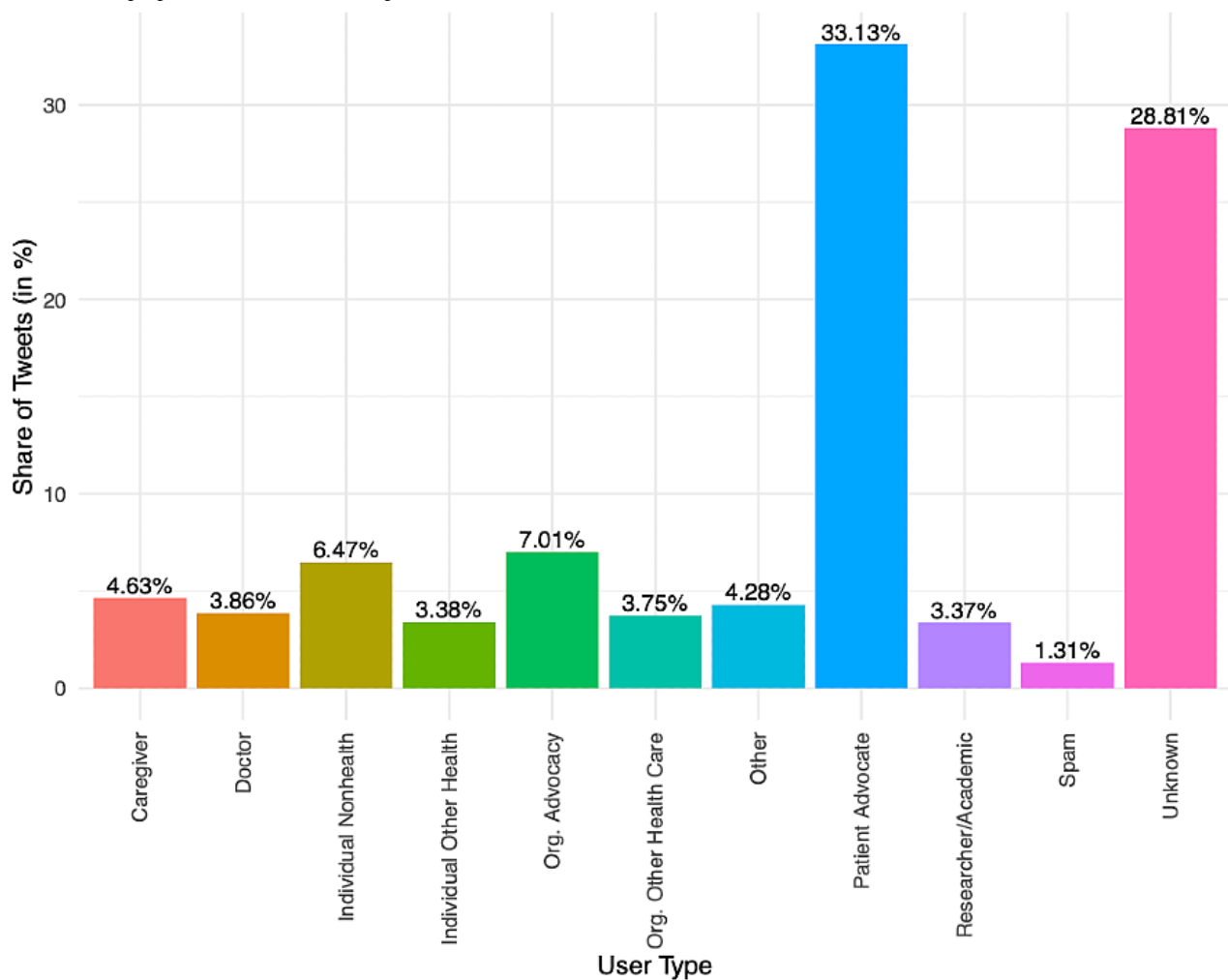
(3.9%; [Figure 1](#)). Although still low, the yearly use of #BTSM by doctors has increased over time from 1.3% to 8.6% between

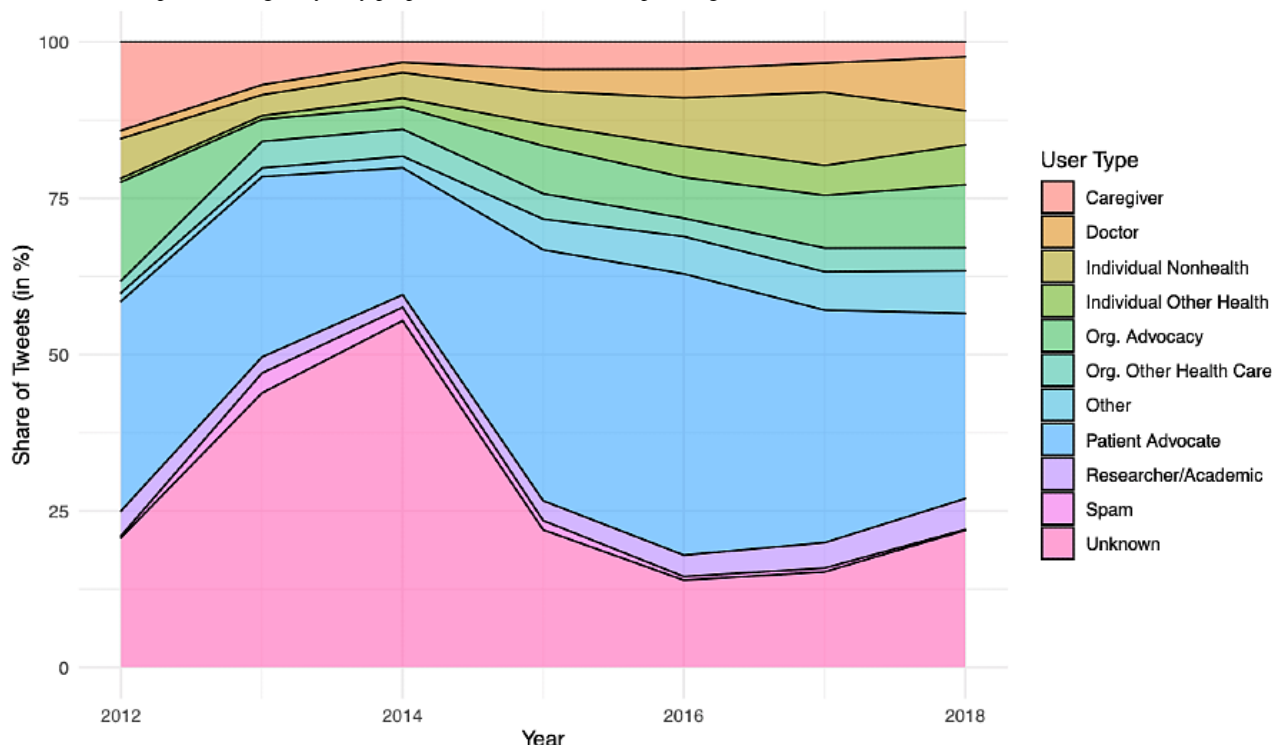
2012 and 2018 ([Figure 2](#)). Persons defined as caregivers showed reduced use over time, declining from 14.2% to 2.4%.

**Table 1.** Key #BTSM user and tweet descriptors for 2012-2018.

Data descriptor	Total count
Unique users	59,764
<b>Tweets generated</b>	298,904
Tweets with mentions	233,658
Tweets that were retweets	159,884
Tweets with links	125,534
Tweets with media (photo, video)	82,679
Tweets that were replies	15,516
Total impression	2,284,771,009

**Figure 1.** Overall proportion of tweet authorship for 2012-2018.



**Figure 2.** A stacked plot showing the yearly proportion of tweet authorship among users for 2012-2018.

### Hashtags for Expanded Reach

We conducted a deeper analysis on #BTSM user participation and growth for the 2018 calendar year. May 2018 data revealed an increase in #BTSM activity generated by tweets combining both #BTSM and #BTAM, a hashtag created by the National Brain Tumor Society (NBTS) to celebrate the Brain Tumor Awareness Month (BTAM) [18]. Out of the 4620 #BTSM tweets in May 2018, roughly 44% contained #BTAM. Other hashtags combined with #BTSM in May 2018 included #BrainTumor, #BrainTumorThursday, #BrainTumorAwarenessMonth, and #Head2Hill2018 (a public policy advocacy day organized by the NBTS) [19].

In the second half of 2018, our exploratory data analysis found #SNO2018 linked to a mid-November surge in #BTSM activity. The annual meeting of the Society of Neuro-Oncology (SNO) was held in New Orleans, Louisiana, from November 15 to 18, 2018 [20]. The meeting's official hashtag, #SNO2018, was promoted and used by conference attendees to promote and share information and research at the event. Of the 140 #BTSM

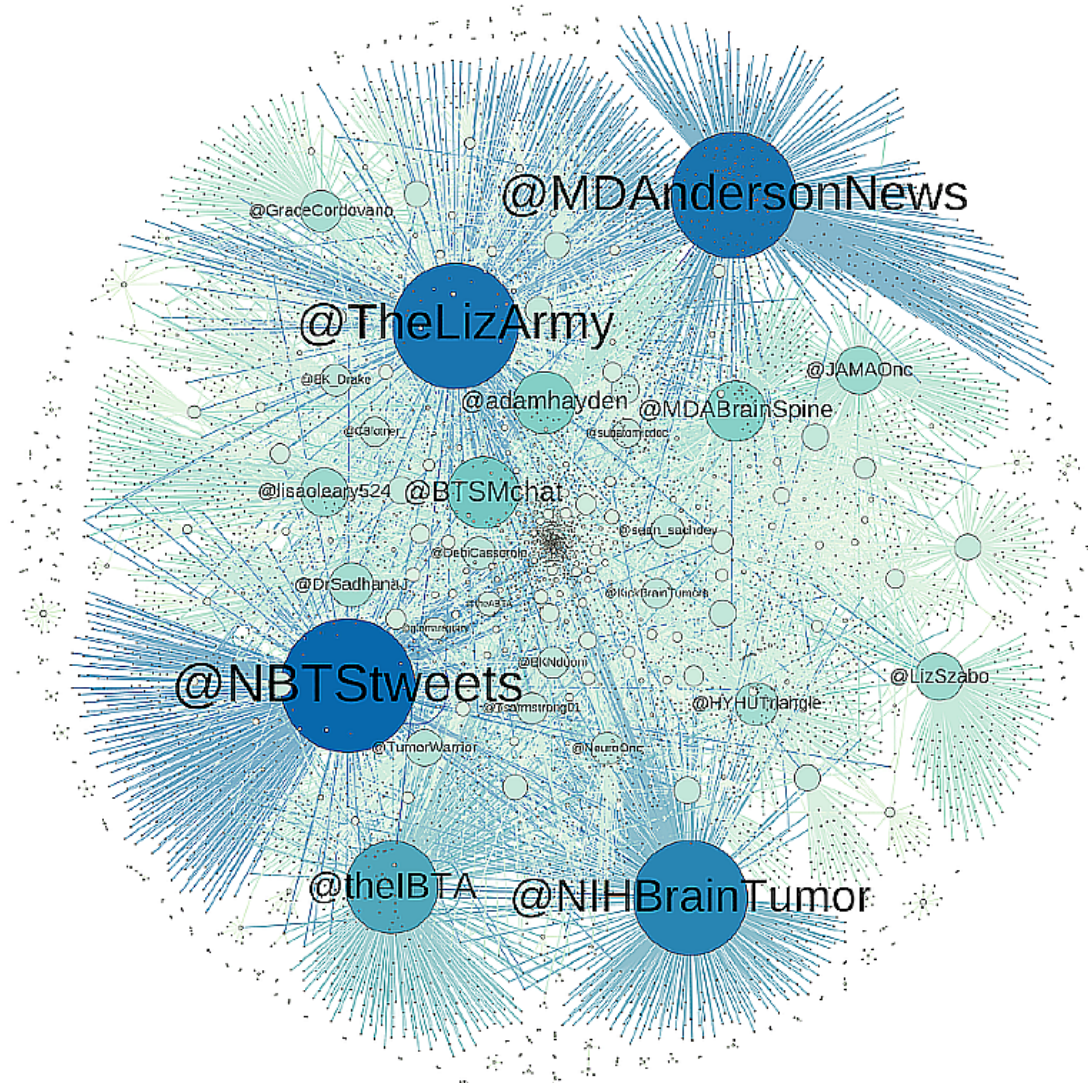
tweets in November 15-18, 85 (roughly 61%) also contained #SNO2018, which demonstrates the potential for neuro-oncology professionals to reach a majority of patient and caregiver audience on Twitter.

### Network Influencers

The 2018 network plot of retweets including #BTSM (Figure 3) identifies a number of key stakeholders from the patient advocate, patient organization, and medical researcher domains and reveals the extent of their reach to other people on Twitter. As indicated by their size and color, 6 of the most connected users in 2018 were @NBTStweets and @theIBTA (patient advocacy organizations), @MDAndersonNews and @NIHBrainTumor (health systems/brain tumor centers), @TheLizArmy (a brain tumor patient), and @BTSMchat (the patient-led community that hosts the monthly #BTSM tweet chats). Further analysis of the 2018 transcripts indicates that a substantial amount of Twitter interactions occurred in the first week of each month, corresponding to the monthly chat hosted by @BTSMchat (Table 2), and includes national and international individuals and organizations.



**Figure 3.** Social networks of most connected #BTSM users in 2018, by retweets.

**Table 2.** Topics Covered by @BTSMchat in 2018.

Month	Topic covered
January	Advance Care Planning
February	N/A <sup>a</sup>
March	Low-Grade Glioma Registry
April	What Does “Quality of Life” Mean to You?
May	Brain Tumor Awareness Month
June	Patient-Driven Symptom Tracking
July	Note to Self: What I Wish I Knew
August	Words Matter: Guilt & the Things We Tell Ourselves
September	Brain Tumors and the Media
October	Patient/Care Partner Relationships
November	Talking About Illness with Family and Friends
December	Breaking Down the Barriers to Clinical Trials

<sup>a</sup>@BTSMchat in February was on break.

## Discussion

This study finds that tweets including the #BTSM hashtag between 2012 and 2018 contained high degrees of Twitter user-to-user engagement where the vast majority of the tweets mentioned at least one other user on Twitter (eg, a reply to a tweet). Although we did not analyze the quality of information shared by users, the study found that sharing content such as website links and media (ie, sharing information) versus text alone corresponds to increased engagement with #BTSM tweets, with 27.66% and 42.00% of the tweets containing visual media (photos/videos) and website links, respectively. Twitter users do not just create unique messages, but also engage with other Twitter users by way of the #BTSM hashtag. As demonstrated in the network plot, major health care organizations (eg, MD Anderson, the National Institutes of Health [NIH], the International Brain Tumour Alliance) now play a central role in #BTSM communication by including the hashtag in their tweets. For future studies, different ways to weigh volume and unique connections, along with user clustering across the network, should be considered to give a more nuanced depiction of the network involved.

A limitation to this study is the focus on metrics that captures participation within #BTSM based on content generation and interaction, which ignores passive Twitter users who might otherwise rely on #BTSM for information without actively engaging through likes and retweets. Twitter users might be searching #BTSM for information without instigating any sort of active and thus trackable method of engagement (eg, tweeting, retweeting, replying, or liking #BTSM tweets). The @BTSMchat community organizers have attempted to track passive viewers during live tweet chats by prompting self-identification through this statement, "If you are here tonight to just listen, please tweet '#BTSM' so we know you're in the audience" [21]. To the best of our knowledge, we are not aware of any study that has thoroughly examined the extent of passive content consumption within #BTSM or any other disease-specific hashtags. Nevertheless, interaction with the #BTSM hashtag is captured by the total impression count of 2,284,771,009, the tally that any user has been served a content in their timeline or search results.

Throughout the year, major brain tumor events and medical conferences play valuable roles in expanding #BTSM's reach by introducing the brain tumor community on Twitter to the hashtag. In 2018, attendees and supporters of both BTAM and the Annual Meeting of the SNO heavily utilized #BTSM in their tweets along with other branded hashtags such as #BTAM and #SNO2018. Although it is unclear whether new Twitter users were converted into active #BTSM participants, tweets from any casual user may still be impactful for the awareness they bring to the larger brain tumor community.

One potential limitation for #BTSM growth is the somewhat older age noted for many persons diagnosed with brain tumors. For 2012-2016, the median age at diagnosis for brain and other central nervous cancers was 59 years in the United States [22], although data suggest that the majority of adults in the United States use some form of social media [23]. Among caregivers

of patients with childhood and early adolescent cancer, Facebook is the most commonly used social media platform over Twitter and Instagram [24]. Even on Twitter, childhood patient networks are also diffused across multiple hashtags such as #PedOnc (pediatric oncology), #PedPC (pediatric palliative care), #AYACSM (adolescent and young adult cancer social media), and #childhoodcancer. Figure 2 highlights that the proportion of tweets from caregivers has steadily decreased from roughly 14.2% in 2012 to 2.4% in 2018, thereby hinting at much-needed work to integrate patient caretakers into the #BTSM community for support and guidance. The observed decline could be related to the high mortality rate of people with malignant brain tumors, caregiver burnout, and the need to distance oneself from the brain tumor community after the death of a loved one [12]. A proactive method to promote #BTSM participation is to integrate more family members and caretakers to the community, and @BTSMchat organizers recruited a caregiver into their team in 2018. Compared to other disease-specific hashtags, family and caretaker involvement may be particularly important for #BTSM as patients with brain tumor are susceptible to cognitive deficits due to the tumor location.

As #BTSM and other disease-specific hashtags continue to expand, it is increasingly important to address the accuracy of health information shared on Twitter as it affects the public and may have patient safety implications. One way to address these issues is to increase the number of brain tumor stakeholders with a blue Twitter-verified profile badge, which Twitter describes as an authenticity indicator as a public interest account. As of April 1, 2020, the majority of #BTSM influencers identified by our network plot lacks this important badge. By increasing the number of verified accounts, it would be easier for the general public to vet individual accounts they might encounter, thereby increasing the likelihood of accurate information consumption among patients and survivors.

A key feature of #BTSM is the high involvement of national organizations and groups such as @BTSMchat in prompting conversations within the hashtag. By hosting a live, facilitated discussion on the first Sunday of each month, @BTSMchat creates an interactive space for various stakeholders within the brain tumor community (eg, patients, caregivers, clinicians, researchers) and forms a structure for sustained engagement. The @BTSMchat chat themes in 2018 ranged from large systemic issues (such as clinical trials and their failure to meet recruitment goals) to more personal topics (such as overcoming guilt, talking with family, holistic patient support through education and peer support, and discussing sensitive topics). @BTSMchat's identified future goals include (1) widening the lived experiences of tweet chat participants with a focus on social diversity, geographic diversity (rural vs urban), and diversity of central nervous system tumor experience; (2) partnering with moderators of brain tumor Facebook Groups by providing Twitter education and cross promoting the tweet chat; and (3) creating and curating resources for patients and care partners who want to pursue or deepen their participation in peer-to-peer support opportunities, including chat moderation, and through community education, conference attendance, public speaking, and research collaboration through traditional medical research pathways.



When social media platforms first entered the national consciousness, health care professionals could not grasp how these new forms of communication could apply to their work. It is now clear that the larger platforms such as Facebook, WeChat, and Twitter are here to stay. Twitter is a 14-year-old platform that continues to connect people not by personal relationship but rather by shared interests. It is only natural that patients, caregivers, clinicians, and researchers all with a shared interest in neuro-oncology would ultimately converge, share,

and form connections, because Twitter was designed to facilitate these forms of information exchange. It is interesting to note that #BTSM was created by patients who saw a needs gap and filled it, without the intervention of a health care structure and is sustained without formal institutional support.

We believe that the growth of the #BTSM hashtag highlights the potential of patient innovation through social media platforms, and serves as a global source of brain tumor information for all brain tumor constituents.

## Acknowledgments

This work was supported by the American Brain Tumor Association, the National Brain Tumor Society, LOGLIO, and STOPhersentumoren.nl. We also thank the Student Association of Yale School of Public Health (SAYPH) for a travel grant (to JF) to present our work at the 2019 SNO Annual Meeting, as well as @BTSMchat organizers who were not authors on this paper: Lisa O'Leary, Esq., Sabine Schwab, PhD, Jeremy Pivor, MA, and Ashley L. Sumrall, MD.

## Conflicts of Interest

None declared.

## References

1. Global Digital Population as of April 2019 (in Millions). URL: <https://www.statista.com/statistics/617136/digital-population-worldwide/> [accessed 2020-09-30]
2. Most Popular Social Networks Worldwide as of April 2019, Ranked by Number of Active Users (in Millions). URL: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> [accessed 2019-05-10]
3. Hu A. Reflections: The Value of Patient Support Groups. *Otolaryngol Head Neck Surg* 2017 Apr;156(4):587-588. [doi: [10.1177/0194599817697030](https://doi.org/10.1177/0194599817697030)] [Medline: [28366109](https://pubmed.ncbi.nlm.nih.gov/28366109/)]
4. Kemp E, Koczwara B, Butow P, Turner J, Girgis A, Schofield P, et al. Online information and support needs of women with advanced breast cancer: a qualitative analysis. *Support Care Cancer* 2018 Oct;26(10):3489-3496. [doi: [10.1007/s00520-018-4206-1](https://doi.org/10.1007/s00520-018-4206-1)] [Medline: [29693203](https://pubmed.ncbi.nlm.nih.gov/29693203/)]
5. Cheung B, Wong CL, Gardhouse A, Frank C, Budd L. #CGS2015: An Evaluation of Twitter Use at the Canadian Geriatrics Society Annual Scientific Meeting. *Can Geriatr J* 2018 Jun;21(2):166-172 [FREE Full text] [doi: [10.5770/cgj.21.302](https://doi.org/10.5770/cgj.21.302)] [Medline: [29977432](https://pubmed.ncbi.nlm.nih.gov/29977432/)]
6. Twitter: How to Use Hashtags. URL: <https://help.twitter.com/en/using-twitter/how-to-use-hashtags> [accessed 2019-05-18]
7. Katz MS, Utengen A, Anderson PF, Thompson MA, Attai DJ, Johnston C, et al. Disease-Specific Hashtags for Online Communication About Cancer Care. *JAMA Oncol* 2016 Mar;2(3):392-394. [doi: [10.1001/jamaoncol.2015.3960](https://doi.org/10.1001/jamaoncol.2015.3960)] [Medline: [26539640](https://pubmed.ncbi.nlm.nih.gov/26539640/)]
8. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter Social Media is an Effective Tool for Breast Cancer Patient Education and Support: Patient-Reported Outcomes by Survey. *J Med Internet Res* 2015;17(7):e188 [FREE Full text] [doi: [10.2196/jmir.4721](https://doi.org/10.2196/jmir.4721)] [Medline: [26228234](https://pubmed.ncbi.nlm.nih.gov/26228234/)]
9. Halkin T. Newton, MA: National Brain Tumor Society URL: <https://blog.braintumor.org/ga-how-social-media-hashtags-and-dedicated-advocates-are-shaping-the-brain-tumor-community/> [accessed 2019-06-02]
10. Katz MS, Anderson PF, Thompson MA, Salmi L, Freeman-Daily J, Utengen A, et al. Organizing Online Health Content: Developing Hashtag Collections for Healthier Internet-Based People and Communities. *JCO Clin Cancer Inform* 2019 Jun;3:1-10. [doi: [10.1200/CCI.18.00124](https://doi.org/10.1200/CCI.18.00124)] [Medline: [31251658](https://pubmed.ncbi.nlm.nih.gov/31251658/)]
11. Salmi L. How to Participate in the #Btsm Tweet Chat. 2013. URL: <https://thelizarmy.com/2013/05/how-to-participate-in-the-btsm-tweet-chat/> [accessed 2020-06-01]
12. Cutshall NR, Kwan BM, Salmi L, Lum HD. "It Makes People Uneasy, but It's Necessary. #BTSM": Using Twitter to Explore Advance Care Planning among Brain Tumor Stakeholders. *J Palliat Med* 2020 Jan;23(1):121-124. [doi: [10.1089/jpm.2019.0077](https://doi.org/10.1089/jpm.2019.0077)] [Medline: [31170019](https://pubmed.ncbi.nlm.nih.gov/31170019/)]
13. Symplur LLC. Symplur - Overview. URL: <https://symplur.com> [accessed 2020-09-20]
14. R Foundation. The R Project for Statistical Computing. Vienna, Austria: R Foundation URL: <https://www.r-project.org/> [accessed 2020-09-20]
15. Twitter: About your activity dashboard. URL: <https://help.twitter.com/en/managing-your-account/using-the-tweet-activity-dashboard> [accessed 2020-06-01]
16. The Gephi Consortium. The Open Graph Viz Platform. URL: <https://gephi.org/> [accessed 2020-09-19]

17. Lie OV, van Mierlo P. Seizure-Onset Mapping Based on Time-Variant Multivariate Functional Connectivity Analysis of High-Dimensional Intracranial EEG: A Kalman Filter Approach. *Brain Topogr* 2017 Jan;30(1):46-59. [doi: [10.1007/s10548-016-0527-x](https://doi.org/10.1007/s10548-016-0527-x)] [Medline: [27722839](https://pubmed.ncbi.nlm.nih.gov/27722839/)]
18. LaFond S. BTAM 2017: Be a Part of Our Team. Newton, MA: National Brain Tumor Society; 2017. URL: <https://blog.braintumor.org/btam-2017-be-a-part-of-our-team/> [accessed 2020-09-30]
19. Head to the Hill - The Experience. Newton, MA: National Brain Tumor Society URL: <https://braintumor.org/your-voice/head-to-the-hill-experience/> [accessed 2020-06-01]
20. Society for Neuro-Oncology. Preliminary Program and Registration Information - 23rd Annual Meeting and Education Day of the Society for Neuro-Oncology. 2018. URL: <https://www.soc-neuro-onc.org/SNO/2018Annual/2018AnnualHome.aspx> [accessed 2020-05-01]
21. @BTSMchat: If you are here tonight to just listen, please tweet ?#BTSM? so we know you're in the audience #btsm. 2017. URL: <https://twitter.com/BTSMchat/status/937508820661743626> [accessed 2020-06-01]
22. Howlader N, Noone A, Krapcho M, Miller D, Brest A, Yu M. SEER Cancer Statistics Review, 1975-2016. Bethesda, MD: National Cancer Institute; 2019. URL: [https://seer.cancer.gov/csr/1975\\_2016/](https://seer.cancer.gov/csr/1975_2016/) [accessed 2020-09-30]
23. Pew Research Center. Social Media Fact Sheet. Washington, DC: Pew Research Center; 2019. URL: <https://www.pewresearch.org/internet/fact-sheet/social-media/> [accessed 2020-06-01]
24. Wilford JG, Osann K, Wenzel LB. Online social network use among parents of young childhood cancer survivors. *JCO* 2017 Feb 10;35(5\_suppl):82. [doi: [10.1200/jco.2017.35.5\\_suppl.82](https://doi.org/10.1200/jco.2017.35.5_suppl.82)]

## Abbreviations

**BCSM:** Breast Cancer Social Media  
**BTAM:** Brain Tumor Awareness Month  
**BTSM:** Brain Tumor Social Media  
**NBTS:** National Brain Tumor Society  
**NIH:** National Institutes of Health

*Edited by G Eysenbach; submitted 01.07.20; peer-reviewed by D Attai, S Jackson; comments to author 15.07.20; revised version received 08.08.20; accepted 10.08.20; published 08.10.20.*

*Please cite as:*

*Feliciano JT, Salmi L, Blotner C, Hayden A, Nduom EK, Kwan BM, Katz MS, Claus EB*

*Brain Tumor Discussions on Twitter (#BTSM): Social Network Analysis*

*J Med Internet Res 2020;22(10):e22005*

URL: <http://www.jmir.org/2020/10/e22005/>

doi: [10.2196/22005](https://doi.org/10.2196/22005)

PMID: [33030435](https://pubmed.ncbi.nlm.nih.gov/33030435/)

©Josemari T Feliciano, Liz Salmi, Charlie Blotner, Adam Hayden, Edjah K Nduom, Bethany M Kwan, Matthew S Katz, Elizabeth B Claus. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Experiences of Serving and Ex-Serving Members With the PTSD Coach Australia App: Mixed Methods Study

Jane Shakespeare-Finch<sup>1</sup>, PhD; Karolina K Alichniewicz<sup>2</sup>, PhD; Esben Strodl<sup>1</sup>, PhD; Kelly Brown<sup>1,3</sup>, MA; Catherine Quinn<sup>4</sup>, PhD; Leanne Hides<sup>4</sup>, PhD; Angela White<sup>5</sup>, PhD; Gabriel Gossage<sup>6</sup>, MA; Loretta Poerio<sup>6</sup>, PhD; Dimitri Batras<sup>6</sup>, PhD; Samantha Jackson<sup>7</sup>, BSc; Jess Styles<sup>7</sup>, BSc; David Kavanagh<sup>1</sup>, PhD

<sup>1</sup>Queensland University of Technology, Brisbane, Australia

<sup>2</sup>Ramsey Health, Brisbane, Australia

<sup>3</sup>Gallipoli Medical Research Foundation, Greenslopes Private Hospital, Brisbane, Australia

<sup>4</sup>University of Queensland, Brisbane, Australia

<sup>5</sup>Royal Brisbane and Womens Hospital, Brisbane, Australia

<sup>6</sup>Department of Veterans Affairs, Canberra, Australia

<sup>7</sup>Mental Health Strategy & Research, Joint Health Command, Canberra, Australia

**Corresponding Author:**

Jane Shakespeare-Finch, PhD

Queensland University of Technology

Victoria Park Road, Kelvin Grove

Brisbane, 4059

Australia

Phone: 61 0731384932

Email: [j.shakespeare-finch@qut.edu.au](mailto:j.shakespeare-finch@qut.edu.au)

## Abstract

**Background:** PTSD Coach Australia is an app for serving and ex-serving defense members and was adapted for the Australian context in 2013 from PTSD Coach, which was created in the United States.

**Objective:** This study aimed to provide a user-centered evaluation of the app from the perspective of serving and ex-serving members of the Australian Defence Force.

**Methods:** Qualitative data were collected in response to questions to participants in 1 of 5 workshops (n=29) or in telephone interviews (n=24). Quantitative data were collected using the user version of Mobile Apps Rating Scale (uMARS).

**Results:** Analysis of the qualitative data demonstrated mixed support for the app. While some people found it extremely useful, especially as an adjunct to therapy, others pointed out limitations and cautioned against the app potentially triggering symptoms in people with PTSD. This perceived risk was usually found to stem from frustration with the app's functionality rather than its content. Participants spoke about the helpful and unhelpful aspects of the app and barriers to its use and made suggestions for improvement. Many participants encouraged its continued use and highlighted the need for it to be promoted more broadly, as many were not aware of it until they were invited to participate in this research.

**Conclusions:** PTSD Coach Australia was seen in a positive light by some participants, but others thought it had too much text and the potential to trigger a traumatic response in users with PTSD. A need to update the app was also a common comment as was the need to increase awareness of the app's existence.

(*J Med Internet Res* 2020;22(10):e18447) doi:[10.2196/18447](https://doi.org/10.2196/18447)

**KEYWORDS**

PTSD; stress disorders, posttraumatic; self-management; evaluation, qualitative; health, veterans; armed forces personnel

## Introduction

Current research suggests that serving defense members have higher levels of posttraumatic stress disorder (PTSD) than

members of the general community [1]. However, perceived stigma and concerns about potential consequences of a diagnosis of mental disorder remain particularly important barriers to help seeking among serving defense members and veterans [2,3],



and self-management of symptoms is a more attractive alternative [2]. A partial solution to these challenges involves digital mental health resources. Web-based interventions for mental health issues have demonstrated benefits for self-management of a wide variety of problems, including PTSD [4]. Smartphone apps offer additional advantages over web programs because of their potentially constant accessibility and high community uptake [5,6]. While there is a wide range of mental health apps to support serving and ex-serving defense members [7], there is limited evidence about their acceptability and impact.

The United States Department of Veterans Affairs treatment providers developed the “PTSD Coach” app in 2011 in collaboration with veterans who had PTSD [8]. Subsequently, the Department of Veterans Affairs in Australia received permission to modify the app for use with an Australian audience. For example, audio was now read by people with Australian accents, Australian prevalence rates of PTSD were used, and some images were changed to reflect the Australian landscape. The resultant “PTSD Coach Australia” was jointly released by the Australian Departments of Defence and Veterans Affairs in February 2013.

There have been some evaluations of the original PTSD Coach app [9–11], including research suggesting its acceptability and benefit for self-management of posttraumatic stress symptoms [12,13]. However, it was not clear whether the Australian version of the app was well accepted and meeting the needs of Australian defense members. Hence, the aim of this research was to assess serving and ex-serving defense members’ perceptions of PTSD Coach Australia, using a combination of qualitative and quantitative methodologies. The research provided an opportunity to determine whether the app retained a high level of acceptability and perceived utility after being in use for several years and adapted to a different national context.

## Methods

### Ethical Approval

The study was approved by The Departments of Defence and Veterans’ Affairs Human Research Ethics Committee (DDVA HREC; E017/004), Queensland University of Technology Human Research Ethics Committee (1700000173), Greenslopes Research and Ethics Committee (17/17), and Townsville Mater Human Research Ethics Committee (E017-004); command approval for participation was obtained from all Australian defense forces.

### Participants

A total of 53 serving and ex-serving defense members volunteered to participate in workshops or telephone interviews. Of these 53 participants, 29 participants (n=8 serving, n=20 ex-serving, and n=1 unknown) took part in 1 of 5 user-experience workshops. Almost all were male (n=28; one indicated gender as “Other”). Of the workshop participants, 18 were from the Army, 4 were from the Air Force, 4 were from Navy, and 3 did not disclose the Australian Defence Force arm they served in. Of the 29 workshop members, 27 provided their age. The age range was 35–79 years (median 62, SD 12.4 years).

Only 3 of the workshop participants were aged under 40 years. The average service time was 14.1 (SD 10.5) years. Most (24/29, 83%) had been deployed, and the average number of deployments was 2.2 (SD 2.5). Of all workshop members, 20 (69%) had been deployed to warlike operations and 15 (52%) had been in a combat role. All had experienced at least one traumatic event in their lifetime, and all but 4 (85%) had received treatment for PTSD.

Further, 24 defense members (n=8 full-time serving, n=15 ex-serving, and n=1 reservist) participated in telephone interviews. Of these, 4 (17%) were female. Of the interview participants, 11 were from the Army, 8 were from the Navy, and 5 were from the Air Force. Their ages ranged from 31 to 80 years with an average age of 55.8 (SD 12.7) years, with only 1 interviewee under the age of 40 years. The average service time was 20.4 (SD 12.8) years. Most had been deployed (21/24, 88%), and their average number of deployments was 4.0 (SD 5.6). Of all telephone interview respondents, 14 (58%) had been deployed to participate in warlike operations and 11 (46%) had been in combat roles. All 24 reported experiencing a traumatic event in their life, and 10 (42%) had received treatment for PTSD.

### Procedure

Recruitment was undertaken through the Department of Veterans’ Affairs’ At Ease website, social media, and community support advisors; Department of Defence websites and newspapers; Defense-orientated organizations such as Mates 4 Mates, White Cloud, the Gallipoli Medical Research Foundation, Defence Families Australia, and the Defence Community Organisation; and treatment and support services including Greenslopes Private Hospital, the Veterans and Veterans Families Counselling Services (now Open Arms – Veterans & Families Counselling), PTSD Trauma Recovery Group Programs, Garrison Health Services, Part Two, Career Shift, Go2 Human Performance, and private practices. The use of focus groups and interviews is closely aligned with the protocol developed by Townsend and colleagues [14] for assessing user experience of digital health resources.

### Workshops

The 2-hour user-experience workshops involved 4–10 people and were facilitated by 2 members of the researcher team, including a registered psychologist with clinical training. Initially, participants were asked to open the app or download it if they had not used it before. Those without a smartphone were provided a smartphone or tablet with the app already loaded and given written instructions on how to download them in the future. After spending approximately 20 minutes investigating the app, they completed the user version of the Mobile Apps Rating Scale (uMARS) [15], which has 4 subscales (Engagement, Functionality, Aesthetics, Information) and 4 subjective quality items (Would you recommend the app? How many times would you use it? Would you pay for this app? Overall star rating). uMARS has been found to be a reliable and valid measure [15,16].

A focus group discussion was then facilitated by one of the research team members. Focal issues for all groups included

their overall experience of using the app, specific aspects of the app that were most and least helpful, what role (if any) the app may have in supporting people with trauma-related symptoms, potential barriers and enablers to its use, how these might be addressed, and how the app might be improved.

## Interviews

As conducting workshops around the country was not possible, telephone interviews were offered to increase participation from people in different geographical areas. As workshop facilitators were not on hand to assist with app download and exploration, participants who had not used the app in the past were given an opportunity to download it and become familiar with its content 2 weeks prior to the interview. This approach was also taken, as participants in interviews were likely to be using their phones for the interview, potentially the same device they were downloading the app onto. The individual 20- to 40-minute structured telephone interviews were conducted by registered psychologists with clinical training. Topics were the same as for the user-experience workshops. Interviews were audiotaped and continued until saturation of themes was obtained.

## Qualitative Data Analysis

All audio of the user-experience workshops and the interviews was professionally transcribed; themes were extracted from the data and grouped under the 6 questions posed, adhering to Braun and Clarke's [17] Thematic Analysis approach. Analysis

involved initially becoming familiar with the data through reading and re-reading transcripts and taking notes in the margins. Subsequent readings involved a process of constant comparison within the transcript and then between transcripts, combining notes into lower-order themes that were eventually grouped into higher-order themes. This approach focused on patterns of meaning across the whole group of participants and within subgroups (ie, serving and ex-serving defense members). Transcript data were supported by observational data, whereby notes were taken by facilitators about how participants were seen to be using the app.

## Expert Ratings of the App

Two raters, including the lead author on the paper regarding the development of the expert version of the Mobile Apps Rating Scale (MARS) [16], provided ratings of the app.

## Results

### Quantitative Ratings

Results of the uMARS and MARS are presented in Table 1. The app was rated moderately high on the uMARS, with its information and functionality being the most highly endorsed features. An expert rated aesthetics, subjective quality, and functionality somewhat lower than the uMARS mean for defense members, but otherwise the expert ratings were similar.

**Table 1.** Ratings of PTSD Coach Australia on the uMARS and MARS.

Subscales	uMARS <sup>a</sup> , mean (SD)	MARS <sup>b</sup> scores - Rater 1	MARS scores - Rater 2
<b>Objective subscales<sup>c</sup></b>			
Engagement	3.48 (0.63)	3.60	3.60
Functionality	4.09 (0.58)	3.75	3.25
Aesthetics	3.91 (0.67)	3.67	2.33
Information	4.10 (0.57)	4.71	4.33
Overall mean	3.87 (0.46)	3.93	3.38
<b>Subjective subscales<sup>c</sup></b>			
Subjective quality	3.54 (0.57)	2.50	3.25
Perceived impact	3.36 (0.97)	3.67	3.33

<sup>a</sup>uMARS: Mobile Apps Rating Scale (User Version).

<sup>b</sup>MARS: Mobile Apps Rating Scale.

<sup>c</sup>All uMARS and MARS items were rated on a 5-point scale: 1 - very poor, 2 - poor, 3 - average, 4 - good, 5 - excellent.

## Qualitative Responses

As the same 6 questions were posed to all workshop and interview participants, results are grouped according to the 6 questions. The themes were spoken about in every workshop group and by all interviewees.

### Knowledge and Experience of the App

Approximately half of the participants had used the app (most interviewees, but fewer workshop participants). Participants who had used the app (some of whom had used it for years) were generally introduced to it by their psychologist or

psychiatrist, or by the Veterans and Veterans Families Counselling Services (now known as Open Arms-Veterans & Families Counselling). It was seen as an app that was easily accessed and used: "Anyone can access it, the programme, and it was easy to use." Most interviewees also described having a positive experience with the app, finding it user-friendly, easy to navigate, and written in a clear and understandable language.

Some participants who had never used the app spoke with some self-deprecation: "I have no idea how to (use it), for idiots like myself" and "We're all dinosaurs." Most of these non-app users appreciated that the app might be useful if they had some digital knowledge, for example,

*If you know what you're doing, it would be effortless" and "It'd be helpful for people who know how to use one (app).*

Another participant said,

*I think for someone who's never used it before, I just need time to sit down on my own and go through it, and really try and understand how it all works, and then make some sort of opinion on it.*

Some participants observed that the purpose of PTSD Coach Australia may be unclear, given that it contains a self-assessment tool as well as strategies to manage symptoms. For example,

*...You need to define what you're going to use the application for. The application seems to me like it's got two purposes. One is to actually, 'Do I have PTSD?' and 'Here are the tools I need to manage PTSD.' To me, that can be dangerous in the fact that you don't want people to think that they have PTSD and then start manage, self-managing. What you prefer is that, um, you want people that have been diagnosed with PTSD and are using this as an additional tool to actually help manage their PTSD.*

### Helpful Characteristics

There were various aspects of PTSD Coach that were found to be helpful. One person explained that they liked the ability to put in reminders: "...so it gives me space to separate my mental health care from trying to remember everything else." Others commented that they "liked the quotes" referring to some being from "Janis Joplin and a Swedish proverb" that were "thought provoking." Participants generally appreciated the audios. For example, one participant said,

*I think going down the road, good to listen to, but also it allows you the opportunity to do it (listen) discreetly by just having your ear phones in," and another said, "I personally didn't read it, I listened to it.*

Others liked to read, so they could go back over material easily if needed.

Another aspect that seemed to be helpful was the ability to monitor stress:

*I've found it useful to let me know where my stress levels are...actually are.*

This feature had been used for a number of reasons, including to self-monitor. For example, if a person was not feeling great:

*Why (am I) such a cranky s\*\*t this week? And I'll look at it and I'll say: oh here I am, I'm up...I'm up here, um and calm down a little bit...so it was a little bit of a diagnostic tool for me*

Others used the stress monitoring feature when making a claim for veterans' support. For example,

*I've also used it (Subjective Units of Distress) as a tool to help in my claim and they've actually photographed my stress levels through the app.*

Many people commented positively about the amount of information. The learning function was seen to be most helpful for people who were not very aware of PTSD symptoms and how they may manifest:

*...the learning function is good. It will certainly help people that don't have any understanding or exposure to PTSD, so like family members and colleagues and that sort of thing.*

For those who had been in therapy, the information was not new, but could be a useful reminder. For example,

*I think there's plenty of information in there and you can spend as little or as much time in there just discovering what's out and about*

*Once you drill down into every folder and sub-folder there's a lot of info in there and it's just stuff that most people already know but it's good to have the reminder.*

Participants also commented on the "nice clinical feel" to give the app credibility: "it looks credible, it looks professional."

Explaining the self-assessment section was useful, an interviewee said,

*...so I can go back and, and have a look at that, in terms of how I might be feeling that day, if I've got something coming up, ... I guess it's a preparatory thing in my mind, to prepare myself for an environment that I may not feel comfortable in.*

People reported using this element to track their well-being such as "... (the app) helps identify how you're travelling."

Most participants who had used the app had accessed the Manage Symptoms and Self-Assessment sections. They found these helpful in managing their symptoms and to increase awareness of how they were feeling:

*...mostly using it to manage symptoms, like just to remind yourself. ...if you start thinking about all your crap, you can go on here and go, let's sort this out now instead of festering on it.*

Features aimed at assisting relaxation and sleep were generally seen to be positive:

*I've used it a fair bit just for sleep and relaxation...some of the sleep coaches and the relaxing sounds, and just grounding techniques and stuff like that.*

One participant said,

*I found it very straightforward, very intuitive...the font was great...the spoken instructions, and the sort of coaching was all very clear.*

Another participant said,

*...look honestly, a very useful tool, I've done it twice last week, and I'll probably do it again later today, and on those occasions where I think I might be feeling a bit down, I might use it, as well.*

Although some older participants in the workshops struggled with the technology, most of the older participants in the



interviews found the app easy to use. For example, one older participant said:

*...I was quite surprised when I went through this app, at how friendly it was. So, first up, it's a very friendly app. It's very easy to use. Pretty straight forward.*

This difference may in part have been due to interview participants having time to download the app pre-interview, whereas some workshop participants had never used technology before, let alone an app.

Consistent with the workshop data, some interview participants said that the app was not perceived as a standalone tool to manage symptoms of PTSD or to replace a mental health professional. Rather, it was seen as more appropriately used to reinforce learning and strategies introduced in therapy sessions.

*So, there's the app itself, I don't believe (it) is a sole source of solution.*

*...it's a good tool to help as well. I still think, you know, depending on the severity of the PTSD, they would need counselling and they would need, well, professional help...but it's a great little pocket tool to be able to you know, look at in times of crisis when you can't necessarily see a professional immediately.*

### **Least Helpful Characteristics**

Some participants did not like the app or found it unhelpful and frustrating when needing a quick strategy in a crisis situation. For example,

*Um, it became extremely frustrating. So then knowing how sometimes I feel when I have an attack, that wasn't good. It kept going skip, next, skip, next, skip, next. And I thought, well it's not really giving me any help.*

*I think when you're angry, reading something, and especially if you've got to read it, understand it, and put it into practice, is not the best way to do that. ...when I am annoyed, I have to read something two or three times before my brain comprehends it.*

Consistent with views expressed in the workshops, some interviewees with PTSD thought aspects of the app would trigger their symptoms:

*...with PTSD, you, it's a high degree and an instant degree of triggered frustration.*

Some participants experienced technical glitches with the app when they had difficulty downloading it onto their device, uploading songs and pictures, and listening to some of the audio. For example, one participant said,

*Bearing in mind that if you're vaguely thinking about this thing, it's gotta be a smooth entry, which in my case it wasn't.*

Some ex-serving defense members who had experience using apps and in particular, PTSD Coach Australia, voiced concerns addressing technical issues such as flexibility and the potential to personalize the app. For example, one participant's perception was as follows:

*I wanted to write a doctor's appointment (in it), there was no way for me to put it in.*

Another user agreed that the app “doesn't utilise all the capabilities of my (brand of phone).” Lack of flexibility was a widespread theme:

*Learn (from menu) is very important, that should be at the top, but for someone who has been using it for some months or years, they should be able to drag the learning component down at the bottom.*

A history function could also be used as a reminder of what a person was doing when they had a good day: activity icons could cue memories and be selected instead of reading and typing information in about activities undertaken during a good day. Some people suggested that the app was not as personalized as they would like and that different audiences (age groups, sex) needed different information or information communicated in different ways. If it were “...individualised and I can go – oh yeah that, that really resonates with me,” that was expected to increase uptake. Essentially, the app was seen to have outdated functionality when compared with current phone and tablet technology.

Some issues with the stability of the app's functions were identified. One person said,

*I was just trying to set reminders for things then (during a workshop) and I don't know if it's something to do with my phone but it just keeps shutting down every time I try and set a reminder.*

Further probing revealed the person was using a particular brand of Android phone, and some other users of the same brand of phone had the same experience. During the workshops, facilitators also noticed that histories were sometimes lost. For example, a participant did the assessment task, but the app would not let the user see the results—“I don't know where it went. Just disappeared.” This was frustrating, because using the app to monitor their health was seen as important.

Some participants saw the app as an agent of authority. For example,

*In the first paragraph it says... you agree to the terms and conditions. I went right through that front page, and nowhere is there a hidden panel or any sub-element of terms and conditions...I love how authority works.*

Another said the tone was suited to serving members because “they're telling you what to do, and it's like really, like being back in the army with your sergeant...” and a third explained, “as soon as you get out (of the military) you will not be told what to do.” Essentially the app was seen as “very militarized,” and that was a connection that many ex-serving members did not want.

Many people spoke about the density of information, for example: “I found the menus are wordy, lengthy, boring.” Color coding was a suggestion to “make it a little more accessible” such as having different colors for health, learning, and self-assessment. Others suggested increasing the density of information as a user progressed through the layers of the app:

*The first layer, so the initial layer, entry layer looks good, then your second layer, then once you step into the sub-elements, so tertiary and beyond, it then gets very overloaded.*

It was pointed out that this level of complexity and denseness of information was not helpful for people with a mental disorder. When becoming overloaded a person would “switch off and not really care about it.” Some people found that the denseness of information elevated their emotional arousal:

*There's a lot of reading in there, and I switched off. You know, black and white, black and white, read, read, read, reference to X, Y, Z. I don't care about X, Y, and Z, what does this bloody thing do? ... so, I switched off and went away, and had to calm down and then come back to it. It made me angry.*

For some people, the Self-assessment section seemed to evoke distress. For example,

*So I used the self-assessment and I got halfway through it and the questions were actually, for me they started to become disturbing, so I stopped.*

The Learn section of the app was seen as not thorough enough by some interviewees:

*...I must say my first reaction to it (Learn section) was, I thought, oh, this is pretty superficial, you know, what's PTSD, and how it develops, and how common is it...and the stuff like that.*

Others identified Soothing Pictures and Soothing Audio as least helpful:

*Now I've been through those resources that are listed, and basically found them, for me anyway...no use.*

As already noted, some participants found the inspiring quotes useful, but others did not; for example:

*Ah, yeah, it's the last thing I want (laughs)...I'm not into those touchy-feely things....I suppose it's just the word 'inspiring' sort of put me off straight away.*

There were also a number of comments about the app's design:

*(The) one that comes to mind most of all is the computer-generated voice. ...I found that to be, uh, um, not easy to understand, and also very impersonal.*

One person found “some of the text is actually overlaid on top of the scroll bar, the tool bar at the bottom.” Another said: “Once again, too much information. Text is too small.” This view was commonly asserted in both the telephone interviews and the workshop groups and was linked to unwanted emotions. For example,

*...to navigate through and looking at a small screen, I, if I spend more than 20 minutes, I just end up getting really angry with it.*

### **Potential Role of the App in Supporting People With Trauma-Related Symptoms**

Most participants who had used the app had done so with the help of their clinician. PTSD Coach Australia was seen by them as a positive adjunct to therapy: “...more helpful if you are

seeing a psychologist or psychiatrist.” However, as already mentioned, many participants commented about the potential for material to retrigger their symptoms. For example,

*You want to be careful you don't stir it all up...reliving it and rehashing it and all that. You know?*

Some people did not think using the app would “push people to get help,” whereas others thought recognizing symptoms through the app may spur people on to seeking help: “you say, s\*\*t this is happening to me so you're going to do something about it.” Many of the older participants thought it unlikely that they would access an app or that others from their cohort would, stating that when a person is experiencing symptoms of anxiety or depression “you'll want to talk to somebody that's going to give you advice...”

The dangers of self-diagnosis were also raised:

*People self-diagnose and they read the symptoms and say, yeah I have PTSD. You really have to go see an expert and psych to get information. You shouldn't do it yourself.*

A perceived lack of consistency in the content of the app with what clients were taught about PTSD and its management in therapy was also discussed as a retriggering feature:

*Well you know, me personally, I've come a long way from hating the world, and I'm comfortable with where I am now, and then I come in here and this (the app) contradicts what I'm being taught, which has made the world a better place for me. So that makes me confused, angry, emotional, frustrated.*

### **Potential Barriers and Enablers to the Greater Use of the App**

The main enabler to the greater use of the PTSD Coach Australia was that it is available for free, which was good “because I wouldn't pay for it.” A barrier for some was their lack of familiarity with technology in general and apps in particular. Some suggestions were offered to make access easier:

*Get the RSL [Returned Services League] to set up trainee workshops through their sub-branches...work in a group and kind of demystify it because it's actually quite simple when you know the trail.*

For many older (ie, Vietnam veteran) participants, the technology was foreign and using it was overwhelming.

Others pointed out that an app of this sort would be difficult to negotiate if they had additional physical complications such “as a stroke or a chronic illness such as Parkinson's.” Another suggested the app be available in “A DVD that you can just stick in...and watch and listen to it.” These quotes came from older participants and may further indicate discomfort with the technology. A lack of visible marketing of the app was noted.

The name of the app was raised by many participants, and the icon was also discussed. There were 2 issues in this theme. The first was that the app's title indicated that it was for people who have PTSD, but that the app could be used by a much broader population. The second issue concerned stigma. Having an icon



on their device that others may see (eg, on the train or bus) that suggests they had PTSD was considered to reveal personal information. Stigma regarding mental health was still seen as a problem, and more education was needed:

*If we can educate our new members and our current members, well, you know, there is still a lot of prejudice against mental health (problems) within Defence.*

The name of the app did not seem to be useful in breaking down that stigma. When asked what would stop someone using the app, one participant summed up the sentiment as, “shame, sense of shame.” When asked about potential barriers to use, others said,

*Just its name, PTSD Coach. I don't know what you could call it, but people don't want to be labelled.*

*My nephew wouldn't download it. My friends wouldn't download it. Uh, most people won't download it if it's called PTSD coach. ...if you call it Wellness Coach...*

Other suggestions for renaming it included “Mental Health Coach” but some doubted whether that would reduce perceived stigma. One suggested that the information screen “could say this can help if you are having trouble sleeping, if you've got anger issues, blah blah...”

Many participants stated that they could not find any other barriers or challenges in using the app, as it was informative and well set up: “...I haven't found it to be difficult at all. It's just been interesting...it's easy to read, it's easy to access and it's easy to understand.” Another participant said, “No barriers. I was fine. I can see that the app is useful ...when it's (used) in conjunction, obviously, with some sort of help as well.”

### Ways to Improve the App

The majority of participants stated that no improvements to the app were needed: “I think it's great. I think, for what it's meant to do, I think it's fantastic.” However, others suggested that including more strategies and interventions such as crisis intervention, mindfulness, mood tracking, and elements of exposure therapy would be beneficial: “So, as it stands now, the app can assist, but it wouldn't be able to assist through a crisis moment.” Another common subtheme was that participants suggested expanding information in the Learn section, for example, by including links to additional information, including videos with psychoeducation and presentations of techniques:

*(Include) a hyperlink to a webpage or something that may be able to help people if they want to do further research.*

*There should be, uh, there should be video examples, rather than trying to find written descriptions of how you deep breath.*

Most participants noted the need for increased awareness about the app and the need for its promotion to assist serving and ex-serving defense members who had experienced trauma:

*...how it can be used is by making people aware that it exists, seriously. I think that's the priority*

*...to actually send out a thing that says, “Here's a link to an app, and you can download it, and it doesn't cost you anything...”*

One person who described himself as belonging to the younger generation said,

*...it probably wouldn't hurt to have a YouTube video explaining what the purpose of it (is) and how to use it and how to maximise the benefit out of it.*

Rather than a video explaining how to use the app, others suggested

*...just a little video clip, a YouTube like video clip on it (the app itself). Nothing that takes up space or time, but just enough to get the gist of what's going on...maybe the odd skit.*

Using the app as a way of connecting with peers was suggested. When speaking about the possibility of the app being a platform for that purpose, one participant explained,

*A group amongst veterans is the strongest form of medicine you can have, the one-to-one with your peers.*

This potential function may also be a way of keeping track of each other; if a person had not heard from a peer for some time, it may prompt them to contact them to check how they were going.

Although many people liked the fact that they could load their own music and pictures, others suggested that “maybe having a library in there of pictures and sounds rather than having to upload your own into the system.” A younger participant said that he “liked that you could alter and change images to what you like. I had a bit of fun like, putting all my kids in.” However, there were some restrictions in the functionality of the app when trying to alter images for things like relaxation, for example, the app “doesn't allow you enough freedom to perhaps get something else off the internet without you having to do a separate action and then upload it to the app.” Others found some of the images insufficient; they were perceived as “a little bit sterile.” One participant was specific in his feedback about some of the images:

*...a picture of a person's head with brain cogs, okay it was clear, but I find it a bit offensive actually. It's like...I'm on the app that's gonna help me deal, I don't want to feel like I'm being analysed or I'm a nut case, or a statistic by the other icon with a line coming up and a lightning bolt going down.*

Some people remarked that the images on the app were not relaxing to them:

*There is a dog in a purple field in a yoga position...what I find is that is not relaxing, it's sort of more something like, you know, something a hippie would use.*

## Discussion

Many responses to PTSD Coach Australia by the serving and ex-serving defense members who participated in this research

were positive, with uMARS ratings in the “average” to “good” range and many positive comments. Most participants said the app was user-friendly, easy to navigate, and helpful in managing symptoms of PTSD. This feedback was consistent with studies evaluating the American version of the app [8,13,18,19]. Key components of the app (eg, Manage Symptoms, Find Support, and Learn) appeared to be maintaining the acceptability and perceived benefit that were observed in earlier research on its American version [18], and many participants continued to value the ability to track changes in stress or other symptoms. However, scheduling of reminders appeared less valued by some users than in past studies [12,18]. Inspiring quotes were generally but not universally valued.

Some participants suggested its use as an adjunct to face-to-face treatment rather than as a stand-alone tool. Consistent with broader evidence on the benefits of having a therapist or coach to guide and support the use of digital mental health tools [18], a pilot study comparing self-managed versus clinician-guided use of PTSD Coach found that clinician support appeared to increase the app’s symptomatic benefits [4], although both modes of application improved PTSD symptoms. Some participants also expressed concerns that self-managed use of the app may generate frustration or trigger trauma symptoms in some users and may not be well suited to crisis management. Improvements to the app’s functionality and reduction of text could avoid triggering frustration and distress in vulnerable users. A recent systematic review and rating of apps designed to assist with PTSD [20] noted the absence of safety planning in the PTSD Coach and PTSD Coach Australia apps. While the app has a listing of crisis services and an ability to set up individualized support contacts, additional safety planning within the app may be indicated.

Consistent with previous US research [8], many participants reported that the app was easy to use, although some wanted faster navigation to favorite tools. Many now found the app to have suboptimal functionality, have some out-of-date content, and lack engaging graphics and flexibility. These observations may reflect changes in user expectations. Importantly, current participants also experienced significant technical problems when using the app on some Android devices, suggesting that its updates had not kept up the pace with developments in

devices and software. Interestingly, previous research suggested that the app would appeal even to technology novices and older individuals [8], whereas (despite the passage of time) current respondents expressed the view that people who had little experience with digital technology may struggle with its use (and some saw this as applying to themselves). Some also noted that use of the app would be difficult for people with a physical disability or cognitive deficit and for those experiencing a heightened state of emotional arousal. Applying recent advances in app design and functionality could help address these issues.

While the availability of PTSD Coach Australia free of charge was seen as an enabler of its use, participants noted a lack of awareness of its existence, suggesting that greater promotion of the app was needed. Some participants proposed that this and other mental health apps could be discussed and used prior to deployments to assist with stress management.

A limitation in the current research was that the participants were predominantly older than the average age of currently serving personnel, which may limit the representativeness of findings in relation to use of the app by younger people. There was also a wide range of digital literacy regardless of age. A difference in the procedures used for interviews and workshops could be seen as a limitation: Interviewees were asked to download the app prior to the interview due to the likelihood of using the same device for the download as they used for the interview, whereas workshop participants examined the app at the start of the workshop, giving facilitators an opportunity to observe any difficulties they had in downloading or using the app. This difference in procedure may have resulted in workshop participants having a more recent experience with the app, although themes were consistent across the two approaches to data collection.

Overall, participants attested to the value of PTSD Coach Australia, and in particular, its potential benefit to both serving and ex-serving defense members who were receiving treatment for posttrauma reactions, and for those with less severe problems. While the app could be improved further, its current functionality was viewed mostly positively. A change in its name and increased promotion of the app could substantially improve its uptake and impact, in addition to an update of aesthetics and design to bring it in line with modern apps.

## Acknowledgments

This project was funded by the Department of Veterans’ Affairs (DVA MH EVAL 01/2016). Material presented in the paper is the responsibility of the authors and does not necessarily reflect the views of the Australian Government. We acknowledge the guidance of other steering committee members, including Michael Correll, Kate Henderson, Maureen Montalban, Tracey Noble, and Brian Riches, and the contributions of the Queensland University of Technology staff, including Professor Ross Young, Dr Madhavan Mani, Stoyan Stoyanov, Susan Lingard, and Daniel Krosch. The assistance of Open Arms Veterans & Families Counselling (Veterans and Veterans Families Counselling Services), Mater Health Services North Queensland, Greenslopes Private Hospital, Garrison Health Services, Part Two, Career Shift, Go2 Human Performance, Mates4Mates, White Cloud, the Gallipoli Medical Research Foundation, Defence Families Australia and the Defence Community Organisation is also acknowledged, as is the time and contributions of participating defense members.

## Conflicts of Interest

None declared.

## References

1. Bleier J, McFarlane A, McGuire A, Treloar S, Waller M, Dobson A. Risk of adverse health outcomes associated with frequency and duration of deployment with the Australian Defence Force. *Mil Med* 2011 Feb;176(2):139-146. [doi: [10.7205/milmed-d-10-00241](https://doi.org/10.7205/milmed-d-10-00241)] [Medline: [21366074](https://pubmed.ncbi.nlm.nih.gov/21366074/)]
2. Adler AB, Britt TW, Riviere LA, Kim PY, Thomas JL. Longitudinal determinants of mental health treatment-seeking by US soldiers. *Br J Psychiatry* 2015 Oct;207(4):346-350. [doi: [10.1192/bjp.bp.114.146506](https://doi.org/10.1192/bjp.bp.114.146506)] [Medline: [25858176](https://pubmed.ncbi.nlm.nih.gov/25858176/)]
3. Britt TW, Jennings KS, Cheung JH, Pury CLS, Zinzow HM. The role of different stigma perceptions in treatment seeking and dropout among active duty military personnel. *Psychiatr Rehabil J* 2015 Jun;38(2):142-149. [doi: [10.1037/prj0000120](https://doi.org/10.1037/prj0000120)] [Medline: [25799296](https://pubmed.ncbi.nlm.nih.gov/25799296/)]
4. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. *Gen Hosp Psychiatry* 2016;38:94-98. [doi: [10.1016/j.genhosppsych.2015.09.005](https://doi.org/10.1016/j.genhosppsych.2015.09.005)] [Medline: [26589765](https://pubmed.ncbi.nlm.nih.gov/26589765/)]
5. Chan S, Godwin H, Gonzalez A, Yellowlees PM, Hilty DM. Review of Use and Integration of Mobile Apps Into Psychiatric Treatments. *Curr Psychiatry Rep* 2017 Oct 30;19(12):96. [doi: [10.1007/s11920-017-0848-9](https://doi.org/10.1007/s11920-017-0848-9)] [Medline: [29082425](https://pubmed.ncbi.nlm.nih.gov/29082425/)]
6. Anthes E. Mental health: There's an app for that. *Nature* 2016 Apr 07;532(7597):20-23. [doi: [10.1038/532020a](https://doi.org/10.1038/532020a)] [Medline: [27078548](https://pubmed.ncbi.nlm.nih.gov/27078548/)]
7. Weingardt KR, Greene CJ. New Electronic Tools for Veterans. *N C Med J* 2015;76(5):332-334 [FREE Full text] [doi: [10.18043/ncm.76.5.332](https://doi.org/10.18043/ncm.76.5.332)] [Medline: [26946869](https://pubmed.ncbi.nlm.nih.gov/26946869/)]
8. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. *Mil Med* 2014 Jan;179(1):12-18. [doi: [10.7205/MILMED-D-13-00271](https://doi.org/10.7205/MILMED-D-13-00271)] [Medline: [24402979](https://pubmed.ncbi.nlm.nih.gov/24402979/)]
9. Erbes CR, Stinson R, Kuhn E, Polusny M, Urban J, Hoffman J, et al. Access, utilization, and interest in mHealth applications among veterans receiving outpatient care for PTSD. *Mil Med* 2014 Nov;179(11):1218-1222. [doi: [10.7205/MILMED-D-14-00014](https://doi.org/10.7205/MILMED-D-14-00014)] [Medline: [25373044](https://pubmed.ncbi.nlm.nih.gov/25373044/)]
10. Kuhn E, Eftekhari A, Hoffman JE, Crowley JJ, Ramsey KM, Reger GM, et al. Clinician perceptions of using a smartphone app with prolonged exposure therapy. *Adm Policy Ment Health* 2014 Nov;41(6):800-807. [doi: [10.1007/s10488-013-0532-2](https://doi.org/10.1007/s10488-013-0532-2)] [Medline: [24398700](https://pubmed.ncbi.nlm.nih.gov/24398700/)]
11. Miller CJ, McInnes DK, Stolzmann K, Bauer MS. Interest in Use of Technology for Healthcare Among Veterans Receiving Treatment for Mental Health. *Telemed J E Health* 2016 Oct;22(10):847-854. [doi: [10.1089/tmj.2015.0190](https://doi.org/10.1089/tmj.2015.0190)] [Medline: [26982279](https://pubmed.ncbi.nlm.nih.gov/26982279/)]
12. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: A pilot randomized controlled trial with community trauma survivors. *Psychol Trauma* 2016 May;8(3):384-392. [doi: [10.1037/tra0000092](https://doi.org/10.1037/tra0000092)] [Medline: [27046668](https://pubmed.ncbi.nlm.nih.gov/27046668/)]
13. Keen SM, Roberts N. Preliminary evidence for the use and efficacy of mobile health applications in managing posttraumatic stress disorder symptoms. *Health Systems* 2017 Dec 19;6(2):122-129. [doi: [10.1057/hs.2016.2](https://doi.org/10.1057/hs.2016.2)]
14. Townsend A, Adam P, Li LC, McDonald M, Backman CL. Exploring eHealth Ethics and Multi-Morbidity: Protocol for an Interview and Focus Group Study of Patient and Health Care Provider Views and Experiences of Using Digital Media for Health Purposes. *JMIR Res Protoc* 2013 Oct 17;2(2):e38 [FREE Full text] [doi: [10.2196/resprot.2732](https://doi.org/10.2196/resprot.2732)] [Medline: [24135260](https://pubmed.ncbi.nlm.nih.gov/24135260/)]
15. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). *JMIR Mhealth Uhealth* 2016 Jun 10;4(2):e72 [FREE Full text] [doi: [10.2196/mhealth.5849](https://doi.org/10.2196/mhealth.5849)] [Medline: [27287964](https://pubmed.ncbi.nlm.nih.gov/27287964/)]
16. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015 Mar 11;3(1):e27 [FREE Full text] [doi: [10.2196/mhealth.3422](https://doi.org/10.2196/mhealth.3422)] [Medline: [25760773](https://pubmed.ncbi.nlm.nih.gov/25760773/)]
17. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp0630a](https://doi.org/10.1191/1478088706qp0630a)]
18. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. *Internet Interventions* 2014 Oct;1(4):205-215 [FREE Full text] [doi: [10.1016/j.invent.2014.08.003](https://doi.org/10.1016/j.invent.2014.08.003)]
19. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the Wild: Using Novel Data to Examine the Reach, Use, and Impact of PTSD Coach. *JMIR Ment Health* 2015;2(1):e7 [FREE Full text] [doi: [10.2196/mental.3935](https://doi.org/10.2196/mental.3935)] [Medline: [26543913](https://pubmed.ncbi.nlm.nih.gov/26543913/)]
20. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. *J Consult Clin Psychol* 2017 Mar;85(3):267-273. [doi: [10.1037/ccp0000163](https://doi.org/10.1037/ccp0000163)] [Medline: [28221061](https://pubmed.ncbi.nlm.nih.gov/28221061/)]

## Abbreviations

**MARS:** Mobile Apps Rating Scale

**PTSD:** Posttraumatic Stress Disorder**uMARS:** Mobile Apps Rating Scale (User Version)

*Edited by G Eysenbach; submitted 27.02.20; peer-reviewed by K Possemato, D Hilty; comments to author 30.03.20; revised version received 25.05.20; accepted 26.07.20; published 08.10.20.*

*Please cite as:*

*Shakespeare-Finch J, Alichniewicz KK, Strodl E, Brown K, Quinn C, Hides L, White A, Gossage G, Poerio L, Batras D, Jackson S, Styles J, Kavanagh D*

*Experiences of Serving and Ex-Serving Members With the PTSD Coach Australia App: Mixed Methods Study*

*J Med Internet Res 2020;22(10):e18447*

URL: <https://www.jmir.org/2020/10/e18447>

doi: [10.2196/18447](https://doi.org/10.2196/18447)

PMID: [33030438](https://pubmed.ncbi.nlm.nih.gov/33030438/)

©Jane Shakespeare-Finch, Karolina K Alichniewicz, Esben Strodl, Kelly Brown, Catherine Quinn, Leanne Hides, Angela White, Gabriel Gossage, Loretta Poerio, Dimitri Batras, Samantha Jackson, Jess Styles, David Kavanagh. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# COVID-19 and the “Film Your Hospital” Conspiracy Theory: Social Network Analysis of Twitter Data

Wasim Ahmed<sup>1</sup>, BA, MSc, PhD; Francesc López Seguí<sup>2</sup>, MSc; Josep Vidal-Alaball<sup>3,4</sup>, MD, PhD; Matthew S Katz<sup>5</sup>, MD

<sup>1</sup>Department of Marketing, Operations and Systems, Newcastle University Business School, Newcastle University, Newcastle Upon Tyne, United Kingdom

<sup>2</sup>TIC Salut Social, Generalitat de Catalunya, Barcelona, Spain

<sup>3</sup>Health Promotion in Rural Areas Research Group, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

<sup>4</sup>Unitat de Suport a la Recerca de la Catalunya Central, Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Sant Fruitós de Bages, Spain

<sup>5</sup>Department of Radiation Medicine, Lowell General Hospital, Lowell, MA, United States

**Corresponding Author:**

Wasim Ahmed, BA, MSc, PhD

Department of Marketing, Operations and Systems

Newcastle University Business School

Newcastle University

5 Barrack Rd

Newcastle Upon Tyne, NE1 4SE

United Kingdom

Phone: 44 1912081704

Email: [wasim.ahmed@newcastle.ac.uk](mailto:wasim.ahmed@newcastle.ac.uk)

## Abstract

**Background:** During the COVID-19 pandemic, a number of conspiracy theories have emerged. A popular theory posits that the pandemic is a hoax and suggests that certain hospitals are “empty.” Research has shown that accepting conspiracy theories increases the likelihood that an individual may ignore government advice about social distancing and other public health interventions. Due to the possibility of a second wave and future pandemics, it is important to gain an understanding of the drivers of misinformation and strategies to mitigate it.

**Objective:** This study set out to evaluate the #FilmYourHospital conspiracy theory on Twitter, attempting to understand the drivers behind it. More specifically, the objectives were to determine which online sources of information were used as evidence to support the theory, the ratio of automated to organic accounts in the network, and what lessons can be learned to mitigate the spread of such a conspiracy theory in the future.

**Methods:** Twitter data related to the #FilmYourHospital hashtag were retrieved and analyzed using social network analysis across a 7-day period from April 13-20, 2020. The data set consisted of 22,785 tweets and 11,333 Twitter users. The Botometer tool was used to identify accounts with a higher probability of being bots.

**Results:** The most important drivers of the conspiracy theory are ordinary citizens; one of the most influential accounts is a Brexit supporter. We found that YouTube was the information source most linked to by users. The most retweeted post belonged to a verified Twitter user, indicating that the user may have had more influence on the platform. There was a small number of automated accounts (bots) and deleted accounts within the network.

**Conclusions:** Hashtags using and sharing conspiracy theories can be targeted in an effort to delegitimize content containing misinformation. Social media organizations need to bolster their efforts to label or remove content that contains misinformation. Public health authorities could enlist the assistance of influencers in spreading antinarrative content.

(*J Med Internet Res* 2020;22(10):e22374) doi:[10.2196/22374](https://doi.org/10.2196/22374)

**KEYWORDS**

COVID-19; coronavirus; Twitter; misinformation; fake news; social network analysis; public health; social media



## Introduction

Since its detection in China in late 2019, SARS-CoV-2 has spread worldwide and been declared a pandemic, with negative effects on both human health and the global economy [1-3]. The dramatic consequences of the pandemic have led to the appearance of numerous conspiracy theories. One of the most successful theories linked 5G to the spread of the disease, leading to misinformation and the burning of 5G towers in the United Kingdom [4]. Other hoaxes have linked the severity of COVID-19 to the genetics of Spaniards and Italians or to vitamin D deficiency [5,6]. There has also been no shortage of hoaxes suggesting that chlorine dioxide can cure the disease [5].

In the first week of April 2020, another conspiracy theory emerged, which suggested that the pandemic was really an elaborate hoax. Drivers of this conspiracy theory argued that it could not exist because hospitals were empty or were operating as normal and therefore there were fewer COVID-19 cases than what had been reported. Its supporters also claimed that the severity of the disease had been exaggerated by the scientific community and the press. This then led to the encouragement of citizens to go to hospitals and film them to show that they were empty. The hashtag #FilmYourHospital was used when posting such videos on social media [7,8].

Misinformation is a matter of public concern. If citizens believe that COVID-19 is a hoax, they may be more eager for lockdown restrictions to ease and they may refuse future vaccines. Recent research has found that those who endorse or believe in conspiracy theories are less likely to adhere to government recommendations such as staying at home or keeping a safe distance between others [9]. The study also found that those who believed in conspiracy theories were also less likely to accept a future diagnostic test or vaccination [9]. In case of future outbreaks, it is important to understand the drivers behind this conspiracy theory so that future ones can be prevented and fought.

In this context, the aim of this study was to analyze the Film Your Hospital conspiracy theory that argued that COVID-19 is a hoax. Specifically, we set out to address the following questions:

1. Who were the drivers of this conspiracy theory on Twitter?
2. What online sources of information were used as evidence to support the theory, including the most retweeted tweets?
3. What was the ratio of automated accounts to organic accounts in the network?
4. What lessons can be learned to mitigate the spread of such a conspiracy theory in the future?

This study is the first empirical investigation into the #FilmYourHospital conspiracy theory and its novelty lies in the social network analysis, identification of influencers, identification of the most shared URLs and hashtags, and identification of genuine accounts as compared to bots. This study may have practical value for the development of recommendations for public health authorities.

## Methods

A computer running Microsoft Windows 8 was used to retrieve data in Microsoft Excel 2010 (Microsoft Corp) using the professional version of the social media analysis software NodeXL (Social Media Research Foundation, release code +1.0.1.428+), which provides access to Twitter's search application programming interface (API). The study retrieved data from a 7-day period from April 13, 2020, at 14:19 Coordinated Universal Time (UTC) to April 20, 2020, at 15:59 UTC. Users were included in the data set if they sent a tweet during the time the data was retrieved or were mentioned or replied to in these tweets. The keyword FilmYourHospital was used to retrieve data containing mentions of this phrase, including the hashtag #FilmYourHospital. There were 11,333 Twitter users within the network and they generated a total of 22,785 tweets, broken down as follows: 12,905 (56%) retweets, 2736 (12%) replies, 2425 (10.6%) mentions in retweets, 2194 (9.6%) mentions, and 2525 (11%) individual tweets.

Influential users, topics, and web sources were studied, and a social network analysis of the discussion was conducted with NodeXL, using a validated methodology used in previous research [4], which provided an understanding of the shape of the conversation [10-12]. In addition, a network graph was laid out using the Harel-Koren Fast Multiscale layout algorithm [11]. The graph's vertices were grouped by cluster using the Clauset-Newman-Moore algorithm. NodeXL uses Twitter's search API. URLs were automatically expanded within NodeXL. The Botometer tool [13] was used to find out the ratio of real to automated/bot accounts being used to tweet about this conspiracy. This was achieved by taking a 10% random sample of users who had sent original tweets and running the sample through Botometer. In understanding the network graph, the results build upon previous seminal research [14] that identified 6 network shapes and structures that Twitter topics may follow [11]: broadcast networks, polarized crowds, brand clusters, tight crowds, community clusters, and support networks. When analyzing popular websites within the network, the content was interpreted by reading the website or watching the video to which the tweet provided a link.

## Results

### Social Network Analysis

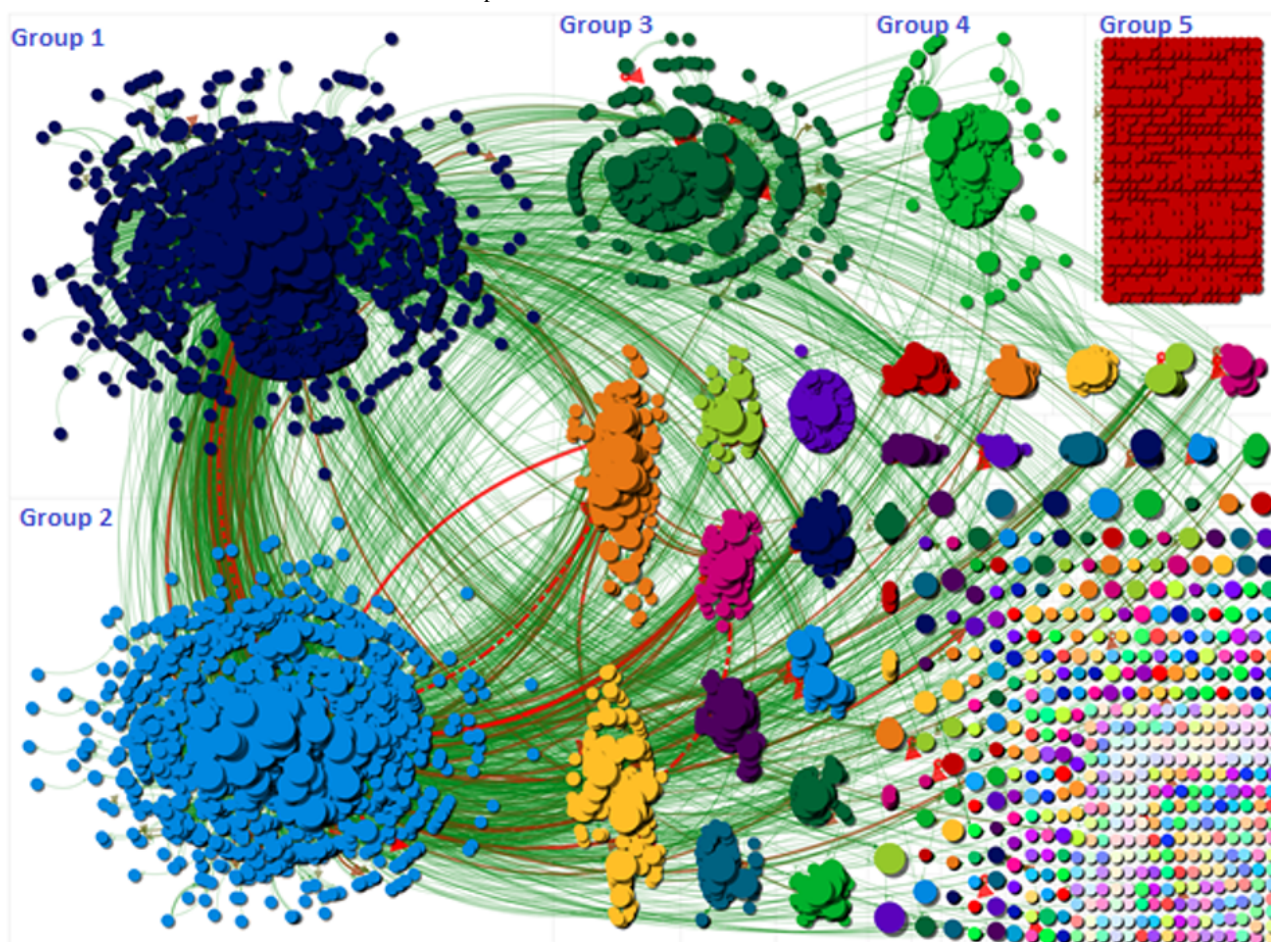
Figure 1 is a social network graph of the discussion surrounding the hashtag #FilmYourHospital. Each small color dot represents a user and a line between them represents an edge. Groups were formed around this topic based on how frequently users mentioned each other. There is an edge for each "replies-to" relationship in a tweet, an edge for each "mentions" relationship in a tweet, and a self-loop edge for each tweet that is not a "replies-to" or "mentions." The size of the nodes are ranked by their betweenness centrality score (BCS) [14], which measures the influence of a vertex over the flow of information between all other vertices under the assumption that information flows over the shortest paths among them. The graph highlights how a number of different communities tweeting about this topic formed on Twitter. It is important to note that the tweets

collected were based on the overall topic and could include a small amount of irrelevant content or users arguing against the conspiracy. The three largest groups (groups 1-3) appeared to be “broadcast groups,” where tweets received a high number of retweets. Group 5 was an “isolates group,” where a number of tweets were sent that did not contain mentions (including retweets). This particular shape appears when a user’s tweet does not contain an @ sign (ie, it is an original tweet expressing a view or opinion that does not refer to any other participant in the discussion). Table 1 provides an overview of the tweets (reworded for anonymity) that contained the most retweets according to the cluster they belonged to within the network.

In groups 1 and 2, we can see that the most retweeted post suggested that the public were being misled about hospitals

being empty and that hospitals were not busy or overrun, as had been suggested by the mainstream media. The tweet then called for users to film their hospitals to show this by using the hashtag #FilmYourHospital and for the hashtag to become a trending topic. This tweet was sent by a verified Twitter user. In group 3, the most popular tweet (posted in Spanish) called for users to watch the videos that were being posted on the #FilmYourHospital hashtag and “draw their own conclusions.” In group 4, the most popular tweet drew attention to an article posted by APNews, an American news agency. The tweet that was most popular in the fifth group drew attention to the #filmyourhospital hashtag. Table 2 provides an overview of the URLs that were most shared during this time.

**Figure 1.** Social network visualization of #FilmYourHospital.



**Table 1.** Most retweeted tweets in groups 1 to 5.

Group	Tweet	Retweet Count
1	<i>Been to two LA hospitals that are supposed to be jam-packed according to the mainstream media. Yet, they are EMPTY and very quiet! Wonder why they are lying to us. As we can't trust the news we should begin covering those – please send pictures of your hospital</i>	21,896
2	<i>Been to two LA hospitals that are supposed to be jam-packed according to the mainstream media. Yet, they are EMPTY and very quiet! Wonder why they are lying to us. As we can't trust the news we should begin covering those – please send pictures of your hospital</i>	21,896
3	<i>Watch the videos that are being published in the #FilmYourHospital tag and draw your conclusions</i>	2178
4	<i>The lame mainstream media has been reporting that the #FilmYourHospital can be traced to Qanon which is a far right group</i>	671
5	<i>Come on people – hospitals are empty – only queues are for shopping supplies. Corona is a Hoax</i>	50

**Table 2.** Most shared URLs.

Rank	Title	URL	Count
1	St Louis Doctor has had enough (video removed by YouTube)	<a href="https://www.youtube.com/watch?v=shBlwyrBii0&amp;app=desktop">https://www.youtube.com/watch?v=shBlwyrBii0&amp;app=desktop</a>	83
2	Health Worker Says EVERYONE Who Dies Has “Corona Virus” on their Death Certificate	<a href="https://www.youtube.com/watch?v=ioUREi1myNs&amp;feature=youtu.be">https://www.youtube.com/watch?v=ioUREi1myNs&amp;feature=youtu.be</a>	63
3	la tendencia que las redes sociales censuraron (English translation: the trend that social networks censored; video removed by YouTube)	<a href="https://www.youtube.com/watch?v=DwY_t1fuuhw&amp;feature=youtu.be">https://www.youtube.com/watch?v=DwY_t1fuuhw&amp;feature=youtu.be</a>	54
4	Polio outbreaks in Africa caused by mutation of strain in vaccine	<a href="https://www.theguardian.com/global-development/2019/nov/28/polio-outbreaks-in-four-african-countries-caused-by-mutation-of-strain-in-vaccine">https://www.theguardian.com/global-development/2019/nov/28/polio-outbreaks-in-four-african-countries-caused-by-mutation-of-strain-in-vaccine</a>	47
5	NHS staff party away in Dudley at Russell Hall Hospital while the world is on lockdown	<a href="https://thetattyjournal.org/2020/04/17/nhs-staff-party-away-in-dudley-at-russell-hall-hospital-while-the-world-is-on-lockdown/?fbclid=IwAR2Y9AoddLDdo-K0Sps40xl_MNBi-AAzhM1jNpYwRLvTlIS5Q59NTWannQoM">https://thetattyjournal.org/2020/04/17/nhs-staff-party-away-in-dudley-at-russell-hall-hospital-while-the-world-is-on-lockdown/?fbclid=IwAR2Y9AoddLDdo-K0Sps40xl_MNBi-AAzhM1jNpYwRLvTlIS5Q59NTWannQoM</a>	25
6	Why are all of these hospital workers dancing in hospitals which appear to be empty?	Removed for anonymity	23
7	Army's Seattle Field Hospital Closes After 3 Days, Without Treating a Single Patient	<a href="https://news.yahoo.com/armys-seattle-field-hospital-closes-165646379.html?soc_src=community&amp;soc_trk=ma">https://news.yahoo.com/armys-seattle-field-hospital-closes-165646379.html?soc_src=community&amp;soc_trk=ma</a>	20
8	Hospitals have slowed down – yet staff at hospitals are able to spend time making TikToks	Removed for anonymity	19
9	Very peculiar situation in the United States at the moment – also wonder why Twitter felt the need to remove the #FilmYourHospital hashtag	Removed for anonymity	18
10	[User handle removed for anonymity] placed in quarantine until May by filling the hospitals. Yesterday the Municipal Hospital of Tatuapé was empty. It is not to preserve lives, but to break São Paulo.	Removed for anonymity	17

In the first video, it is claimed the person speaking is a doctor, but the comments in the video and the office setting indicate that the person in the video may be a chiropractor. In the second video, a health worker notes that, on mainstream television, COVID-19 has been exaggerated and that all deaths in his hospital are being treated as COVID-19–related. The third most frequent URL shows a person who speaks against the conspiracy campaign and requests that people take the pandemic seriously. The fourth URL notes that polio outbreaks in Africa were caused by the mutation of a strain from a vaccine. The fifth URL linked to a news website critical of National Health Service (NHS) staff. Similarly, the sixth most popular URL linked to a tweet

that was critical of medical professionals posting videos of themselves dancing in hospitals on social media. The seventh most popular URL noted that the US army's Seattle Field Hospital closed after 3 days without treating any patients. The eighth most popular URL was a tweet that was critical of medical professionals using TikTok to create videos (the tweet was later deleted). The ninth most shared URL linked to a tweet that questioned why Twitter had removed the trending hashtag #FilmYourHospital. The tenth most shared URL came from a user who claimed that a certain hospital was empty. Table 3 provides insight into the most shared hashtags from this time period that were used alongside the #filmyourhospital hashtag.



The three most frequently used hashtags were #filmyourhospital, #emptyhospitals, and #filmyourhospitals. It was also interesting to see the hashtag #lamestreammedia, which appeared to be critical of the mainstream media. Other hashtags included

#endtheshutdown as well as the hashtag #plandemic, which refers to another viral conspiracy theory. Users sharing #endtheshutdown may have tried to use the conspiracy theory as an argument to end the lockdown.

**Table 3.** Most shared hashtags.

Hashtag	Count
filmyourhospital	13,984
emptyhospitals	1160
filmyourhospitals	1144
filmesuhoospital	1045
covid19	840
coronavirus	759
coronahoax	649
lamestreammedia	609
endtheshutdown	478
plandemic	441

## User Analysis

**Table 4** ranks influential users who have more than 100 followers by their betweenness centrality score. The followers column notes the number of followers the user had. The in-degree coefficient examines the number of inbound connections and is helpful in highlighting the users who are potentially the most trusted. The out-degree coefficient examines the number of outbound connections by users, which is useful in highlighting the number of accounts users were referring to.

The most popular account belonged to a user who identified as a Brexit supporter, followed by two citizens, one of whom was from Brazil. This user was particularly active during this time and had sent out a total of 154 tweets. A number of accounts (such as 6 and 7) appeared to be in support of the idea that COVID-19 is a hoax and these accounts posted a number of tweets using the #FilmYourHospital hashtag. Account number 6 was the most active and had sent out a total of 163 tweets. The most common type of account in the network belonged to citizens supporting this conspiracy theory. Donald Trump did

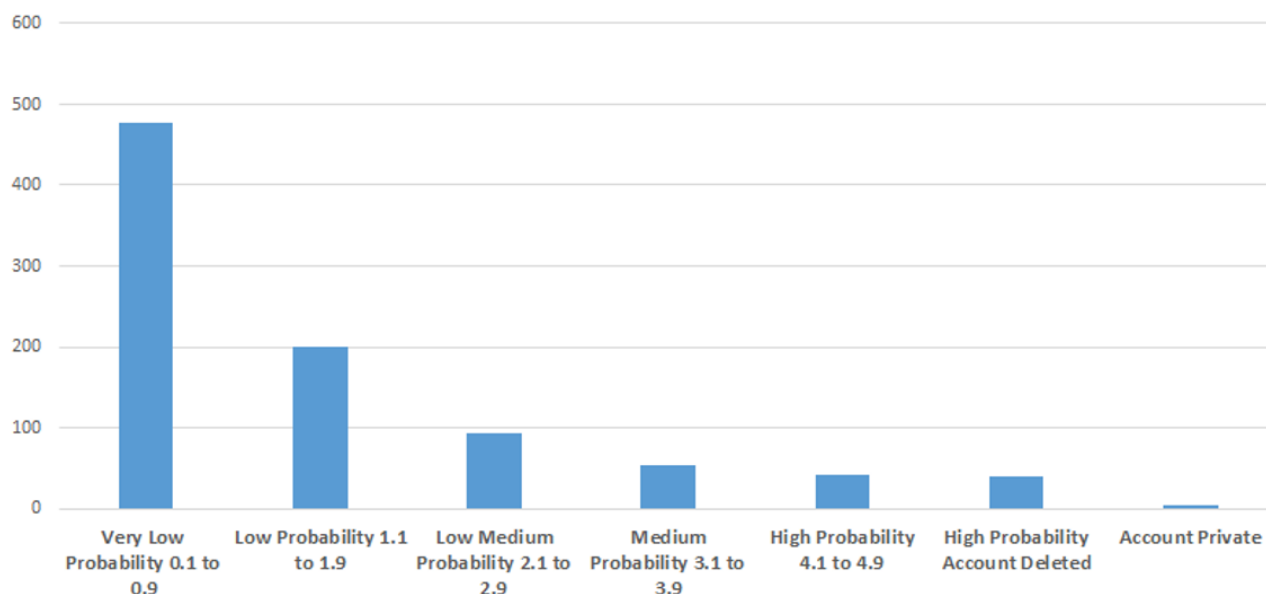
not tweet using the hashtag; however, he was a trusted node in the network, receiving a total of 310 inbound connections.

**Figure 2** presents results from Botometer, which was used to assess the frequency of automated accounts. The results indicated that the rate of automated accounts was low (9.2% of the sample), as the majority of accounts had a score of 0.1 to 0.9 (52% of the sample). The accounts scoring 4 or higher were further examined, as were those that had their tweets deleted. It was found that these accounts had sent 136 tweets during the studied time period, although the total combined number of tweets the bots had sent before their accounts were removed was 323,096. The high-probably bot accounts had received a total of 373,508 likes and had a combined total of 22,855 followers. **Table 5** highlights the most frequently used words that appeared in the user-bios of users who had a score over 4 and/or whose account was deleted.

It can be seen that high-probability automated accounts and/or deleted accounts contained the words “maga” and “trump” with the highest frequencies within their user bios.

**Table 4.** Influential users.

Rank	User	Betweenness centrality	Followers	In-degree	Out-degree	NodeXL Group in <a href="#">Figure 1</a>
1	Brexit supporter (UK)	14,153,758	1219	7	154	2
2	Citizen	12,085,093	194	695	2	1
3	Citizen	12,055,635	2679	631	2	3
4	President Donald Trump	11,194,004	77,802,730	310	0	2
5	George News	11,138,055	141,840	615	1	4
6	Citizen	7,948,257	191	16	163	2
7	Citizen	6,816,583	5904	0	78	6
8	Citizen	6,107,454	176	239	15	1
9	Nongovernmental organization	4,656,161	14,779	14	130	1
10	Citizen	3,616,519	1838	189	5	2

**Figure 2.** Assessing the number of bots on a scale from 0 to 5 using Botometer.**Table 5.** Top 5 most frequently used keywords in bios of users who scored 4 or over and whose account was deleted.

Rank	Word	Frequency
1	trump	7
2	maga	7
3	patriot	4
4	support	4
5	love	4

## Discussion

In regard to the first research question, this study found that the most frequent drivers of the #FilmYourHospital conspiracy theory appeared to be ordinary citizens; one of the most influential users is a Brexit supporter. Previous research on other COVID-19 conspiracies has also found that accounts that appeared to belong to citizens were the most influential [11]. Regarding the second research question, we found that YouTube was the information source most linked to by users. The most retweeted post belonged to a verified Twitter user, suggesting that Twitter tools aimed at identifying verified and trustworthy users do not always work. Regarding the third research question, related to automated accounts, we found that there was a low volume of bots in this social network analysis and a low volume of accounts that were deleted within the network.

YouTube was a popular platform for hosting content used to support this conspiracy. Certain tweets sharing the conspiracy theory also became very popular on Twitter. For instance, the tweet sent by the verified user contained a vast number of replies from other users indicating support of the view that COVID-19 is a hoax. Although our results found that some content had been deleted on Twitter and YouTube, there were still tweets and YouTube videos that remained online.

A number of recommendations can be made to attempt to combat the propagation of misinformation. First, as

recommended in previous research [15], public health authorities could use the hashtag and enlist the assistance of influencers to share the antinarrative (ie, reasons against the conspiracy). The majority of genuine (versus automated) accounts show the nature and popularity of the conspiracy. Second, bots would need to be countered in a more technical manner. It is important to note that most of them already existed before the beginning of this specific conspiracy theory, suggesting a broader problem. However, not all automated bots are set up for malicious purposes. Social media organizations could monitor for suspicious accounts set up to spread misinformation. Third, as our results show, the “citizen-based” #FilmYourHospital hoax should be countered with untargeted, trustworthy information, delivered from public health authorities as well as popular culture influencers (as mentioned above). Previous academic research suggests that explaining flawed arguments and describing scientific consensus to other people may help lessen the effect of misinformation about science [16]. Lastly, public health authorities and governments should enlist the help of the public in using the “report content” features across social media platforms. With a collective effort in rapidly reporting false and misleading information, a coordinated response would allow it to be detected and removed faster. This tactic may be more effective than the public engaging with the content, which may inadvertently raise its profile.

The network shape of this conspiracy theory differs from that of a conspiracy theory linking 5G and COVID-19, as that



theory's network had a large isolated group where users were tweeting without mentioning each other [11]. Conversely, in the FilmYourHospital network, there appear to be denser groups, indicating a larger volume of retweets and mentions. A reason for this is that certain tweets became very popular, were retweeted with high frequency, and received a large number of replies. This demonstrates that this conspiracy theory was larger in size, attracting more users and tweets: when comparing the 7-day time period of this study to that of the COVID-19 and 5G conspiracy [4], "FilmYourHospital" had almost double the number of tweets. The limitations of this study are that a 7-day time period was examined and Twitter's search API was used, which may exclude certain tweets. Future research could seek

to examine the locations of users that had tweeted using the hashtag and correlate this data with the re-emergence of COVID-19. Another important aspect to note is that there were also tweets that were irrelevant, such as ads, and some tweets might have been from users questioning the conspiracy theory. Not all tweets can definitively be interpreted as supporting the conspiracy. Future research could seek to conduct an analysis of tweets to ascertain the percentage of tweets that were related to the conspiracy. Future research could also seek to analyze the social networks of multiple conspiracy theories and look for potential commonalities with respect to key users and information sources.

## Conflicts of Interest

None declared.

## References

1. Wu F, Zhao S, Yu B, Chen Y, Wang W, Song Z, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020 Feb 3;579(7798):265-269. [doi: [10.1038/s41586-020-2008-3](https://doi.org/10.1038/s41586-020-2008-3)]
2. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. Corrigendum to "World Health Organization declares Global Emergency: A review of the 2019 Novel Coronavirus (COVID-19)" [Int J Surg 76 (2020) 71-76]. *Int J Surg* 2020 May;77:217 [FREE Full text] [doi: [10.1016/j.ijsu.2020.03.036](https://doi.org/10.1016/j.ijsu.2020.03.036)] [Medline: [32305321](https://pubmed.ncbi.nlm.nih.gov/32305321/)]
3. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *Int J Surg* 2020 Jun;78:185-193 [FREE Full text] [doi: [10.1016/j.ijsu.2020.04.018](https://doi.org/10.1016/j.ijsu.2020.04.018)] [Medline: [32305533](https://pubmed.ncbi.nlm.nih.gov/32305533/)]
4. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *J Med Internet Res* 2020 May 06;22(5):e19458 [FREE Full text] [doi: [10.2196/19458](https://doi.org/10.2196/19458)] [Medline: [32352383](https://pubmed.ncbi.nlm.nih.gov/32352383/)]
5. Salud Sin Bulos. URL: <https://saludsinbulos.com/?s=covid> [accessed 2020-05-18]
6. Lanham-New SA, Webb AR, Cashman KD, Buttriss JL, Fallowfield JL, Masud T, et al. Vitamin D and SARS-CoV-2 virus/COVID-19 disease. *BMJNP* 2020 May 13;3(1):106-110. [doi: [10.1136/bmjnp-2020-000089](https://doi.org/10.1136/bmjnp-2020-000089)]
7. Zdrozny B, Collins B. Coronavirus deniers take aim at hospitals as pandemic grows. *NBC News*. URL: <https://www.nbcnews.com/tech/social-media/coronavirus-deni> [accessed 2020-05-18]
8. Gruz A, Mai P. Conspiracy theorists are falsely claiming that the coronavirus pandemic is an elaborate hoax. *The Conversation*. URL: <https://theconversation.com/conspiracy-theorists-are-falsely-claiming-that-the-coronavirus-pandemic-is-an-elaborate-hoax-135985> [accessed 2020-05-18]
9. Freeman D, Waite F, Rosebrock L, Petit A, Causier C, East A, et al. Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. *Psychol Med* 2020 May 21:1-13 [FREE Full text] [doi: [10.1017/s0033291720001890](https://doi.org/10.1017/s0033291720001890)]
10. Ahmed W, Marin-Gomez X, Vidal-Alaball J. Contextualising the 2019 E-Cigarette Health Scare: Insights from Twitter. *Int J Environ Res Public Health* 2020 Mar 26;17(7):1 [FREE Full text] [doi: [10.3390/ijerph17072236](https://doi.org/10.3390/ijerph17072236)] [Medline: [32225020](https://pubmed.ncbi.nlm.nih.gov/32225020/)]
11. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *J Med Internet Res* 2020 May 06;22(5):e19458 [FREE Full text] [doi: [10.2196/19458](https://doi.org/10.2196/19458)] [Medline: [32352383](https://pubmed.ncbi.nlm.nih.gov/32352383/)]
12. Ahmed W. Public health implications of #ShoutYourAbortion. *Public Health* 2018 Oct;163:35-41. [doi: [10.1016/j.puhe.2018.06.010](https://doi.org/10.1016/j.puhe.2018.06.010)] [Medline: [30059806](https://pubmed.ncbi.nlm.nih.gov/30059806/)]
13. OSoMe. Botometer. URL: <https://botometer.iuni.iu.edu/#/> [accessed 2020-06-15]
14. White DR, Borgatti SP. Betweenness centrality measures for directed graphs. *Social Networks* 1994 Oct;16(4):335-346. [doi: [10.1016/0378-8733\(94\)90015-9](https://doi.org/10.1016/0378-8733(94)90015-9)]
15. Vidal-Alaball J, Fernandez-Luque L, Marin-Gomez FX, Ahmed W. A New Tool for Public Health Opinion to Give Insight Into Telemedicine: Twitter Poll Analysis. *JMIR Form Res* 2019 May 28;3(2):e13870 [FREE Full text] [doi: [10.2196/13870](https://doi.org/10.2196/13870)] [Medline: [31140442](https://pubmed.ncbi.nlm.nih.gov/31140442/)]
16. Cook J, Lewandowsky S, Ecker U. Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLoS One* 2017;12(5):e0175799 [FREE Full text] [doi: [10.1371/journal.pone.0175799](https://doi.org/10.1371/journal.pone.0175799)] [Medline: [28475576](https://pubmed.ncbi.nlm.nih.gov/28475576/)]

## Abbreviations

**API:** application programming interface  
**BCS:** betweenness centrality score

**UTC:** Coordinated Universal Time

*Edited by G Eysenbach; submitted 10.07.20; peer-reviewed by A Alasmari, J Li; comments to author 13.07.20; revised version received 27.07.20; accepted 15.09.20; published 05.10.20.*

*Please cite as:*

Ahmed W, López Seguí F, Vidal-Alaball J, Katz MS

COVID-19 and the “Film Your Hospital” Conspiracy Theory: Social Network Analysis of Twitter Data

J Med Internet Res 2020;22(10):e22374

URL: <http://www.jmir.org/2020/10/e22374/>

doi: [10.2196/22374](https://doi.org/10.2196/22374)

PMID: [32936771](https://pubmed.ncbi.nlm.nih.gov/32936771/)

©Wasim Ahmed, Francesc López Seguí, Josep Vidal-Alaball, Matthew S Katz. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Low Testosterone on Social Media: Application of Natural Language Processing to Understand Patients' Perceptions of Hypogonadism and Its Treatment

Vadim Osadchiy<sup>1,2</sup>, MD; Tommy Jiang<sup>1,2</sup>, BS; Jesse Nelson Mills<sup>1</sup>, MD; Sriram Venkata Eleswarapu<sup>1,2</sup>, MD, PhD

<sup>1</sup>Division of Andrology, Department of Urology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

<sup>2</sup>Consortium for Health Activity on Social Media, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

**Corresponding Author:**

Sriram Venkata Eleswarapu, MD, PhD

Division of Andrology

Department of Urology

David Geffen School of Medicine, University of California, Los Angeles

10945 Le Conte Avenue

Ueberroth #3361

Los Angeles, CA,

United States

Phone: 1 310 794 3058

Fax: 1 310 206 2369

Email: [SEleswarapu@mednet.ucla.edu](mailto:SEleswarapu@mednet.ucla.edu)

## Abstract

**Background:** Despite the results of the Testosterone Trials, physicians remain uncomfortable treating men with hypogonadism. Discouraged, men increasingly turn to social media to discuss medical concerns.

**Objective:** The goal of the research was to apply natural language processing (NLP) techniques to social media posts for identification of themes of discussion regarding low testosterone and testosterone replacement therapy (TRT) in order to inform how physicians may better evaluate and counsel patients.

**Methods:** We retrospectively extracted posts from the Reddit community r/Testosterone from December 2015 through May 2019. We applied an NLP technique called the meaning extraction method with principal component analysis (MEM/PCA) to computationally derive discussion themes. We then performed a prospective analysis of Twitter data (tweets) that contained the terms low testosterone, low T, and testosterone replacement from June through September 2019.

**Results:** A total of 199,335 Reddit posts and 6659 tweets were analyzed. MEM/PCA revealed dominant themes of discussion: symptoms of hypogonadism, seeing a doctor, results of laboratory tests, derogatory comments and insults, TRT medications, and cardiovascular risk. More than 25% of Reddit posts contained the term doctor, and more than 5% urologist.

**Conclusions:** This study represents the first NLP evaluation of the social media landscape surrounding hypogonadism and TRT. Although physicians traditionally limit their practices to within their clinic walls, the ubiquity of social media demands that physicians understand what patients discuss online. Physicians may do well to bring up online discussions during clinic consultations for low testosterone to pull back the curtain and dispel myths.

(*J Med Internet Res* 2020;22(10):e21383) doi:[10.2196/21383](https://doi.org/10.2196/21383)

## KEYWORDS

hypogonadism; natural language processing; Reddit; social media; testosterone replacement therapy; Twitter

## Introduction

The Testosterone Trials were a coordinated series of placebo-controlled, double-blinded trials intended to elucidate

risks and benefits of testosterone replacement therapy (TRT) in hypogonadal men [1-7]. Despite these recent trials, clinicians continue to be uncomfortable treating these men, in part due to unanswered questions related to cardiovascular outcomes and

cancer risk, as well as how TRT is portrayed in popular culture. Perhaps discouraged by conflicting information from physicians and traditional media, patients sometimes turn to social media platforms to discuss medical concerns with peers [8,9].

Interactive social media channels have emerged as potent resources for individuals to discuss health care concerns [9]. Reddit, an anonymous discussion platform with over 330 million monthly active users, serves as a popular internet destination for discussions of health-related topics [10]. The Reddit forum or subreddit r/Testosterone [11], which boasts over 30,000 active members, is devoted to answering questions, sharing personal accounts, and disseminating resources related to TRT and testosterone levels. Similar discussions occur on other social media sites, including Twitter, a microblogging platform with over 126 million daily active users [12].

We hypothesized that the content of online discussions about low testosterone can be classified into themes that may inform how physicians evaluate, counsel, and treat men with hypogonadism. Here, we apply quantitative natural language processing (NLP) techniques to identify dominant themes of discussions regarding low testosterone and TRT on social media.

## Methods

### Study Design and Sources of Data

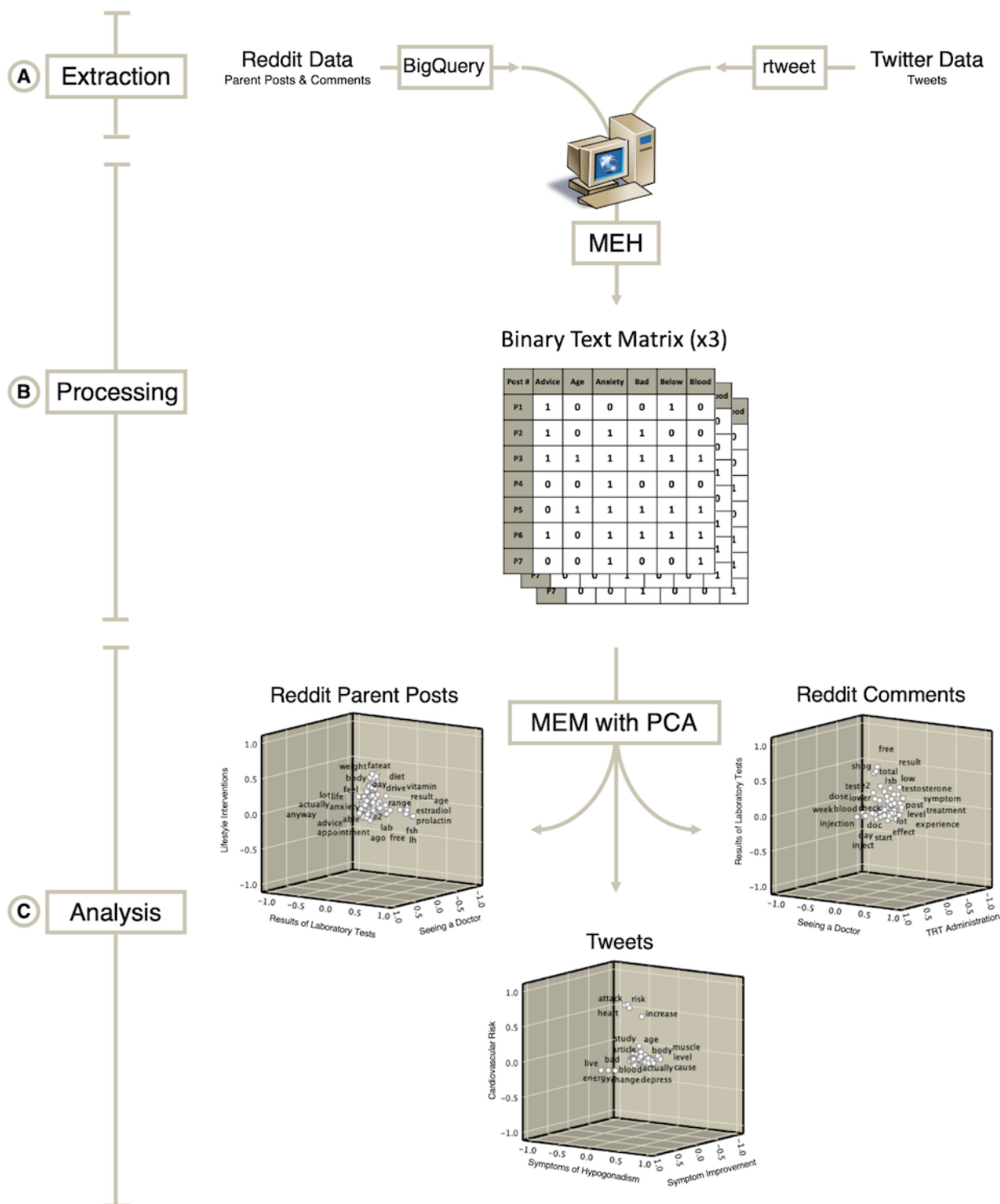
An overview of our methodology is presented in Figure 1. The study comprised three phases: extraction of data from social

media platforms (Figure 1A), automated organization of textual data (Figure 1B), and quantitative analysis of the textual data to identify dominant themes of the text (Figure 1C).

First, we retrospectively processed posts and comments from the Reddit community r/Testosterone from December 2015 through May 2019. Reddit data were extracted using BigQuery (Google LLC), an enterprise data analytics platform, from a dataset uploaded for public use [13] (Figure 1A). We evaluated both parent posts (the main post in a Reddit discussion) and comment posts (submitted in response to a parent post). We applied a word count criterion of >20 words for parent posts to exclude potential spam, deleted text, and posts composed only of links to other websites. As we anticipated the average word count of comment posts to be less, we used a more relaxed word count criterion of >5 words for comment posts.

Next, Twitter data (tweets) were collected prospectively from June through September 2019 using the rtweet application [14], which integrates tweets for processing in RStudio version 1.1.463 (RStudio PBC) (Figure 1A). We extracted tweets containing the terms low testosterone, low T, and testosterone replacement. We applied a word count criterion for tweets (>5 words per tweet), given the character count limitation imposed by the Twitter platform. Retweets (reposts of an identical, previously published tweet) were excluded from analysis.

**Figure 1.** Overview of methods: (A) extraction of Reddit and Twitter data using BigQuery and rtweet, respectively; (B) processing of raw text data using the meaning extraction helper to generate a binary text matrix for each data set; (C) meaning extraction method with principal component analysis generates word clusters for each dataset. Rotated component plots are shown with x-, y-, and z-axes representing the three clusters that capture the greatest variance of the data. MEH: meaning extraction helper; MEM: meaning extraction method; PCA: principal component analysis.



## Natural Language Processing Using the Meaning Extraction Method

Reddit parent posts, Reddit comment posts, and tweets from Twitter were separately subjected to an NLP technique called the meaning extraction method (MEM) [15] with principal component analysis (PCA). MEM/PCA tracks words that cluster

together to derive themes quantitatively [15]. This approach has been previously validated to reveal information about individuals' personalities, communication strategies, and behaviors [16,17].

To automate the MEM, we used the topic modeling application meaning extraction helper version 2 [18] to deconstruct each



post or tweet into its component words. Stop words (eg, articles, prepositions, and transitions) were filtered out. Remaining words were ranked by their frequencies of appearance in each post or tweet (Figure 1B). Words were then subjected to PCA with varimax rotation (Figure 1C) using SPSS Statistics version 25 (IBM Corporation). PCA identified clusters of words that frequently appeared together. Each word was conferred a factor loading, the correlation coefficient between the word and the cluster to which it belonged. Factor loading thresholds of  $>0.20$  are appropriate when performing PCA of text data to capture a sufficient proportion of the variance in the data [19,20]. We assigned a descriptive theme to each cluster based on the words within it.

### Subset Analyses on Key Topics of Interest

Given widespread interest and controversy regarding the potential associations of TRT with cardiovascular disease and prostate cancer risk, we sought to quantitate the appearance of these topics on Reddit and Twitter. Subset analysis was performed to determine the frequencies of the words prostate, cancer, PSA (prostate-specific antigen), heart, attack, stroke, cardiovascular, and death. Furthermore, to identify the degree to which individuals allude to seeking consultation with a health care provider, an additional analysis was performed to determine the frequencies of the relevant terms doctor, urologist, endocrinologist, and appointment.

### Statistical Validity of Principal Component Analysis

To assess applicability of PCA to each dataset, the Kaiser-Meyer-Olkin (KMO) statistic, a measure of sampling adequacy (values  $>0.60$  are adequate), and the Bartlett test for

sphericity, which tests if there are significant correlations among variables of interest, were calculated [21].

### Ethics

Consistent with previous investigations on social media data, this work was exempted by the institutional review board of the University of California, Los Angeles, as it involves publicly available data and does not involve human subjects.

## Results

### Total Number of Posts Extracted From Social Media

From the r/Testosterone community on Reddit, we retrospectively extracted 19,083 parent posts and 218,082 comment posts over the 42-month period of study. After exclusions, 12,665 parent posts and 186,670 comment posts remained. From Twitter, we prospectively extracted 7467 tweets over 4 months; 6659 tweets remained after exclusions.

### Natural Language Processing of Reddit Data

Using MEM for Reddit parent post and comment post data, we identified 5 factors, or thematic word clusters, that included words with factor loadings greater than 0.30 and 0.20, respectively (Tables 1 and 2).

The following themes emerged from NLP of Reddit data: seeing a doctor, results of laboratory tests, administration of TRT, and lifestyle interventions (both parent posts and comment posts); symptoms of hypogonadism (parent posts only); and TRT medications (comment posts only). Table 3 contains representative quotations that feature each Reddit theme. Some quotes have been abridged in the interest of space.

**Table 1.** Thematic clusters, word frequencies, and associated factor loading coefficients derived from the meaning extraction method with principal component analysis of parent posts from the Reddit community r/Testosterone (n=12,665).

Cluster and word	Factor loading coefficient	Frequency
<b>Results of laboratory tests</b>		
LH (luteinizing hormone)	0.72	13.2
FSH (follicle-stimulating hormone)	0.70	11.4
Free	0.67	26.3
Prolactin	0.61	8.8
TSH (thyroid-stimulating hormone)	0.58	7.7
SHBG (sex hormone binding globulin)	0.57	12.3
Total	0.56	24.3
Estradiol	0.54	11.3
Range	0.46	22.1
Result	0.41	26.0
<b>Lifestyle interventions</b>		
Weight	0.55	10.6
Fat	0.55	9.3
Eat	0.53	8.4
Diet	0.50	8.7
Gain	0.49	7.3
Lift	0.48	7.0
Muscle	0.46	9.8
Sleep	0.43	11.3
Gym	0.40	6.7
Body	0.38	11.6
<b>Seeing a doctor</b>		
Doctor	0.43	25.5
Low	0.37	43.3
Told	0.37	8.4
Level	0.35	34.6
Month	0.32	26.4
Appointment	0.32	5.1
Read	0.31	13.1
Endocrinologist	0.31	5.2
Treatment	0.31	7.5
Urologist	0.30	5.7
<b>Testosterone replacement therapy administration</b>		
Week	0.55	40.2
Dose	0.49	13.3
HCG (human chorionic gonadotropin)	0.46	13.5
Protocol	0.42	6.3
Injection	0.42	16.4
Day	0.42	28.4
CYP (cytochrome P450)	0.42	5.9

Cluster and word	Factor loading coefficient	Frequency
Start	0.40	28.7
E2 (estradiol)	0.38	12.3
Twice	0.36	5.8
<b>Symptoms of hypogonadism</b>		
Fog	0.76	5.0
Brain	0.75	5.7
Depress	0.42	11.7
Symptom	0.40	21.0
Anxiety	0.38	7.9
Libido	0.36	16.0
Erection	0.35	8.4
Sex	0.34	14.7
Energy	0.32	10.7
Drive	0.31	9.1

**Table 2.** Thematic clusters, word frequencies, and associated factor loading coefficients derived from the meaning extraction method with principal component analysis of comment posts from the Reddit community r/Testosterone (n=186,670).

Cluster and word	Factor loading coefficient	Frequency
<b>Seeing a doctor</b>		
Doctor	0.39	9.0
TRT (testosterone replacement therapy)	0.38	15.3
Treatment	0.33	2.7
Symptom	0.32	5.9
Life	0.31	4.0
People	0.30	5.9
Issue	0.29	5.3
Help	0.27	6.0
Cause	0.27	5.6
Hormone	0.26	3.9
Prescribe	0.26	2.6
Experience	0.25	2.8
<b>Results of laboratory tests</b>		
Free	0.64	4.9
Total	0.58	4.9
SHBG (sex hormone binding globulin)	0.54	4.6
Range	0.44	5.9
Test	0.42	17.9
Low	0.38	15.4
LH (luteinizing hormone)	0.34	2.8
Lab	0.33	4.5
E2 (estradiol)	0.31	8.0
Normal	0.28	5.6
Result	0.26	4.3
<b>Testosterone replacement therapy administration</b>		
Week	0.60	13.0
Day	0.45	9.7
Injection	0.43	5.8
Dose	0.42	7.6
Inject	0.35	3.9
Start	0.33	9.1
Protocol	0.32	2.8
Feel	0.31	11.6
Month	0.31	7.2
Time	0.30	9.4
<b>Lifestyle interventions</b>		
Fat	0.57	3.0
Eat	0.55	3.1
Diet	0.53	3.4
Weight	0.50	3.2

Cluster and word	Factor loading coefficient	Frequency
Muscle	0.41	2.8
Body	0.34	4.7
Sleep	0.33	3.5
<b>Testosterone replacement therapy medications</b>		
Increase	0.40	4.2
Estrogen	0.37	3.3
Effect	0.36	4.2
Lower	0.32	4.7
Testosterone	0.30	11.9
Clomid	0.26	3.7
HCG (human chorionic gonadotropin)	0.26	5.9



**Table 3.** Representative quotations for each theme derived from the meaning extraction method. Asterisks are part of the quotations and do not refer to anything in the table.

Data source and theme	Representative quotation
<b>Reddit parent posts</b>	
<b>Results of laboratory tests</b>	<p>Here's what came up:</p> <p>FSH<sup>a</sup> 2.1 (1.5-12.4)</p> <p>LH<sup>b</sup> 5.6 (1.7-8.6)</p> <p>Prolactin 15.25 (4.04-15.2)*</p> <p>T, total<sup>c</sup> 311.1 (249-836)*</p> <p>SHBG<sup>d</sup> 33.3 (16.5-55.9)</p> <p>Free testosterone index 32.43 (35.0-92.6)*</p> <p>shbg and dheia still pending. I had to get these results because i have an appointment with neurosurgeon soon and he will need the labs and mri<sup>e</sup>.</p>
<b>Lifestyle interventions</b>	<p>Have been eating super clean. Working with a dietitian/personal trainer. Was dieting mostly high protein / low fat / low carb</p> <p>I work out all the time lifting heavy weights, 3 or 4 times a week on average. I eat a good diet, take my zinc, vitamin D, and get in my fats and essential fats.</p>
<b>Seeing a doctor</b>	<p>I know several people on trt<sup>f</sup>, but they all have the same doc...you walk in, tell him you want to get bigger, stronger, and faster, pay out of pocket for his blood test then buy your meds from his attached pharmacy. That's not what I want. I want to find out what's wrong without a preconceived bias.</p> <p>So I go to the appointment. And the specialist I saw (a urologist) said he wasn't the guy to see about this issue, and ended up referring me to another specialist.</p>
<b>Testosterone replacement therapy administration</b>	<p>T cyp<sup>g</sup> 200 mg/ml - 0.32 mL IM/SQ twice weekly (~130 mg/week)</p> <p>HCG<sup>h</sup> 500 IU SQ twice weekly to prevent testicular atrophy</p> <p>No AI<sup>i</sup> - low E2<sup>j</sup>, monitor</p> <p>DHEA<sup>k</sup> 25 mg every night</p> <p>So I don't know what to do? Take my AI and hope that my E2 is high? Or keep not taking my AI and hope things will get better?</p> <p>I literally can't hold out a week to get another blood test and also I can't afford it right now.</p>
<b>Symptoms of hypogonadism</b>	<p>All the normal symptoms: brain fog, mood swings, low libido, erectile dysfunction, inability to add muscle at the gym despite working out 3x a week.</p> <p>Symptoms: brain fog, very low energy level, lifelessness-zombie feeling most days, very lethargic, mood swings, easy to get angry, grumpy and annoyed at earliest, no libido/sex drive, ED<sup>l</sup>—less frequency, less powerful, minimal to no erections during sex, softer (haven't had sex in years)</p>
<b>Reddit comment posts</b>	
<b>Seeing a doctor</b>	<p>Many doctors—especially PCPs<sup>m</sup>—are not fluent in the endocrine system. They aren't supposed to be. Going to your primary care physician for hormone questions is a mistake. If you knew you had heart issues, wouldn't you go to a cardiologist?</p> <p>My PCP looked super confused and clueless as to what he was supposed to do for me. Doc made me do two more labs fasting to confirm then he referred me out to an endocrinologist. The endo made me do three more fasting labs and a testicular ultrasound to confirm.</p>
<b>Results of laboratory tests</b>	

Data source and theme	Representative quotation
<b>Testosterone replacement therapy administration</b>	<p>Honestly I don't think testosterone is your problem based on Sept 10th 2015 blood results. You have decent midrange total, and free testosterone. SHBG bounces around, so maybe it's a testing error.</p> <p>198 is low as hell for your dad, and even 450 for him would be low. Yours is lowish, but you have definite symptoms.</p> <p>75 mg E5D<sup>n</sup> (105 mg per week). Doesn't require an AI, doesn't give me side effects. I am at ~700 on trough days and feel pretty damn good.</p> <p>I had just moved to a standard TRT dose of Test Cyp, 100 mg/week. At 5'11", 172 lbs, and 17% body fat, taking 1 mg of Arimidex every day tanked my E2. Dropping down to 0.25 mg Arimidex once a week had the same effect.</p>
<b>Lifestyle interventions</b>	<p>Eat good food, lift heavy, and get sleep. Repeat for two years.</p> <p>TRT will not turn you into a bodybuilder. It may tone you a little bit (if everything is in check). But just saying "I eat good" literally means nothing. What are your macros? What's your diet? Etc?</p>
<b>Testosterone replacement therapy medications</b>	<p>HCG is a water-based peptide hormone that can be injected to replace the lost LH hormone that TRT shuts down. Without hCG, the LH receptors in the testes are no longer getting activated. The results: the testes shrink.</p> <p>Clomiphene. What a double-edged sword. First, Clomid will certainly have an effect on your testosterone levels. Usually, it is doses substantially higher than 12.5 mgs daily.</p>
<b>Tweets</b>	
<b>Symptoms of hypogonadism</b>	<p>Keeping your hormone levels up is a crucial part of #health. Low T can lead to all types of adverse effects: - weight gain/belly fat - #LowEnergy - low sex drive</p> <p>This, in turn, causes a lower sex drive, depression, reduced muscle mass, and low levels of energy. Erectile dysfunction is another symptom.</p>
<b>Cardiovascular risk</b>	<p>#Testosterone Replacement Therapy Lowers Heart Attack Risk</p> <p>Aging men with low testosterone levels who take testosterone replacement therapy (TRT) are at a slightly greater risk of experiencing an ischemic stroke</p>
<b>Symptom improvement</b>	<p>Starting testosterone replacement therapy and thyroid medication at the same time is quite the 1-2 punch to the system. Endless energy, great sleep, and able to lift weights heavier and longer.</p> <p>"My energy is back": how testosterone replacement therapy is changing men's lives</p>
<b>Derogatory comments and insults</b>	<p>That little cuck should be the poster boy for low T supplements</p>

Data source and theme	Representative quotation
	I was called effeminate and a low testosterone beta here for defending women's rights.

<sup>a</sup>FSH: follicle-stimulating hormone.

<sup>b</sup>LH: luteinizing hormone.

<sup>c</sup>T, total: total testosterone.

<sup>d</sup>SHBG: sex hormone binding globulin.

<sup>e</sup>MRI: magnetic resonance imaging.

<sup>f</sup>TRT: testosterone replacement therapy.

<sup>g</sup>CYP: cytochrome P450.

<sup>h</sup>HCG: human chorionic gonadotropin.

<sup>i</sup>AI: aromatase inhibitor.

<sup>j</sup>E2: estradiol.

<sup>k</sup>DHEA: dehydroepiandrosterone.

<sup>l</sup>ED: erectile dysfunction.

<sup>m</sup>PCP: primary care physician.

<sup>n</sup>E5D: every 5 days.

The highest frequency word occurrences among parent posts as determined by PCA were low (5484/12,665 [43.30%] of posts), week (5092/12,665, 40.20%), level (4382/12,665, 34.60%), and start (3635/12,665, 28.70%). Among comment posts, the highest frequency word occurrences were test (33,414/186,670, 17.90%), low (28,747/186,670, 15.40%), TRT (28,561/186,670, 15.30%), and week (24,267/186,670, 13.00%).

Parent post and comment post PCA accounted for 15.45% (1957/12,665) and 13.84% (25,835/186,670) of the total variance, respectively. KMO statistic was 0.91 for Reddit parent post data and 0.80 for Reddit comment post data, with Bartlett test <0.01, indicating that the Reddit data were appropriate for factor analysis using PCA.

### Natural Language Processing of Twitter Data

Similarly, MEM for Twitter data identified 4 factors, or thematic word clusters, with factor loadings greater than 0.25 (Table 4). The following themes emerged from NLP of tweets: symptoms of hypogonadism, cardiovascular risk, symptom improvement, and derogatory comments and insults.

The highest frequency word occurrences among tweets as determined by PCA were level (693/6659, 10.40%), male (426/6659, 6.40%), sex (213/6659, 3.20%), and increase (200/6659, 3.00%). Twitter PCA accounted for 9.01% (600/6659) of the total variance. KMO statistic was 0.61 for Twitter data, with Bartlett test <0.01, indicating that the Twitter data were appropriate for factor analysis using PCA. Of note, other studies using MEM/PCA have reported similar percentages of variance as those determined in our analysis of Reddit and Twitter data [22,23].

**Table 4.** Thematic clusters, word frequencies, and associated factor loading coefficients derived from the meaning extraction method with principal component analysis of tweets about low testosterone, low T, or testosterone replacement on Twitter (n=6659).

Cluster and word	Factor loading coefficient	Frequency
<b>Symptoms of hypogonadism</b>		
Muscle	0.54	1.9
Mass	0.48	1.2
Sex	0.41	3.2
Libido	0.39	1.0
Level	0.36	10.4
Drive	0.35	1.8
Fat	0.32	1.3
Hormone	0.31	2.7
Body	0.28	2.3
Weight	0.27	1.1
<b>Cardiovascular risk</b>		
Heart	0.76	1.1
Attack	0.76	1.1
Risk	0.73	2.1
Increase	0.62	3.0
<b>Symptom improvement</b>		
Change	0.69	2.2
Energy	0.69	2.9
Live	0.69	2.0
Life	0.21	2.1
<b>Derogatory comments and insults</b>		
Boy	0.58	2.3
Soy	0.56	2.4
Beta	0.42	2.8
Cuck	0.29	1.0
Male	0.23	6.4
Girl	0.22	1.0
Little	0.22	1.3

### Word Occurrences on Key Topics of Interest

Subset analysis was performed to determine word occurrence frequencies in three key topics of interest that relate to TRT: prostate cancer risk, cardiovascular disease risk, and seeking consultation with a health care professional. These data are presented in [Table 5](#).

In brief, over 1% of Reddit parent posts contain the terms prostate (143/12,665, 1.13%), cancer (143/12,665, 1.13%), PSA (210/12,665, 1.66%), or heart (175/12,665, 1.38%). Over a quarter of Reddit parent posts contain the term doctor (3235/12,665, 25.54%), while over 5% of parent posts refer to either a urologist (732/12,665, 5.78%) or endocrinologist (657/12,665, 5.19%). Frequencies of these terms were higher among Reddit posts than among tweets from Twitter.

**Table 5.** Subset analysis of word frequencies related to prostate cancer risk, cardiovascular disease risk, and seeking a health care consultation.

Concern associated with testosterone replacement therapy	Word frequency (%)		
	Reddit parent post	Reddit comment post	Tweet
<b>Prostate cancer risk</b>			
Prostate	1.13	0.36	0.63
Cancer	1.13	0.56	0.87
PSA <sup>a</sup>	1.66	0.25	0.06
<b>Cardiovascular disease risk</b>			
Heart	1.38	0.61	1.14
Attack	0.89	0.32	1.14
Stroke	0.23	0.13	0.86
Cardiovascular	0.13	0.13	0.44
Death	0.24	0.19	0.21
<b>Seeking health care consultation</b>			
Doctor	25.54	9.00	1.77
Urologist	5.78	1.36	0.20
Endocrinologist	5.19	0.97	0.17
Appointment	5.13	0.74	0.39

<sup>a</sup>PSA: prostate-specific antigen.

## Discussion

### Principal Findings

NLP techniques applied to unfiltered discussions on Reddit and Twitter offer a useful framework for understanding patient priorities outside the doctor's office. We found that men largely turn to social media to learn about symptoms of low testosterone, interpretation of personal lab results, practicalities of TRT, and body changes with treatment. Notably, cardiovascular risk was a major discussion theme, echoing concerns among prescribers, who may be deterred by continued ambiguity despite the publication of the Testosterone Trials. Although NLP analysis did not reveal prostate cancer as a notable theme, a number of posts included text related to this topic, suggesting that this may represent an important discussion point for a subset of online discussions related to TRT.

Our results underscore that patients are searching for medical guidance related to hypogonadism on social media, an environment where anecdotes predominate and advertising often masquerades as medical advice [24]. TRT prescriptions have risen almost 4-fold over the last two decades, which can be attributed, in part, to off-label indications and direct-to-consumer advertising [25]. Even beyond standard TRT, testosterone-boosting supplements with minimal data to support their efficacy are aggressively marketed and readily available online [26]. But still, social media represents an enormous opportunity for the medical community to improve how we engage with our patients and to do so in a meaningful and impactful way. Potential interventions that may inoculate against coercive direct-to-consumer marketing practices include disseminating high-quality, open-access information related to

hypogonadism. For example, Halpern et al [27] recently published a JAMA Patient Page article on hypogonadism. This single-page handout written in easily accessible language includes an infographic highlighting symptoms of hypogonadism and potential adverse effects of TRT, in addition to information related to etiology of hypogonadism and a discussion of potential cardiovascular and prostate cancer risks associated with TRT—all topics that emerged as major themes of discussion from our data.

Social media platforms, including Reddit and Twitter, create a space for patients not only to obtain answers to questions that they are either uncomfortable or unwilling to ask in a face-to-face clinical setting but also to connect with others going through similar experiences. However, not all health-related discussions online are productive. Twitter featured the theme of derogatory comments and insults, highlighting an undertone of stigma, which may compound existing barriers preventing men from accessing care [28]. In contrast, the seeing a doctor theme only emerged on Reddit, with more than 25% of parent posts mentioning the word doctor, compared with less than 2% on Twitter. This may reflect inherent differences among the two social media platforms, as Twitter is constrained by a strict character count limitation and is overall less anonymous, with discussants frequently using their true identities in their display usernames and account photos.

Although clinician engagement with the online hypogonadism community will become increasingly important in the coming years, improving the in-office clinical experience of our patients cannot be overemphasized. Our data reveal that many of the online discussions featured personal questions related to interpretation of lab results. This is consistent with a previous study exploring Reddit discussions of male factor infertility,



where nearly 20% of all posts featured a question related to personal semen analysis results [29]. Such discussions related to lab results cannot be addressed by disseminating a primer on hypogonadism and TRT, but instead demand the expertise of a clinician trained in managing male endocrinology and the related sexual, reproductive, and psychological comorbidities. Creating an in-office experience where men feel comfortable and safe to ask their questions and voice their concerns should be a priority for any outpatient clinical setting, but especially one that caters to men with suspected hypogonadism. Both outpatient primary care settings and urological outpatient clinics can learn from the success of the emerging multidisciplinary men's health clinic [30].

Here we offer valuable insight into primarily patient concerns in a forum that allows for honest and unfiltered patient feedback as it relates to these discussants' experiences with hypogonadism. Clinically, these data highlight that patients worry most about comorbidities, lifestyle factors impacted by low testosterone, and treatment options. While other aspects of hypogonadism can be discussed, these data highlight the most salient hypogonadism-related concerns for our patients. Additionally, this study can further improve on patients' in-office experiences by informing how physicians can lead discussions to highlight aspects of low testosterone that patients may feel are not being adequately addressed.

### Limitations

Our study is not without limitations. Although NLP techniques allowed us to analyze a large volume of discrete social media posts, generalizability of MEM is limited by the absence of

contextual valence (positivity or negativity). However, this does not impair overall thematic identification. Additionally, discussants who turn to social media for health care information may be different with respect to demographics, health care priorities, and information preferences compared with those who do not; our results should therefore be interpreted within this context [31]. It should also be noted that some individuals use social media as a platform to vent about their experiences with health care professionals as they relate to hypogonadism care. This is an important distinction to make because it may not necessarily represent a lack of communication between patients and their physician but rather a discussant's opportunity to share. Future studies may consider investigating to other Reddit communities, expanding Twitter search terms, or exploring other social media platforms.

### Conclusions

This study represents the first evaluation of the social media landscape surrounding hypogonadism and TRT using NLP techniques. Our analysis of more than 200,000 discrete social media posts revealed dominant themes of discussion, which may inform how physicians evaluate and counsel men with hypogonadism. Understanding the complex internet landscape of hypogonadism discussions represents the first step in creating well-informed and clinically meaningful change. Although physicians traditionally limit their practices to within their clinic walls, the ubiquity of social media demands that physicians engage patients where they are, including online. Practicing physicians may do well to bring up online discussions during clinic consultations, to pull back the curtain and dispel myths.

### Acknowledgments

SVE is supported by a Research Scholar Award from the Urology Care Foundation and American Urological Association. These organizations played no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Authors' Contributions

VO was responsible for concept and design; acquisition, analysis, or interpretation of data; drafting the manuscript; critical revision of the manuscript for important intellectual content; statistical analysis; and approval of the manuscript. TJ was responsible for concept and design; acquisition, analysis, or interpretation of data; critical revision of the manuscript for important intellectual content; and approval of the manuscript. JNM was responsible for concept and design; acquisition, analysis, or interpretation of data; critical revision of the manuscript for important intellectual content; administrative, technical, or material support; and approval of the manuscript. SVE was responsible for concept and design; acquisition, analysis, or interpretation of data; drafting the manuscript; critical revision of the manuscript for important intellectual content; administrative, technical, or material support; supervision; and approval of the manuscript.

### Conflicts of Interest

SVE serves as a consultant for Metuchen Pharmaceuticals. JNM serves as a consultant for Antares Pharma, Boston Scientific Corporation, and Endo Pharmaceuticals. The remaining authors report no conflicts of interest.

### References

1. Snyder PJ, Bhasin S, Cunningham GR, Matsumoto AM, Stephens-Shields AJ, Cauley JA, et al. Lessons from the Testosterone Trials. *Endocr Rev* 2018 Jun 01;39(3):369-386 [FREE Full text] [doi: [10.1210/er.2017-00234](https://doi.org/10.1210/er.2017-00234)] [Medline: [29522088](https://pubmed.ncbi.nlm.nih.gov/29522088/)]
2. Bhasin S, Ellenberg SS, Storer TW, Basaria S, Pahor M, Stephens-Shields AJ, et al. Effect of testosterone replacement on measures of mobility in older men with mobility limitation and low testosterone concentrations: secondary analyses of the Testosterone Trials. *Lancet Diabetes Endocrinol* 2018 Nov;6(11):879-890 [FREE Full text] [doi: [10.1016/S2213-8587\(18\)30171-2](https://doi.org/10.1016/S2213-8587(18)30171-2)] [Medline: [30366567](https://pubmed.ncbi.nlm.nih.gov/30366567/)]

3. Mohler ER, Ellenberg SS, Lewis CE, Wenger NK, Budoff MJ, Lewis MR, et al. The effect of testosterone on cardiovascular biomarkers in the Testosterone Trials. *J Clin Endocrinol Metab* 2018 Feb 01;103(2):681-688 [FREE Full text] [doi: [10.1210/jc.2017-02243](https://doi.org/10.1210/jc.2017-02243)] [Medline: [29253154](https://pubmed.ncbi.nlm.nih.gov/29253154/)]
4. Resnick SM, Matsumoto AM, Stephens-Shields AJ, Ellenberg SS, Gill TM, Shumaker SA, et al. Testosterone treatment and cognitive function in older men with low testosterone and age-associated memory impairment. *JAMA* 2017 Feb 21;317(7):717-727 [FREE Full text] [doi: [10.1001/jama.2016.21044](https://doi.org/10.1001/jama.2016.21044)] [Medline: [28241356](https://pubmed.ncbi.nlm.nih.gov/28241356/)]
5. Budoff MJ, Ellenberg SS, Lewis CE, Mohler ER, Wenger NK, Bhasin S, et al. Testosterone treatment and coronary artery plaque volume in older men with low testosterone. *JAMA* 2017 Feb 21;317(7):708-716 [FREE Full text] [doi: [10.1001/jama.2016.21043](https://doi.org/10.1001/jama.2016.21043)] [Medline: [28241355](https://pubmed.ncbi.nlm.nih.gov/28241355/)]
6. Cunningham GR, Stephens-Shields AJ, Rosen RC, Wang C, Bhasin S, Matsumoto AM, et al. Testosterone treatment and sexual function in older men with low testosterone levels. *J Clin Endocrinol Metab* 2016 Aug;101(8):3096-3104 [FREE Full text] [doi: [10.1210/jc.2016-1645](https://doi.org/10.1210/jc.2016-1645)] [Medline: [27355400](https://pubmed.ncbi.nlm.nih.gov/27355400/)]
7. Snyder PJ, Bhasin S, Cunningham GR, Matsumoto AM, Stephens-Shields AJ, Cauley JA, Testosterone Trials Investigators. Effects of testosterone treatment in older men. *N Engl J Med* 2016 Feb 18;374(7):611-624 [FREE Full text] [doi: [10.1056/NEJMoa1506119](https://doi.org/10.1056/NEJMoa1506119)] [Medline: [26886521](https://pubmed.ncbi.nlm.nih.gov/26886521/)]
8. Nobles AL, Leas EC, Althouse BM, Dredze M, Longhurst CA, Smith DM, et al. Requests for diagnoses of sexually transmitted diseases on a social media platform. *JAMA* 2019 Nov 05;322(17):1712-1713. [doi: [10.1001/jama.2019.14390](https://doi.org/10.1001/jama.2019.14390)] [Medline: [31688875](https://pubmed.ncbi.nlm.nih.gov/31688875/)]
9. fox S, Duggan M. Health Online 2013. Washington: Pew Internet and American Life Project; 2013. URL: [https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/PIP\\_HealthOnline.pdf](https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf) [accessed 2020-09-22]
10. Cassis C, Bassett J. Reddit's year in review. 2018. URL: <https://redditblog.com/2018/12/04/reddit-year-in-review-2018/> [accessed 2019-10-06]
11. subreddit r/Testosterone. URL: <https://www.reddit.com/r/Testosterone/> [accessed 2020-09-22]
12. Shaban H. Twitter reveals its daily active user numbers for the first time. Washington Post. 2019 Feb 07. URL: <https://www.washingtonpost.com/technology/2019/02/07/twitter-reveals-its-daily-active-user-numbers-first-time/> [accessed 2019-10-06]
13. subreddit directory. URL: <https://files.pushshift.io/reddit/> [accessed 2020-09-22]
14. Kearney M. rtweet: collecting Twitter data (R package version 0.6.7). 2018. URL: <https://cran.r-project.org/package=rtweet> [accessed 2020-09-22]
15. Chung C, Pennebaker J. Revealing dimensions of thinking in open-ended self-descriptions: an automated meaning extraction method for natural language. *J Res Pers* 2008 Feb;42(1):96-132 [FREE Full text] [doi: [10.1016/j.jrp.2007.04.006](https://doi.org/10.1016/j.jrp.2007.04.006)] [Medline: [18802499](https://pubmed.ncbi.nlm.nih.gov/18802499/)]
16. Boyd RL, Pennebaker JW. Did Shakespeare write double falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychol Sci* 2015 May;26(5):570-582. [doi: [10.1177/0956797614566658](https://doi.org/10.1177/0956797614566658)] [Medline: [25854277](https://pubmed.ncbi.nlm.nih.gov/25854277/)]
17. Barrett A, Murphy M, Blackburn K. "Playing hooky" health messages: apprehension, impression management, and deception. *Health Commun* 2018 Mar;33(3):326-337. [doi: [10.1080/10410236.2016.1266578](https://doi.org/10.1080/10410236.2016.1266578)] [Medline: [28059569](https://pubmed.ncbi.nlm.nih.gov/28059569/)]
18. Boyd R. MEH: Meaning Extraction Helper (Version 2.1.06). 2018. URL: <https://www.ryanboyd.io/software/meh/> [accessed 2020-09-22]
19. Blackburn KG, Yilmaz G, Boyd RL. Food for thought: exploring how people think and talk about food online. *Appetite* 2018 Apr 01;123:390-401. [doi: [10.1016/j.appet.2018.01.022](https://doi.org/10.1016/j.appet.2018.01.022)] [Medline: [29407531](https://pubmed.ncbi.nlm.nih.gov/29407531/)]
20. Jiang T, Osadchiy V, Mills JN, Eleswarapu SV. Is it all in my head? Self-reported psychogenic erectile dysfunction and depression are common among young men seeking advice on social media. *Urology* 2020 May 11:1. [doi: [10.1016/j.urology.2020.04.100](https://doi.org/10.1016/j.urology.2020.04.100)] [Medline: [32437776](https://pubmed.ncbi.nlm.nih.gov/32437776/)]
21. Pett M, Lackey N, Sullivan J. Making Sense of Factor Analysis. Thousand Oaks: Sage Publications; 2003.
22. Wolf M, Chung CK, Kordy H. Inpatient treatment to online aftercare: e-mailing themes as a function of therapeutic outcomes. *Psychother Res* 2010 Jan;20(1):71-85. [doi: [10.1080/10503300903179799](https://doi.org/10.1080/10503300903179799)] [Medline: [19821186](https://pubmed.ncbi.nlm.nih.gov/19821186/)]
23. Stanton AM, Boyd RL, Pulverman CS, Meston CM. Determining women's sexual self-schemas through advanced computerized text analysis. *Child Abuse Negl* 2015 Aug;46:78-88 [FREE Full text] [doi: [10.1016/j.chiabu.2015.06.003](https://doi.org/10.1016/j.chiabu.2015.06.003)] [Medline: [26146161](https://pubmed.ncbi.nlm.nih.gov/26146161/)]
24. Kravitz RL. Direct-to-consumer advertising of androgen replacement therapy. *JAMA* 2017 Mar 21;317(11):1124-1125. [doi: [10.1001/jama.2017.1364](https://doi.org/10.1001/jama.2017.1364)] [Medline: [28324072](https://pubmed.ncbi.nlm.nih.gov/28324072/)]
25. Bandari J, Ayyash OM, Emery SL, Wessel CB, Davies BJ. Marketing and testosterone treatment in the USA: a systematic review. *Eur Urol Focus* 2017 Oct;3(4-5):395-402. [doi: [10.1016/j.euf.2017.10.016](https://doi.org/10.1016/j.euf.2017.10.016)] [Medline: [29174614](https://pubmed.ncbi.nlm.nih.gov/29174614/)]
26. Balasubramanian A, Thirumavalavan N, Srivatsav A, Yu J, Lipshultz LI, Pastuszak AW. Testosterone imposters: an analysis of popular online testosterone boosting supplements. *J Sex Med* 2019 Feb;16(2):203-212 [FREE Full text] [doi: [10.1016/j.jsxm.2018.12.008](https://doi.org/10.1016/j.jsxm.2018.12.008)] [Medline: [30770069](https://pubmed.ncbi.nlm.nih.gov/30770069/)]
27. Halpern JA, Brannigan RE. Testosterone deficiency. *JAMA* 2019 Sep 17;322(11):1116. [doi: [10.1001/jama.2019.9290](https://doi.org/10.1001/jama.2019.9290)] [Medline: [31529009](https://pubmed.ncbi.nlm.nih.gov/31529009/)]
28. Gott M, Hinchliff S. Barriers to seeking treatment for sexual problems in primary care: a qualitative study with older people. *Fam Pract* 2003 Dec;20(6):690-695. [doi: [10.1093/fampra/cm612](https://doi.org/10.1093/fampra/cm612)] [Medline: [14701894](https://pubmed.ncbi.nlm.nih.gov/14701894/)]

29. Osadchiy V, Mills JN, Eleswarapu SV. Understanding patient anxieties in the social media era: qualitative analysis and natural language processing of an online male infertility community. *J Med Internet Res* 2020 Mar 10;22(3):e16728 [FREE Full text] [doi: [10.2196/16728](https://doi.org/10.2196/16728)] [Medline: [32154785](https://pubmed.ncbi.nlm.nih.gov/32154785/)]
30. Houman JJ, Eleswarapu SV, Mills JN. Current and future trends in men's health clinics. *Transl Androl Urol* 2020 Mar;9(Suppl 2):S116-S122 [FREE Full text] [doi: [10.21037/tau.2019.08.33](https://doi.org/10.21037/tau.2019.08.33)] [Medline: [32257852](https://pubmed.ncbi.nlm.nih.gov/32257852/)]
31. Koch-Weser S, Bradshaw YS, Gualtieri L, Gallagher SS. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. *J Health Commun* 2010;15 Suppl 3:279-293. [doi: [10.1080/10810730.2010.522700](https://doi.org/10.1080/10810730.2010.522700)] [Medline: [21154099](https://pubmed.ncbi.nlm.nih.gov/21154099/)]

## Abbreviations

**KMO:** Kaiser-Meyer-Olkin statistic  
**MEM:** meaning extraction method  
**NLP:** natural language processing  
**PCA:** principal component analysis  
**PSA:** prostate-specific antigen  
**TRT:** testosterone replacement therapy

*Edited by G Eysenbach; submitted 12.06.20; peer-reviewed by A Balasubramanian, X Ding; comments to author 04.07.20; revised version received 27.07.20; accepted 03.08.20; published 07.10.20.*

### *Please cite as:*

Osadchiy V, Jiang T, Mills JN, Eleswarapu SV

*Low Testosterone on Social Media: Application of Natural Language Processing to Understand Patients' Perceptions of Hypogonadism and Its Treatment*

*J Med Internet Res* 2020;22(10):e21383

URL: <https://www.jmir.org/2020/10/e21383>

doi: [10.2196/21383](https://doi.org/10.2196/21383)

PMID: [33026354](https://pubmed.ncbi.nlm.nih.gov/33026354/)

©Vadim Osadchiy, Tommy Jiang, Jesse Nelson Mills, Sriram Venkata Eleswarapu. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 07.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis

Nicolò Gozzi<sup>1</sup>, MSc; Michele Tizzani<sup>2</sup>, PhD; Michele Starnini<sup>2</sup>, PhD; Fabio Ciulla<sup>3</sup>, PhD; Daniela Paolotti<sup>2</sup>, PhD; André Panisson<sup>2</sup>, PhD; Nicola Perra<sup>1</sup>, PhD

<sup>1</sup>University of Greenwich, London, United Kingdom

<sup>2</sup>ISI Foundation, Torino, Italy

<sup>3</sup>Quid Inc, San Francisco, CA, United States

**Corresponding Author:**

Nicolò Gozzi, MSc

University of Greenwich

Old Royal Naval College

Park Row

London, SE10 9LS

United Kingdom

Phone: 44 020 8331 8000

Email: [n.gozzi@gre.ac.uk](mailto:n.gozzi@gre.ac.uk)

## Abstract

**Background:** The exposure and consumption of information during epidemic outbreaks may alter people's risk perception and trigger behavioral changes, which can ultimately affect the evolution of the disease. It is thus of utmost importance to map the dissemination of information by mainstream media outlets and the public response to this information. However, our understanding of this exposure-response dynamic during the COVID-19 pandemic is still limited.

**Objective:** The goal of this study is to characterize the media coverage and collective internet response to the COVID-19 pandemic in four countries: Italy, the United Kingdom, the United States, and Canada.

**Methods:** We collected a heterogeneous data set including 227,768 web-based news articles and 13,448 YouTube videos published by mainstream media outlets, 107,898 user posts and 3,829,309 comments on the social media platform Reddit, and 278,456,892 views of COVID-19-related Wikipedia pages. To analyze the relationship between media coverage, epidemic progression, and users' collective web-based response, we considered a linear regression model that predicts the public response for each country given the amount of news exposure. We also applied topic modelling to the data set using nonnegative matrix factorization.

**Results:** Our results show that public attention, quantified as user activity on Reddit and active searches on Wikipedia pages, is mainly driven by media coverage; meanwhile, this activity declines rapidly while news exposure and COVID-19 incidence remain high. Furthermore, using an unsupervised, dynamic topic modeling approach, we show that while the levels of attention dedicated to different topics by media outlets and internet users are in good accordance, interesting deviations emerge in their temporal patterns.

**Conclusions:** Overall, our findings offer an additional key to interpret public perception and response to the current global health emergency and raise questions about the effects of attention saturation on people's collective awareness and risk perception and thus on their tendencies toward behavioral change.

(*J Med Internet Res* 2020;22(10):e21597) doi:[10.2196/21597](https://doi.org/10.2196/21597)

**KEYWORDS**

social media; news coverage; digital epidemiology; infodemiology; infoveillance; infodemic; digital epidemiology; data science; topic modeling; pandemic; COVID-19; Reddit; Wikipedia; information; response; risk perception; behavior

## Introduction

### Background

*In the next influenza pandemic, be it now or in the future, be the virus mild or virulent, the single most important weapon against the disease will be a vaccine. The second most important will be communication.*

This evocative sentence was written in May 2009 by John M Barry [1] in the early phases of what would soon become the 2009 H1N1 pandemic. In his essay, Barry summarized the mishandling of the deadly 1918 Spanish influenza, highlighting the importance of precise, effective, and honest information at the onset of health crises.

Eleven years later, the world is facing another pandemic. The cause is not a novel strain of influenza; however, unfortunately, Barry's words are still extremely relevant. In fact, as SARS-CoV-2 spreads worldwide and a vaccine may still be far in the future, the most important weapons to reduce the burden of the disease are nonpharmaceutical interventions [2,3]. Social distancing has become paramount, gatherings have been cancelled, and mobility within and across countries has been dramatically reduced. While these measures have been enforced to different extents across nations, they all rely on compliance. The effectiveness of these measures is linked to risk and susceptibility perception [4]; thus, the information to which citizens are exposed is of fundamental importance.

History repeats itself, and humanity appears to not be able to learn from its past mistakes. As happened in 1918, despite early evidence from China [5,6], the virus was first equated by many people with common seasonal influenza. As in 1918, many national and regional governments organized campaigns aimed at encouraging social activities (and thus local economies) while actively attempting to convince people that their cities were safe and that the spread of the disease was isolated in faraway locations. For example, the hashtag #MilanoNonSiFerma ("Milan does not stop") was coined to invite citizens in Milan to go out and live normally, while free aperitifs were offered in Venice. In hindsight, of course, it is easy to criticize the initial response in Italy. In fact, the country was one of the first to experience rapid growth of hospitalization [7]. However, the Mayor of London, 12 days before the national lockdown and a few days after the extension of the *cordon sanitaire* to the entire country in Italy, affirmed via his official Facebook page [8] that "we should carry on doing what we've been doing." More generally, in several western countries, news reports from other countries reporting concerning epidemic outbreaks were not considered to be relevant to the local situation. This initial phase aimed at conveying low local risk and boosting confidence in national safety was repeated, at different times, across countries. A series of surveys conducted in late February provide a glimpse of the possible effects of these approaches. These surveys report that citizens of several European countries, despite the grim news coming from Asia, were overly optimistic about the health emergency, placing their risk of infection at 1% or less [9]. As in 1918, countries that reacted earlier rather than later were

better able to control the virus, with significantly fewer infections [10-14].

History repeats itself; however, the context is often radically different. In 1918, news circulated slowly via newspapers, controlled by editorial choices; of course, news also spread by word of mouth. In 2009, we witnessed the first pandemic in the social media era. Newspapers and television were still very important sources of information; however, Twitter, Facebook, YouTube, and Wikipedia started to become relevant for decentralized news consumption, boosting of peer discussions, and spreading of misinformation. Currently, these platforms and websites are far more popular and integral parts of society, and they are instrumental sources of national and international news circulation. Together with traditional news media, these platforms and websites are the principal sources of information for the public. As such, they are fundamental drivers of people's perception and opinions and thus of their behaviors. This is particularly relevant for health issues. For example, approximately 60% of adults in the United States consulted web-based sources to gather health information [15].

Furthermore, some platforms are acknowledging their growing responsibility in media consumption and have introduced specific features to increase users' awareness and levels of information.

### Prior Work

With respect to past epidemics and pandemics, studies on traditional news coverage of the 2009 H1N1 pandemic highlighted the importance of framing and its effect on people's perception, behaviors (eg, vaccination intent), and stigmatization of cultures at the epicenter of the outbreak, as well as how these factors differ across countries and cultures [16-21]. During the Zika virus epidemic in 2016, public attention was synchronized across US states, driven by news coverage about the outbreak and independent of the real local risk of infection [22]. With respect to the COVID-19 pandemic itself, a recent study clearly showed how Google searches for *coronavirus* in the United States spiked significantly immediately after the announcement of the first confirmed case in each state [23]. Several studies based on Twitter data have also highlighted how misinformation and low-quality information about COVID-19, although limited overall, spread before the local outbreak and rapidly expanded once local epidemics started [24-26]. In the current landscape, this spread of misinformation has the potential to encourage irrational, unscientific, and dangerous behaviors. On the other hand, despite some important limitations [27], modern media has become a key data source to observe and monitor health. In fact, posts on Twitter [28-33], Facebook [34], and Reddit [35,36], page views in Wikipedia [37,38], and searches on Google [39,40] have been used to study, nowcast, and predict the spreading of infectious diseases as well as the prevalence of noncommunicable illnesses. Therefore, in the current full-fledged digital society, information is not only key to inform people's behavior but can also be used to develop an unprecedented understanding of these behaviors as well as of the phenomena driving them.



## Goal of This Study

The context in which COVID-19 is unfolding is very heterogeneous and complex. Traditional and social media are integral parts of public perception and opinions, and they have potential to trigger behavior changes and thus influence the spread of the pandemic. This complex landscape must be characterized to understand the public attention and response to media coverage. Here, we addressed this challenge by assembling a heterogeneous data set that includes 227,768 news reports and 13,448 YouTube videos published by traditional media, 278,456,892 views of topical Wikipedia pages, and 107,898 submissions and 3,829,309 comments from 417,541 distinct users on Reddit, as well as epidemic data in four different countries: Italy, the United Kingdom, the United States, and Canada. First, we explored how media coverage and epidemic progression influence public attention and response. To achieve this, we analyzed news volume and COVID-19 incidence with respect to volumes of Wikipedia page views and Reddit comments. Our results show that public attention and response are mostly driven by media coverage rather than by disease spread. Furthermore, we observed the typical saturation and memory effects of public collective attention. Moreover, using an unsupervised topic modeling approach, we explored the different topics framed in traditional media and in Reddit discussions. We show that while the attention of news outlets and internet users toward different topics is in good accordance, interesting deviations emerge in their temporal patterns. Also, we highlight that at the end of our observation period, general interest grew toward topics about the resumption of activities after lockdown, the search for a vaccine against SARS-CoV-2, acquired immunity, and antibody tests. Overall, the research presented here offers insights to interpret public perception and response to the current global health emergency and raises questions about the effects of attention saturation on collective awareness and risk perception and thus on tendencies toward behavioral change.

## Methods

### Data Set

#### News Articles and Videos

We collected news articles using News API, a service that allows free downloads of articles published on the internet in a variety of countries and languages [41]. For each country considered, we downloaded all relevant articles published on the internet by selected sources in the period from February 7 to May 15, 2020. We selected relevant articles by considering those citing one of the following keywords: *coronavirus*, *covid19*, *covid-19*, *ncov-19*, and *sars-cov-2*. Note that for each article, we could access the title, a description, and a preview of the full text. In total, our data set consisted of 227,768 news articles; 71,461 were published by Italian media, 63,799 by UK media, 82,630 by US media, and 9878 by Canadian media.

Additionally, we collected all videos published on YouTube by major news organizations in the four countries under investigation via their official YouTube channels using the official application programming interface (API) [42]. In this

process, we downloaded the titles and descriptions of all the videos and selected as relevant those that mentioned one of the following keywords: *coronavirus*, *virus*, *covid*, *covid19*, *sars*, *sars-cov-2*, and *sarscov2*. The reach of each channel (measured by the number of subscribers) varied drastically, from more than 9 million for CNN (United States) to approximately 12,000 for Ansa (Italy). In total, the YouTube data set consisted of 13,448 videos; 3325 were published by Italian channels, 3525 by UK channels, 6288 by US channels, and 310 by Canadian channels.

It is important to underline that while there is good overlap between the sources of news articles and videos, some do not match. This is due to the fact that not all news organizations have a YouTube channel, while others do not produce traditional articles. In [Multimedia Appendix 1](#), we provide a complete list of the news outlets and YouTube channels we considered.

#### Reddit Posts

Reddit is a social content aggregation website on which users can post, comment, and vote on content. It is structured in subcommunities (ie, subreddits) that are centered around a variety of topics. Reddit has already been proven to be suitable for a variety of research purposes, ranging from the study of user engagement and interactions between highly related communities [43,44] to postelection political analyses [45]. Moreover, it has been used to study the impact of linguistic differences in news titles [46] and to explore recent web-related issues such as hate speech [47] and cyberbullying [48] as well as health-related issues such as mental illness [49]; it also provides insights into the opioid epidemic [50].

We used the Reddit API to collect all submissions and comments published in Reddit under the subreddit r/Coronavirus from February 15 to May 15, 2020. After cleaning the data by removing entries deleted by authors and moderators, we retained only submissions with scores >1 to avoid spam. We removed comments with <10 characters and with >3 duplicates to avoid including automatic messages from moderators. The final data set contained 107,898 submissions and 3,829,309 comments from 417,541 distinct users.

To characterize the topics discussed on Reddit, we then selected entries with links to English-language news outlets. The contents of the URLs were extracted using the available implementation of the method described in [51], resulting in 66,575 valid documents.

Reddit does not provide explicit information about users' locations; therefore, we used self-reporting via regular expression to assign locations to users. Reddit users often declare geographical information about themselves in submissions or comment texts. We used the same approach described in [50], in which the use of regular expressions was found to be reliable, resulting in high correlation with census data in the United States; however, we acknowledge a potential higher bias at the country level due to heterogeneities in Reddit population coverage and user demographics. We selected all English-language texts containing expressions such as "I am from" or "I live in" and extracted candidate expressions from the text that followed the expressions to identify texts that

represented country locations. By removing inconsistent self-reporting, we were able to assign a country to 789,909 distinct users, among which 41,465 had posted at least one comment in the subreddit r/Coronavirus (13,811 from the United States, 6870 from Canada, 3932 from the United Kingdom, and 445 from Italy).


### Wikipedia Page Views

Wikipedia has become a popular digital data source to study health information-seeking behavior [52] and to monitor and forecast the spreading of infectious diseases [53,54]. Here, we used the Wikimedia API [55] to collect the number of visits per day to Wikipedia articles and the total monthly visits to a specific project from each country. We considered language to be indicative of a specific country, suggesting that the relevant projects for our analysis would be written in English and Italian (ie, *en.wikipedia* and *it.wikipedia*, respectively). We chose articles directly related to COVID-19 and those in the “See also” section of each page at the time of the analysis (February 7 to May 15, 2020), including country-specific articles (see [Multimedia Appendix 1](#) for the full list of webpages considered).

Except for Italian, where the language is highly indicative of the location, the number of visits to English pages is almost evenly distributed among English-speaking countries. To normalize the signal related to each country, we weighted the number of daily visits to a single article from a specific project  $p$ ,  $S_p(d)$ , with the total number of monthly visits from a country

$c$ , to the related Wikipedia project , such that the number of daily page views for a given Wikipedia project and country is:

$$(1) \quad \text{where the denominator is the total number of views of the specific Wikipedia project. The total volume of views on day } d \text{ from country } c \text{ is then given by the sum over all the articles } a \text{ and projects } p, \text{ namely:}$$

(2) 

### Media Coverage and Collective Web-Based Response

With our data set, we aimed to provide an overview of media coverage and a proxy of public attention and response. On the one hand, the study of news articles and videos enabled us to estimate the exposure of the public to information about the COVID-19 pandemic in traditional news media. On the other hand, the study of users' discussions and responses on social media (through Reddit) and information-seeking (through Wikipedia page views) allowed us to quantify the reaction of individuals both to the COVID-19 pandemic and to news exposure. As mentioned in the Introduction, previous studies showed the usefulness of social media, internet use, and search trends to analyze health-related information streams and monitor public reaction to infectious diseases [56-60]. Hence, we considered the volume of comments of geolocalized users on the subreddit r/Coronavirus to explore the public discussion in reaction to media coverage of the epidemic in the various

countries; meanwhile, we considered the number of views of relevant Wikipedia pages about the COVID-19 pandemic to quantify users' interest. It is important to stress that Reddit and Wikipedia provide different aspects of internet users' behavior and collective response. In fact, while Reddit posts can be regarded as a general indicator of the web-based discussion surrounding the global health emergency, the number of visits to COVID-19-related Wikipedia pages is a proxy of health information-seeking behavior. Health information-seeking behavior is the act by which individuals retrieve and acquire new knowledge about a specific topic related to health [61,62]; it is likely to be triggered on a population scale by a disrupting event, such as the threat of a previously unknown disease [63,64].

### Linear Regression Approach to Model Collective Attention

To analyze the relationship between media coverage, epidemic progression, and users' collective web-based response, we considered a linear regression model that predicts the public response for each country given the amount of news exposure. To include “memory effects” in the public response to media coverage, we also considered a modified version of this simple model, in which we weight a cumulative news articles volume time series with an exponential decay term [22]. Formally, we define the new variable as:

$$(3) \quad \text{where } \tau \text{ is a free parameter that sets the memory time scale and is tuned by comparing different variants of the linear regression with } \tau \in [1, 45] \text{ in terms of the adjusted coefficient of determination } R^2 [65] \text{ (results for the best } \tau \text{ are displayed). These two models were compared to a linear regression that considers only COVID-19 incidence to predict public collective attention. Then, the models considered are:}$$

$$\text{Model I: } y_t = \alpha_1 \text{incidence}_t + u_t$$

$$\text{Model II: } y_t = \alpha_1 \text{news}_t + u_t$$

$$\text{Model III: } y_t = \alpha_1 \text{news}_t + \alpha_2 \text{newsMEM}_t + u_t \quad (4)$$

where  $y_t$  can be the volume of Reddit comments of geolocalized users or of country-specific Wikipedia visits, and  $u_t$  is the error term. In [Multimedia Appendix 1](#), we provide more details on the model diagnostics and fitting procedure.

### Topic Modeling

Topic modeling has emerged as one of the most effective methods for classifying, clustering, and retrieving textual data, and it has been the object of extensive investigation in the literature. Many topic analysis frameworks are extensions of well-known algorithms that are considered to be state-of-the-art for topic modeling. Latent Dirichlet allocation (LDA) [66] is the reference for probabilistic topic modeling. Nonnegative matrix factorization (NMF) [67] is the counterpart of LDA for matrix factorization. Although there are many approaches to temporal and hierarchical topic modeling [68-70], we chose to apply NMF to the data set and then build time-varying intensities

for each topic using the publication dates of the articles. Starting from a data set  $D$  containing the news articles shared in Reddit, we extracted words and phrases with the methodology described in [71], discarding terms with frequencies  $>10$ , to form a vocabulary  $V$  with approximately 60,000 terms. Each document was then represented as a vector of term counts in a bag-of-words approach. We applied term frequency–inverse document frequency (TF-IDF) normalization [72] and extracted a total of  $K=64$  topics through NMF:

$$\| \mathbf{X} \|_F$$

(5)

where  $\| \cdot \|_F$  is the Frobenius norm and  $\mathbf{X} \in \mathbf{R}^{|D| \times |V|}$  is the matrix resulting from TF-IDF normalization, subject to the constraint that the values in  $\mathbf{W} \in \mathbf{R}^{|D| \times K}$  and  $\mathbf{H} \in \mathbf{R}^{K \times |V|}$  must be nonnegative. The nonnegative factorization was achieved using the projected gradient method with sparseness constraints, as described in [73,74]. The matrix  $\mathbf{H}$  was then used as a transformation basis for other data sets (eg, with a new matrix  $\mathbf{D}^{(t)}$ , we fixed  $\mathbf{H}$  and calculated a new  $\mathbf{X}$  according to Equation 5). For each topic  $k$ , we built a time series  $s_k$  for each data set  $D$ , where  $\mathbf{X}$  is the strength of topic  $k$  at time  $t$ . For the news outlets data set,  $\mathbf{X}$ , where  $D^{(t)}$  is the set of all documents shared at time  $t$  in news outlets. For Reddit, we weighted each shared document by its number of comments, and  $\mathbf{X}$ , where  $D^{(t)}$  is the set of all documents shared at time  $t$  in Reddit and  $c_i$  is the number of comments associated with document  $i$ . Finally, we defined the relevance  $R$  of a topic as the integral in time of the strength. Therefore, given  $t_0$  and  $t_f$  as the start and end of our analysis interval,  $\mathbf{X}$ . In Multimedia Appendix 1, we show that choosing  $K=64$  as the number of extracted topics provides a good balance between sufficient captured topic strength and good topic coherence.

## Results

### Impact of Media Coverage and Epidemic Progression on Collective Attention

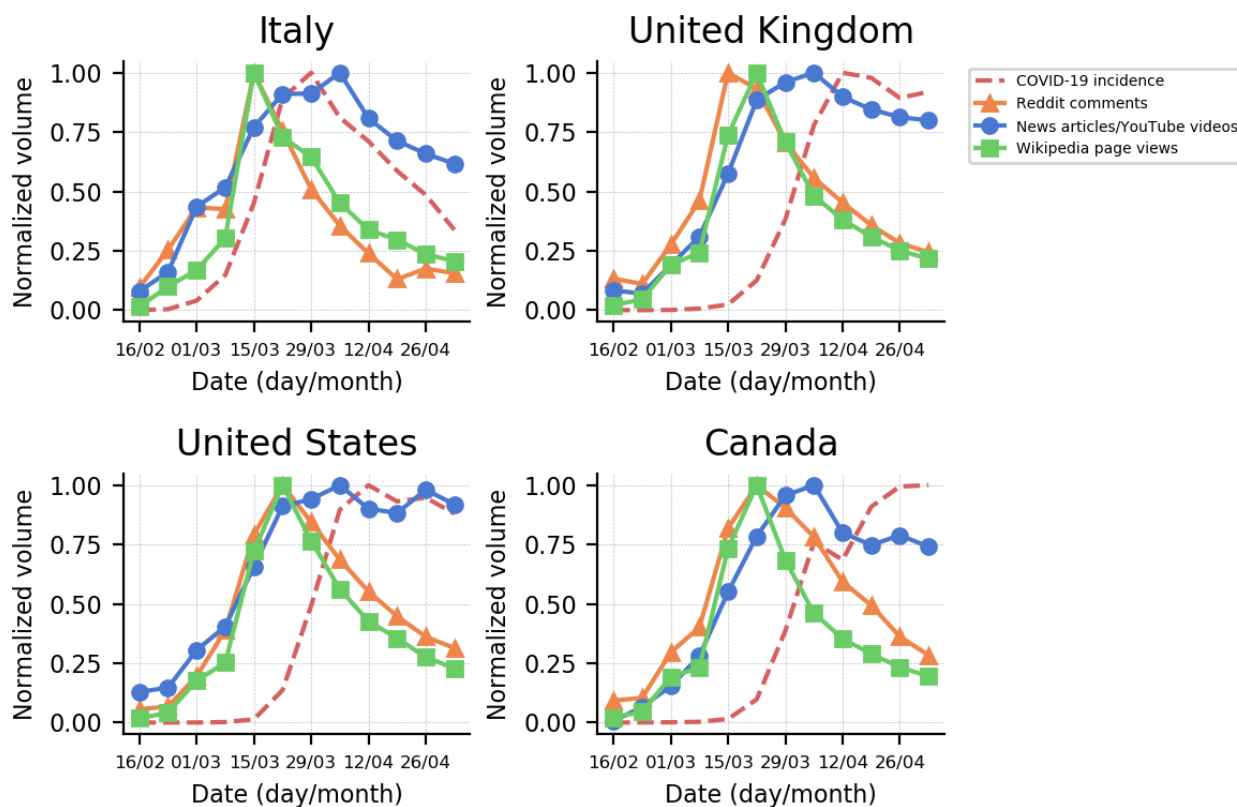
To answer the important question of how collective attention is shaped by news media coverage and epidemic progression, we started by comparing the weekly volumes of news stories and videos published on YouTube, Wikipedia page views, and

Reddit comments of geolocalized users with the weekly COVID-19 incidence in the four countries considered (Figure 1). It can be seen that as COVID-19 spread, both media coverage and public interest grew with time. However, public attention, quantified by the number of Reddit comments and Wikipedia page views, sharply decreased after reaching a peak, even though the volume of news stories and the incidence of COVID-19 remained high. Furthermore, the peak in public attention consistently anticipated the maximum media exposure and maximum COVID-19 incidence.

The correlation between media coverage, public attention, and progression of the epidemic is quantified in more detail in Table 1. The table shows that news coverage of each country is strongly correlated with COVID-19 incidence (both global and domestic) and slightly less correlated with the volumes of Reddit comments and Wikipedia views, which in turn are much less correlated with COVID-19 incidence (both global and domestic). This result was observed for all countries under consideration; it highlights how the spread of COVID-19 triggered media coverage as well as how public response was more likely to be driven by news exposure in each country than by the progression of COVID-19.

Beyond these observations, Figure 2 shows the share of citations of Chinese versus home country locations by Italian, UK, US, and Canadian news outlets before and after the first COVID-19 death occurred in those countries; the geographic locations were extracted from the text using the methods described in [75,76]. Interestingly, Italy is the only country where the news volume shows a higher correlation with domestic incidence than with global incidence (ie, news references to China). This suggests that Italian media coverage follows internal evolution more closely than global evolution, in contrast to other countries. This is probably due to the fact that Italy is the location of the first COVID-19 outbreak outside Asia. This observation is supported by Figure 2, which shows the citation share of Italian locations by Italian news media before and after the first COVID-19 death was confirmed in Italy on February 23, 2020. After this date, Italian locations represent about 74% of all places cited by Italian media (in our data set), with an increase of 45% with respect to the same statistics calculated before. Similar effects, although generally less intense, were observed in the other countries. Therefore, while media coverage is generally well synchronized with global COVID-19 incidence, the media attention gradually shifts toward the internal evolution of the pandemic as soon as domestic outbreaks erupt.

**Figure 1.** Normalized weekly volumes of news articles and YouTube videos, Reddit comments, and Wikipedia page views related to the COVID-19 pandemic and the incidence of COVID-19 in different countries.



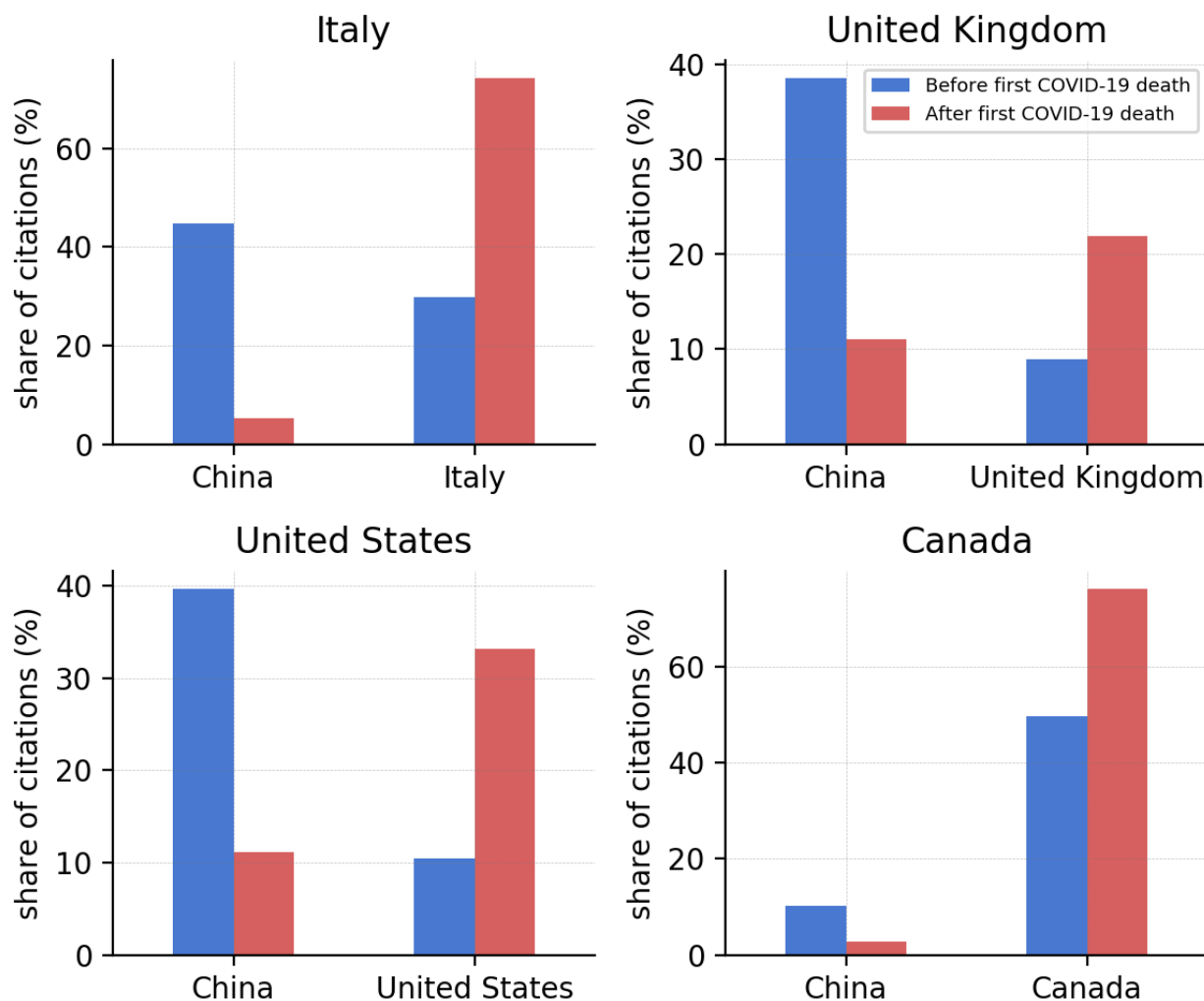
**Table 1.** Country-specific Pearson correlation coefficients for news coverage and global and domestic COVID-19 incidence, volumes of Reddit comments, and volumes of Wikipedia page views; domestic COVID-19 incidence and volumes of Reddit comments and Wikipedia views; and global COVID-19 incidence and volumes of Reddit comments and Wikipedia views.

Country	Global incidence of COVID-19	<i>P</i> value	Country incidence of COVID-19	<i>P</i> value	Reddit comments	<i>P</i> value	Wikipedia page views	<i>P</i> value
<b>Italy</b>								
News	0.59	.04	0.92	<.001	0.43	.17	0.71	.009
Global incidence of COVID-19	1	N/A <sup>a</sup>	— <sup>b</sup>	N/A	−0.42	.18	−0.01	.97
Country incidence of COVID-19	—	N/A	1	N/A	0.30	.34	0.64	.02
<b>United Kingdom</b>								
News	0.83	<.001	0.74	.006	0.50	.10	0.62	.03
Global incidence	1	N/A	—		−0.04	.90	0.09	.77
Country incidence	—	N/A	1	N/A	−0.15	.64	−0.04	.91
<b>United States</b>								
News	0.84	<.001	0.79	.002	0.70	.01	0.64	.03
Global incidence	1	N/A	—	N/A	0.25	.44	0.17	.60
Country incidence	—	N/A	1	N/A	0.16	.62	0.08	.81
<b>Canada</b>								
News	0.82	.001	0.71	.01	0.73	.007	0.59	.04
Global incidence	1	N/A	—	N/A	0.23	.46	0.06	.85
Country incidence	—	N/A	1	N/A	0.05	.87	−0.10	.76

<sup>a</sup>N/A: not applicable.<sup>b</sup>—: not determined.



**Figure 2.** Shares of citations of China versus home country locations by Italian, UK, US, and Canadian news outlets before and after the first COVID-19 death occurred in each country.



To more systematically explore the relationship between media coverage, public attention, and epidemic progression, we considered a linear regression model to nowcast the collective public attention for each country (quantified by the number of comments by geolocalized Reddit users or visits to relevant Wikipedia pages) using the volume of media coverage or the COVID-19 incidence as independent variables. We also included “memory effects” on the public attention by considering an exponential decaying term in the news time series [22]. We

compared the three models, where the independent variables are the domestic incidence, the news volume, and the news volume plus a memory term, using the adjusted coefficient of determination ( $R^2$ ) [65]. We found that the model that considered only COVID-19 incidence performed worse than the models that considered media coverage (Table 2). This enforces the idea that collective attention is mainly driven by media coverage rather than by COVID-19 incidence. In addition, we found that including memory effects improved the model performance.

**Table 2.** Adjusted  $R^2$  values for the three linear regression models applied to predict Reddit comments and Wikipedia page views ( $P < .001$ ).

Country	Model I		Model II		Model III	
	Reddit comments	Wikipedia page views	Reddit comments	Wikipedia page views	Reddit comments	Wikipedia page views
Italy	0.52	0.65	0.68	0.73	<i>0.82<sup>a</sup></i>	<i>0.79</i>
United Kingdom	0.27	0.27	0.72	0.74	<i>0.82</i>	<i>0.85</i>
United States	0.42	0.35	0.82	0.74	<i>0.89</i>	<i>0.82</i>
Canada	0.35	0.23	0.83	0.71	<i>0.90</i>	<i>0.82</i>

<sup>a</sup>Italics indicate the superior performance of Model III.

More formally, we compared Model I to Model III using the Cox test [77] for nonnested models, and we compared Model II to Model III using the  $F$  test [78] for nested models. In all cases we obtained  $P$  values  $<.001$ , providing strong statistical evidence that Model III actually outperforms the other models. Not surprisingly, the coefficients of the “memory effects” term reported in Table 3 are negative for all countries. This implies

that public attention actually saturates in response to news exposure and enables us to quantify the rate at which this phenomenon occurs.

In the next section, we will characterize the media coverage and internet users’ response more specifically in terms of content produced and consumed.

**Table 3.** Coefficient estimates (95% CI) for Model III (news plus memory effects). All coefficients are significant with  $P<.001$ .

Country	News		News plus memory effects	
	Reddit comments	Wikipedia page views	Reddit comments	Wikipedia page views
Italy	0.87 (0.60 to 1.14)	0.43 (0.29 to 0.58)	−0.41 (−0.59 to −0.23)	−0.15 (−0.26 to −0.04)
United Kingdom	0.95 (0.62 to 1.27)	0.99 (0.68 to 1.30)	−0.44 (−0.71 to −0.18)	−0.47 (−0.70 to −0.23)
United States	1.03 (0.79 to 1.27)	0.83 (0.58 to 1.09)	−0.51 (−0.77 to −0.24)	−0.46 (−0.73 to −0.19)
Canada	1.12 (0.89 to 1.36)	1.06 (0.67 to 1.44)	−0.40 (−0.59 to −0.22)	−0.45 (−0.72 to −0.18)

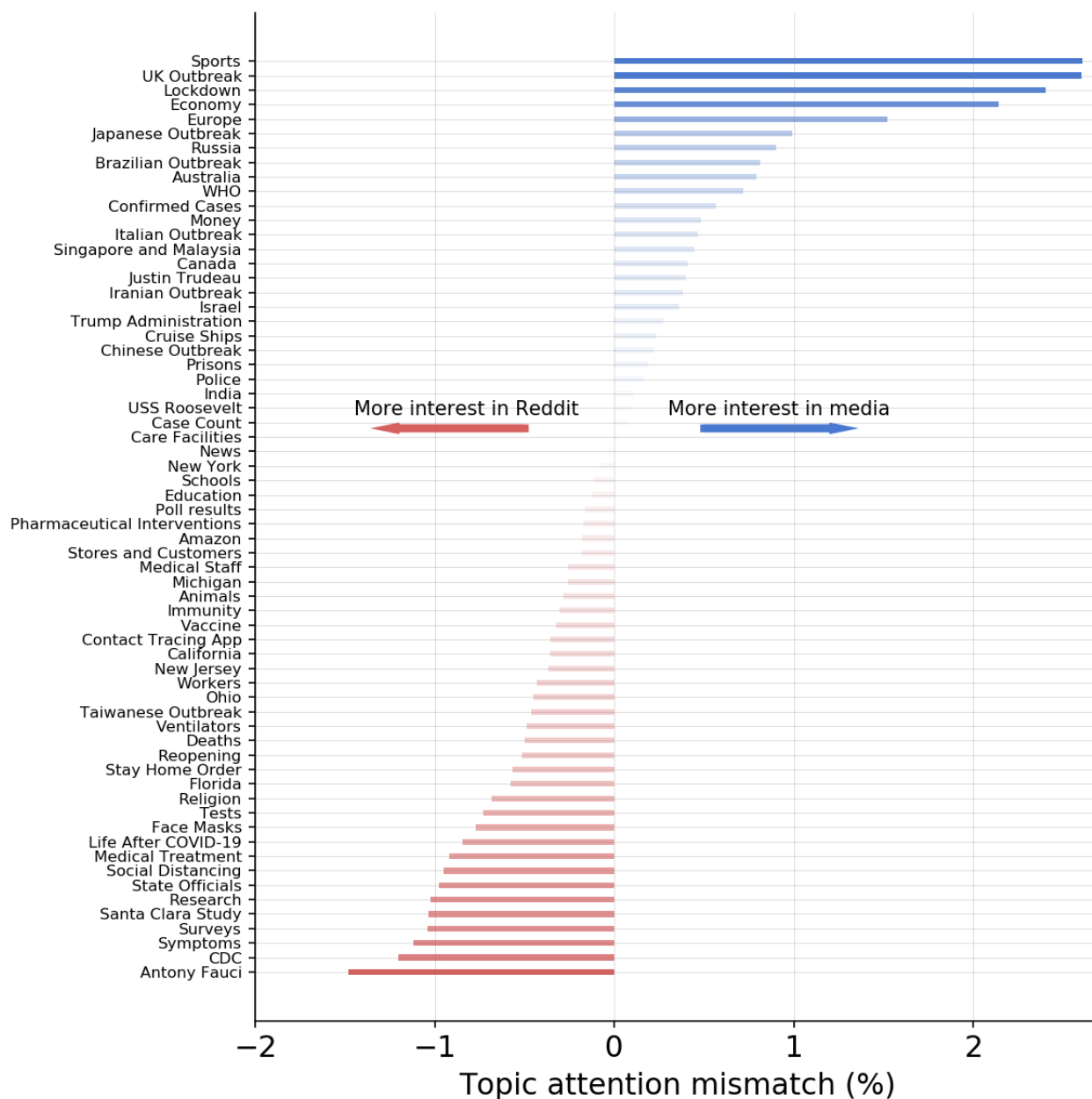
### Dynamics of Content Production and Consumption

While collective attention and media coverage are well correlated in terms of volume, the content and topics discussed by media and consumed by internet users may not be as synchronized [79,80]. To shed light on this issue, we adopted an unsupervised topic modeling approach to extract prevalent topics in the news articles mentioned and discussed on Reddit. Indeed, users on Reddit often post submissions containing news articles, and discussion unfolds in the comments under the submissions. Differently from the previous section and to provide a comprehensive overview of the topics discussed, in this section, we do not take any geographical context into account. However, in Multimedia Appendix 1, we provide some additional insights into the specific topics discussed by users in different countries.

We characterized the main topics discussed on Reddit by considering all submissions that included a news article in English. We then applied a topic modelling approach to the content of this news article set. Specifically, we extracted topics using NMF [67], a popular method for this type of task. In this way, we extracted the 64 most relevant topics in the news articles shared on Reddit. As a second step, we applied the model trained on the Reddit news to the set of articles published

by mainstream media. That is, we characterized the news published by media outlets in terms of the topics discussed on Reddit. This choice enabled us to directly compare the topics covered by the media to the public discussion around this news exposure. A complete list of the 64 topics extracted with the most frequent words is provided in Multimedia Appendix 1. We considered the number of articles published on a certain topic as a proxy of general interest of traditional media in that topic; meanwhile, we measured the collective interest of Reddit users by the number of comments under the news articles on a specific topic. Figure 3 shows an overview of the topics extracted and a comparison of the interest of media and Reddit users. We obtained a diverse and heterogeneous set of topics, including the global spread of the virus (Outbreaks, WHO [World Health Organization], CDC [US Centers for Disease Control and Prevention]); COVID-19 symptoms, treatment, hospitals and care facilities (Symptoms, Medical Treatment, Medical Staff, Care Facilities); the economic impact of the pandemic and responses from the governments to the upcoming crisis (Economy, Money); different societal aspects (Sports, Religious Services, Education); and possible interventions to mitigate the spread of the virus (Face Masks, Social Distancing, Tests, Vaccine).

**Figure 3.** Differences in interest percentage shares of different topics by traditional media outlets and Reddit users. For example, +2% on the x-axis indicates that traditional media dedicates proportionally 2% more attention to that specific topic than Reddit users. CDC: US Centers for Disease Control and Prevention; UK: United Kingdom; WHO: World Health Organization.



Overall, the levels of attention of traditional media outlets and Reddit users toward the different topics are in good accordance. Indeed, in Figure 3, we represent the difference between interest shares toward different topics in media and Reddit submissions. That is, we computed the percentage share of attention dedicated by news outlets and Reddit users to each topic, and we subtracted these two quantities. We observed a maximum absolute mismatch in interest share of 2.61%. However, we observed that Reddit users are slightly more interested in topics regarding health (Symptoms, Medical Treatment), nonpharmaceutical interventions and personal protective equipment (Social Distancing, Face Masks), studies and information on the epidemic (Research, Surveys, Santa Clara Study, CDC), and specific public figures such as Anthony Fauci. Interestingly, the Santa Clara Study topic refers to the discussion

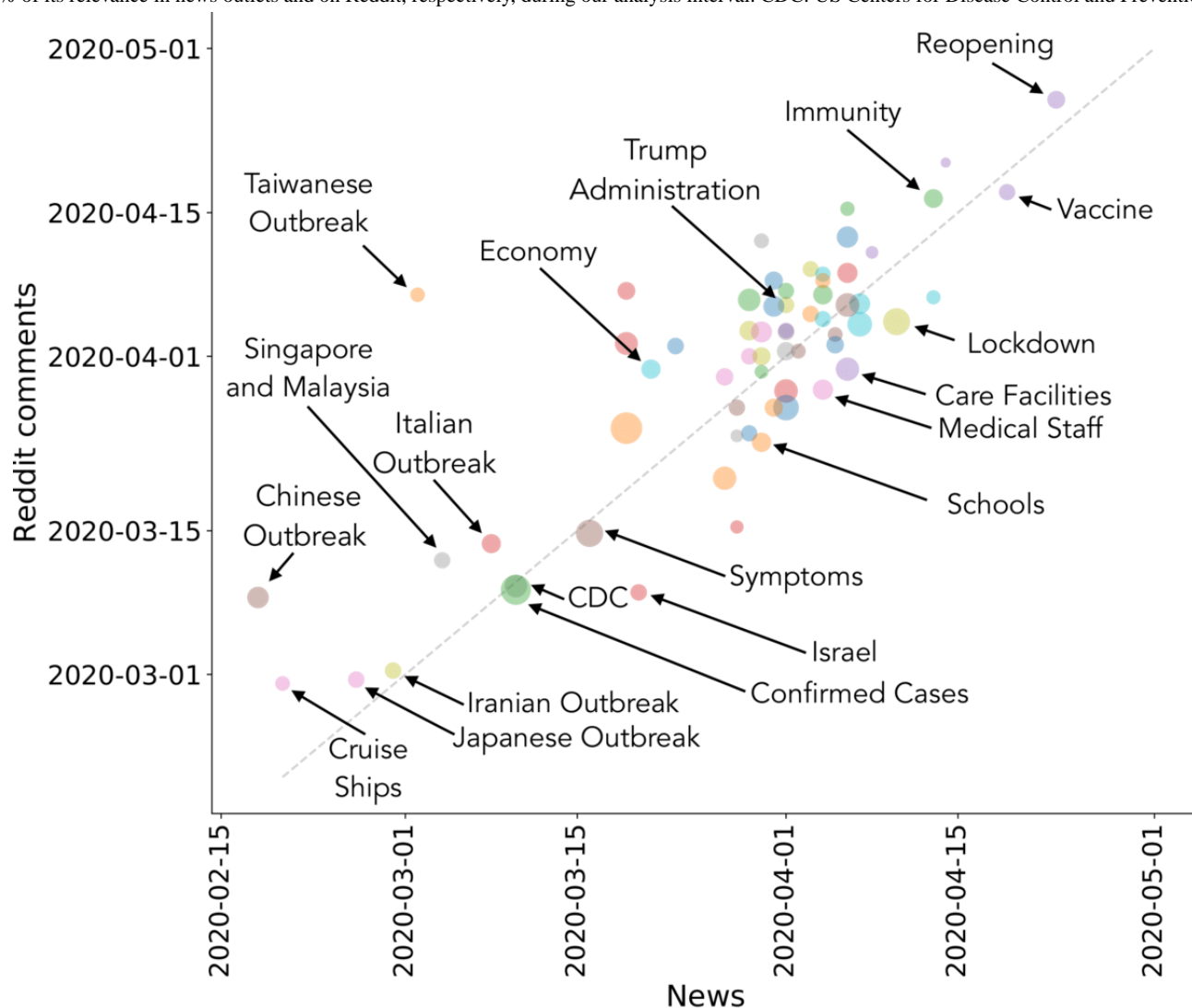
about a controversial scientific paper suggesting that a much higher fraction of the population in Santa Clara County was infected with respect to what was originally thought [81]. Because the study suggested a lower mortality rate, the preprint was quickly leveraged to support protest against lockdowns [82]; meanwhile, substantial flaws have been detected in the scientific methodology of the paper [83].

The overview of topics presented here does not take temporal dynamics of interest into account. However, topics showing similar overall statistics may present a mismatch in temporal patterns. Hence, in the following, we take into account the temporal evolution of interest toward different topics. In Figure 4, we represent each topic as a single point: the x-coordinate and y-coordinate indicate the  $t_{1/2}$  when the topic reached 50% of its total relevance  $R$  in news outlets and on Reddit,

respectively, during the analysis interval. Therefore, topics in the bottom left region became relevant very early in the public discussion. Among these, we recognize themes centered on early COVID-19 outbreaks (ie, Chinese, Japanese, Iranian, and Italian outbreaks), events related to cruise ships, specific countries (ie, Israel, Singapore, and Malaysia), and topics regarding (early) health issues (ie, Symptoms, Confirmed Cases, and CDC). In contrast, topics in the top right region became relevant toward the end of the analysis interval (early May).

Reasonably, this region contains topics about the resumption of activities after lockdown (ie, Reopening), the feasibility and timing of a possible vaccine against SARS-CoV-2 (ie, Vaccine), and discussions regarding acquired immunity and antibody tests (ie, Immunity). All other topics are clustered around the end of March and mid-April 2020, which is the period in which the general discussion surrounding the COVID-19 pandemic increased sharply, as also shown in Figure 1.

**Figure 4.** Scatter plot with the 64 topics extracted via nonnegative matrix factorization. The x-axis and y-axis coordinates indicate when a topic achieved 50% of its relevance in news outlets and on Reddit, respectively, during our analysis interval. CDC: US Centers for Disease Control and Prevention.



Note that the diagonal in Figure 4 (plotted as a dashed line) separates topics according to their temporal evolution. Above and below the diagonal, we find topics in which interest on Reddit grows slowly and quickly, respectively, with respect to the media coverage. Therefore, above the diagonal, the interest of Reddit users is mainly triggered by media exposure, while below it, the interest grows faster and declines rapidly despite sustained media exposure. The top left and bottom right regions are empty, indicating that as a first approximation, temporal patterns of attention by traditional media and Reddit users are well synchronized; however, interesting deviations from the diagonal are observable. For example, above the diagonal, one can mainly find topics related to various outbreaks, economics,

and politics, for which the interest on Reddit follows the media coverage. Below the diagonal, we observe topics more related to everyday life, such as Schools, Medical Staff, Care Facilities, and Lockdown, for which the attention on Reddit accelerates with respect to media coverage and then declines rapidly. Note that our view of the topics discussed on Reddit is limited, as we only considered topics from news articles shared in submissions and did not explicitly take content expressed in comments into account. This ensures a proper comparison with the topics extracted from published news reports and explains the absence of points in the bottom right corner of Figure 4.

## Discussion

### Principal Results

In this work, we characterized the response of internet users to both media coverage and COVID-19 pandemic progression. As a first step, we focused on the impact of media coverage on collective attention in different countries, characterized as volumes of country-specific Wikipedia page views and comments of geolocalized Reddit users. We showed that collective attention was mainly driven by media coverage rather than epidemic progression, rapidly became saturated, and decreased despite media coverage and COVID-19 incidence remaining high. These results are in very good accordance with findings obtained in previous contexts related to epidemics and pandemics. Indeed, a similar media-driven spiky unfolding of public attention, measured through the information-seeking and public discussions of internet users, was observed during the 2009 H1N1 influenza pandemic [84,85], the 2016 Zika virus outbreak [86], influenza season [87], and more localized public health emergencies such as the 2013 measles outbreak in the Netherlands [88]. Our findings confirm the central role of the media, showing how media exposure is capable of shaping and driving collective attention during a national and global health emergency. Media exposure is another important factor that can influence individual risk perception as well [79,89-91]. The timing and framing of the information disseminated by media can actually modulate the attention and, ultimately, the behavior of individuals [2]. This becomes an even greater concern in a context where the most effective strategy to fight the spread of disease involves containment measures based on individuals' behavior.

Also, we showed how media coverage sharply shifted to the domestic situation as soon as the first death was confirmed in the home country. Arguably, this may have played an important role in individual risk perception. We can speculate that reframing the emergency within a national dimension can amplify the perceived susceptibility of individuals [92,93] and thus increase the adoption of behavioral changes [4,94]. Indeed, previous studies showed that at the beginning of February 2020, people were overly optimistic regarding the risks associated with the new virus circulating in Asia, and their perception sharply changed after the first cases were confirmed in their countries [9,95].

As a second step, we focused on the dynamics of content production and consumption. We modeled topics published in mainstream media and discussed on Reddit, showing that Reddit users were generally more interested in health, data regarding the new disease, and interventions needed to halt the spreading with respect to media exposure. By taking into account the dynamics of the extracted topics, we showed that while their temporal patterns are generally synchronized, the public attention to topics related to politics and economics is mainly triggered by media exposure, while the interest in topics more related to daily life increases on Reddit with respect to media coverage.

### Limitations

Of course, our research comes with limitations. First, we characterized the exposure of individuals to the COVID-19 pandemic by considering only news articles and YouTube videos published on the internet by major news outlets. However, individuals are also exposed to relevant information through other channels, with television being the most important [96]. Second, a 2013 Pew Internet Study found that Reddit users are more likely to be young men [97]; it was shown that around 15% of male internet users aged 18 to 29 years report using Reddit, compared to 5% of women in the same age range and 8% of men aged 30 to 49 years. Similarly, informal surveys proposed to users [98] showed that most respondents were males in their "late teens to mid-20s" and that female users were "very much in the minority." Furthermore, Reddit is much more popular among urban and suburban residents than among individuals living in rural areas [97]. In addition to sociodemographic biases, other studies have suggested that Reddit has become an increasingly self-referential community, reinforcing the tendency to focus on its own contents rather than external sources [99]. Thus, the perceptions, interests, and behaviors of Reddit users may differ from those of the general population. A similar argument can be raised for Wikipedia searches. Indeed, the use of the internet, especially for information-seeking purposes, can vary across people with different sociodemographic backgrounds [100-102]. Additionally, we extracted Reddit users' geographic location using a method based on regular expressions that has been successfully used in previous work [50]. However, because we have no ground truth data for comparison, we must consider the quality of location detection to be a possible limitation. Finally, our view on internet users' reactions is partial. Indeed, we did not consider other popular digital data sources, such as Twitter. The reasons for this choice are twofold. First, many studies have already characterized public response during current and past health emergencies through the lens of Twitter [25,58,60,85,86,103,104]. Second, several studies have reported a high prevalence of bots as drivers of low-quality information and discussions on COVID-19 on this platform [24,25,105-107]. Thus, careful and challenging additional steps would be necessary to isolate, identify, and distinguish organic Twitter discussions and reactions that originated from traditional media from those sparked by social bots. We leave this for future work.

### Conclusions

Our work offers further insights to interpret public response to the current global health emergency and raises questions about possible undesired effects of communication. On one hand, our results confirm the pivotal role of media during health emergencies, showing how collective attention is mainly driven by media coverage. Therefore, because people are highly reactive to the news they are exposed to at the beginning of an outbreak, the quality and type of information provided may have critical effects on risk perception and behaviors, which will ultimately affect the unfolding of the outbreak. However, we also found that collective internet attention saturates and declines rapidly, even when media exposure and disease circulation remain high. Attention saturation has the potential to affect collective awareness and perceived risk, which



ultimately affects the propensity toward virtuous individual behavioral changes aimed at mitigating the spread of disease. Furthermore, especially in the case of unknown viruses, attention saturation may exacerbate the spreading of low-quality information, which is likely to spread in the early phases of the outbreak when the characteristics of the disease are uncertain. Future work is needed to characterize the actual effects of attention saturation on human perceptions during a global health emergency. Our findings suggest that public health authorities should consider reinforcing specific communication channels, such as social media platforms, to compensate for the natural phenomenon of attention saturation. Indeed, these channels have the potential to create more durable engagement with people through a continuous loop of direct interactions. Currently, public health authorities are regularly issuing declarations on social media. However, the CDC did not even

have a Twitter account in 2009 during the H1N1 pandemic (the account was created in May 2010). While this is just one example, it underlines that the communication of these global health emergencies through social media platforms is relatively new. Therefore, there is great need to further reinforce these channels and engage people through them. Simultaneously, public health authorities should consider strengthening additional communication channels. One example of this is the participatory surveillance platforms that are appearing worldwide, such as Influenzanet, Flu Near You, and FluTracking [108-110], which can deliver in-depth targeted information to individuals during public health emergencies and promote the exchange of information between people and public health authorities; this has potential to enhance the level of engagement in the community [111].

## Acknowledgments

The authors would like to thank the startup company Quick Algorithm for providing the platform where the data collected during the COVID-19 pandemic were visualized in real time [112]. DP and MT acknowledge support from the Lagrange Project of the Institute for Scientific Interchange Foundation (ISI Foundation) funded by Fondazione Cassa di Risparmio di Torino (Fondazione CRT). MT acknowledges support from EPIPOSE (Epidemic intelligence to minimize COVID-19's public health, societal and economic impact) H2020-SC1-PHE-CORONAVIRUS-2020 call. MS and AP acknowledge support from the Research Project "Casa Nel Parco" (POR FESR 14/20 - CANP - Cod. 320 - 16 - Piattaforma Tecnologica "Salute e Benessere") funded by Regione Piemonte in the context of the Regional Platform on Health and Wellbeing. AP acknowledges partial support from Intesa Sanpaolo Innovation Center. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. NG acknowledges support from the Doctoral Training Alliance.

## Authors' Contributions

NG, MS, DP, AP, and NP conceptualized the study. NG, NP, AP, and MT collected the data. NG, AP, and FC performed analyses. NG, MS, and NP wrote the initial draft of the manuscript. NG and AP provided visualization. All authors (NG, NP, DP, MS, AP, MT, and FC) discussed the research design, reviewed, edited, and approved the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary information.

[DOCX File, 12651 KB - [jmir\\_v22i10e21597\\_app1.docx](#)]

## References

1. Barry JM. Pandemics: avoiding the mistakes of 1918. *Nature* 2009 May 21;459(7245):324-325 [FREE Full text] [doi: [10.1038/459324a](#)] [Medline: [19458695](#)]
2. Funk S, Salathé M, Jansen VAA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J R Soc Interface* 2010 Sep 06;7(50):1247-1256 [FREE Full text] [doi: [10.1098/rsif.2010.0142](#)] [Medline: [20504800](#)]
3. Verelst F, Willem L, Beutels P. Behavioural change models for infectious disease transmission: a systematic review (2010-2015). *J R Soc Interface* 2016 Dec;13(125):20160820 [FREE Full text] [doi: [10.1098/rsif.2016.0820](#)] [Medline: [28003528](#)]
4. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. *Health Educ Q* 1988 Sep 04;15(2):175-183. [doi: [10.1177/109019818801500203](#)] [Medline: [3378902](#)]
5. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 2020 Apr 19;26(4):506-510 [FREE Full text] [doi: [10.1038/s41591-020-0822-7](#)] [Medline: [32284616](#)]
6. WHO Timeline - COVID-19 (Archived). World Health Organization. URL: <https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19> [accessed 2020-05-11]

7. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. World Health Organization. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/> [accessed 2020-05-11]
8. #Coronavirus: Today I met with England's Chief Medical Officer Professor Chris Whitty @CMO\_England for the latest update. Sadiq Khan Facebook page. 2020 Mar 11. URL: <https://www.facebook.com/sadiqforlondon/posts/3025766374142796> [accessed 2020-05-11]
9. Raude J, Debin M, Souty C, Guerrisi C, Turbelin C, Falchi A, et al. Are people excessively pessimistic about the risk of coronavirus infection? PsyArXiv. Preprint posted online March 8, 2020 . [doi: [10.31234/osf.io/364qj](https://doi.org/10.31234/osf.io/364qj)]
10. Kraemer MUG, Yang C, Gutierrez B, Wu C, Klein B, Pigott DM, Open COVID-19 Data Working Group, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 2020 May 01;368(6490):493-497 [FREE Full text] [doi: [10.1126/science.abb4218](https://doi.org/10.1126/science.abb4218)] [Medline: [32213647](https://pubmed.ncbi.nlm.nih.gov/32213647/)]
11. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* 2020 May 15;368(6492):742-746 [FREE Full text] [doi: [10.1126/science.abb4557](https://doi.org/10.1126/science.abb4557)] [Medline: [32269067](https://pubmed.ncbi.nlm.nih.gov/32269067/)]
12. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* 2020 Mar;395(10228):931-934. [doi: [10.1016/s0140-6736\(20\)30567-5](https://doi.org/10.1016/s0140-6736(20)30567-5)]
13. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, et al. COVID-19: towards controlling of a pandemic. *Lancet* 2020 Mar;395(10229):1015-1018. [doi: [10.1016/s0140-6736\(20\)30673-5](https://doi.org/10.1016/s0140-6736(20)30673-5)]
14. Colbourn T. COVID-19: extending or relaxing distancing control measures. *Lancet Public Health* 2020 May;5(5):e236-e237. [doi: [10.1016/s2468-2667\(20\)30072-4](https://doi.org/10.1016/s2468-2667(20)30072-4)]
15. Fox S, Duggan M. Health Online 2013. Pew Research Center. 2013 Jan 15. URL: <https://www.pewresearch.org/internet/2013/01/15/health-online-2013/> [accessed 2020-09-21]
16. Lee ST. Predictors of H1N1 Influenza Pandemic News Coverage: Explicating the Relationships between Framing and News Release Selection. *Int J Strateg Commun* 2014 Sep 09;8(4):294-310. [doi: [10.1080/1553118x.2014.913596](https://doi.org/10.1080/1553118x.2014.913596)]
17. McCauley M, Minsky S, Viswanath K. The H1N1 pandemic: media frames, stigmatization and coping. *BMC Public Health* 2013 Dec 03;13:1116 [FREE Full text] [doi: [10.1186/1471-2458-13-1116](https://doi.org/10.1186/1471-2458-13-1116)] [Medline: [24299568](https://pubmed.ncbi.nlm.nih.gov/24299568/)]
18. Lin CA, Lagoe C. Effects of News Media and Interpersonal Interactions on H1N1 Risk Perception and Vaccination Intent. *Commun Res Rep* 2013 Apr;30(2):127-136. [doi: [10.1080/08824096.2012.762907](https://doi.org/10.1080/08824096.2012.762907)]
19. Lee ST, Basnyat I. From press release to news: mapping the framing of the 2009 H1N1 A influenza pandemic. *Health Commun* 2013;28(2):119-132. [doi: [10.1080/10410236.2012.658550](https://doi.org/10.1080/10410236.2012.658550)] [Medline: [22439616](https://pubmed.ncbi.nlm.nih.gov/22439616/)]
20. Jung Oh H, Hove T, Paek H, Lee B, Lee H, Kyu Song S. Attention cycles and the H1N1 pandemic: a cross-national study of US and Korean newspaper coverage. *Asian J Commun* 2012 Apr;22(2):214-232. [doi: [10.1080/01292986.2011.642395](https://doi.org/10.1080/01292986.2011.642395)]
21. Keramarou M, Cottrell S, Evans MR, Moore C, Stiff RE, Elliott C, et al. Two waves of pandemic influenza A(H1N1) 2009 in Wales--the possible impact of media coverage on consultation rates, April-December 2009. *Euro Surveill* 2011 Jan 20;16(3):19772 [FREE Full text] [Medline: [21262184](https://pubmed.ncbi.nlm.nih.gov/21262184/)]
22. Tizzoni M, Panisson A, Paolotti D, Cattuto C. The impact of news exposure on collective attention in the United States during the 2016 Zika epidemic. *PLoS Comput Biol* 2020 Mar;16(3):e1007633. [doi: [10.1371/journal.pcbi.1007633](https://doi.org/10.1371/journal.pcbi.1007633)] [Medline: [32163409](https://pubmed.ncbi.nlm.nih.gov/32163409/)]
23. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. *Proc Natl Acad Sci USA* 2020 May 26;117(21):11220-11222 [FREE Full text] [doi: [10.1073/pnas.2005335117](https://doi.org/10.1073/pnas.2005335117)] [Medline: [32366658](https://pubmed.ncbi.nlm.nih.gov/32366658/)]
24. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of. medRxiv. Preprint posted online on April 16, 2020 . [doi: [10.1101/2020.04.08.20057968](https://doi.org/10.1101/2020.04.08.20057968)]
25. Singh L, Bansal S, Bode L, Budak C, Chi G, Kawintiranon K, et al. A first look at COVID-19 information and misinformation sharing on Twitter. ArXiv. Preprint posted online on March 31, 2020 . [Medline: [32550244](https://pubmed.ncbi.nlm.nih.gov/32550244/)]
26. Cinelli M, Quattrocioni W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 Social Media Infodemic. arXiv. Preprint posted online on March 10, 2020 [FREE Full text]
27. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. *Science* 2014 Mar 14;343(6176):1203-1205. [doi: [10.1126/science.1248506](https://doi.org/10.1126/science.1248506)] [Medline: [24626916](https://pubmed.ncbi.nlm.nih.gov/24626916/)]
28. Culotta A. Towards detecting influenza epidemics by analyzing Twitter messages. In: SOMA '10: Proceedings of the First Workshop on Social Media Analytics. 2020 Jul Presented at: SOMA 2010: Workshop on Social Media Analytics; May 7-21, 2010; Washington, DC p. 115-122. [doi: [10.1145/1964858.1964874](https://doi.org/10.1145/1964858.1964874)]
29. Lamos V, Cristianini N. Tracking the flu pandemic by monitoring the social web. 2010 Presented at: 2nd International Workshop on Cognitive Information Processing; June 14-16, 2010; Elba, Italy p. 411-416. [doi: [10.1109/cip.2010.5604088](https://doi.org/10.1109/cip.2010.5604088)]
30. Zhang Q. Forecasting Seasonal Influenza Fusing Digital Indicators and a Mechanistic Disease Model. In: WWW '17: Proceedings of the 26th International Conference on World Wide Web. 2017 Apr Presented at: 26th International Conference on World Wide Web; April 3-7, 2017; Perth, Australia. [doi: [10.1145/3038912.3052678](https://doi.org/10.1145/3038912.3052678)]
31. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media. 2013 Presented at: Seventh International AAAI Conference on Weblogs and Social Media; July 8-11, 2013; Cambridge, MA.

32. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: WebSci '13: Proceedings of the 5th Annual ACM Web Science Conference. 2013 May Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris, France p. 47-56. [doi: [10.1145/2464464.2464480](https://doi.org/10.1145/2464464.2464480)]
33. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013 Dec 9;8(12):e83672 [FREE Full text] [doi: [10.1371/journal.pone.0083672](https://doi.org/10.1371/journal.pone.0083672)] [Medline: [24349542](https://pubmed.ncbi.nlm.nih.gov/24349542/)]
34. Araujo M, Mejova Y, Weber I, Benevenuto F. Using Facebook Ads Audiences for Global Lifestyle Disease Surveillance: Promises and Limitations. In: WebSci '17: Proceedings of the 2017 ACM on Web Science Conference. 2017 Jun Presented at: 2017 ACM Web Science Conference; June 25-28, 2017; Troy, NY p. 253-257. [doi: [10.1145/3091478.3091513](https://doi.org/10.1145/3091478.3091513)]
35. Park A, Conway M. Tracking Health Related Discussions on Reddit for Public Health Applications. AMIA Annu Symp Proc 2017;1362-1371 [FREE Full text] [Medline: [29854205](https://pubmed.ncbi.nlm.nih.gov/29854205/)]
36. Kumar M, Dredze M, Coppersmith G, De Choudhury M. Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides. HT ACM Conf Hypertext Soc Media 2015 Sep;2015:85-94 [FREE Full text] [doi: [10.1145/2700171.2791026](https://doi.org/10.1145/2700171.2791026)] [Medline: [28713876](https://pubmed.ncbi.nlm.nih.gov/28713876/)]
37. Generous N, Fairchild G, Deshpande A, Del Valle SY, Priedhorsky R. Global disease monitoring and forecasting with Wikipedia. PLoS Comput Biol 2014 Nov 13;10(11):e1003892 [FREE Full text] [doi: [10.1371/journal.pcbi.1003892](https://doi.org/10.1371/journal.pcbi.1003892)] [Medline: [25392913](https://pubmed.ncbi.nlm.nih.gov/25392913/)]
38. Hickmann K, Fairchild G, Priedhorsky R, Generous N, Hyman JM, Deshpande A, et al. Forecasting the 2013-2014 influenza season using Wikipedia. PLoS Comput Biol 2015 May;11(5):e1004239 [FREE Full text] [doi: [10.1371/journal.pcbi.1004239](https://doi.org/10.1371/journal.pcbi.1004239)] [Medline: [25974758](https://pubmed.ncbi.nlm.nih.gov/25974758/)]
39. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634)] [Medline: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)]
40. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. PLoS One 2013;8(2):e56176 [FREE Full text] [doi: [10.1371/journal.pone.0056176](https://doi.org/10.1371/journal.pone.0056176)] [Medline: [23457520](https://pubmed.ncbi.nlm.nih.gov/23457520/)]
41. News API. URL: <https://newsapi.org> [accessed 2020-05-11]
42. YouTube Data API. URL: <https://developers.google.com/youtube/v3> [accessed 2020-05-11]
43. Tan C, Lee L. All Who Wander: On the Prevalence and Characteristics of Multi-community Engagement. In: WWW '15: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: 24th International Conference on World Wide Web; May 18-22, 2015; Florence, Italy. [doi: [10.1145/2736277.2741661](https://doi.org/10.1145/2736277.2741661)]
44. Hessel J, Tan C, Lee L. Science, askscience, and badscience: On the coexistence of highly related communities. In: Proceedings of the 10th International AAAI Conference on Web and Social Media. 2016 Presented at: 10th International AAAI Conference on Web and Social Media; May 17-20, 2016; Cologne, Germany p. 171-180.
45. Barthel M. How the 2016 presidential campaign is being discussed on Reddit. Pew Research Center. 2020 May 26. URL: <https://www.pewresearch.org/fact-tank/2016/05/26/how-the-2016-presidential-campaign-is-being-discussed-on-reddit/> [accessed 2020-09-24]
46. Horne B, Adali S. The impact of crowds on news engagement: A Reddit case study. ArXiv. Preprint posted online on March 03, 2017 [FREE Full text]
47. Saleem H, Dillon K, Benesch S, Ruths D. A web of hate: Tackling hateful speech in online social spaces. ArXiv. Preprint posted online on September 28, 2017 [FREE Full text]
48. Bin Abdur Rakib T, Soon LK. Using the Reddit Corpus for Cyberbully Detection. In: Nguyen N, Hoang D, Hong TP, Pham H, Trawiński B, editors. Intelligent Information and Database Systems. ACIIDS 2018. Lecture Notes in Computer Science, vol 10751. Cham, Switzerland: Springer; 2018.
49. Choudhury MD, De S. Mental Health Discourse on reddit: Self-disclosure, Social Support, and Anonymity. In: Proceedings of the Eighth International Conference on Weblogs and Social Media. 2014 Presented at: Eighth International Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI.
50. Balsamo D, Bajardi P, Panisson A. Firsthand Opiates Abuse on Social Media: Monitoring Geospatial Patterns of Interest Through a Digital Cohort. In: WWW '19: The World Wide Web Conference. 2019 May Presented at: WWW '19: The World Wide Web Conference; May 13-17, 2019; San Francisco, CA p. 2572-2579. [doi: [10.1145/3308558.3313634](https://doi.org/10.1145/3308558.3313634)]
51. Radford A, Wu J, Child R, Luan D, Amodei S, Sutskever I. Language Models are Unsupervised Multitask Learners. SemanticsScholar. URL: <https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe> [accessed 2020-09-24]
52. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009;16(4):471-479 [FREE Full text] [doi: [10.1197/jamia.M3059](https://doi.org/10.1197/jamia.M3059)] [Medline: [19390105](https://pubmed.ncbi.nlm.nih.gov/19390105/)]
53. McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. PLoS Comput Biol 2014 Apr;10(4):e1003581 [FREE Full text] [doi: [10.1371/journal.pcbi.1003581](https://doi.org/10.1371/journal.pcbi.1003581)] [Medline: [24743682](https://pubmed.ncbi.nlm.nih.gov/24743682/)]

54. Generous N, Fairchild G, Deshpande A, Del Valle SY, Priedhorsky R. Global disease monitoring and forecasting with Wikipedia. *PLoS Comput Biol* 2014 Nov;10(11):e1003892 [FREE Full text] [doi: [10.1371/journal.pcbi.1003892](https://doi.org/10.1371/journal.pcbi.1003892)] [Medline: [25392913](https://pubmed.ncbi.nlm.nih.gov/25392913/)]
55. Wikimedia API. URL: <https://wikimedia.org/api/rest> [accessed 2020-05-11]
56. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *Am J Prev Med* 2011 May;40(5 Suppl 2):S154-S158. [doi: [10.1016/j.amepre.2011.02.006](https://doi.org/10.1016/j.amepre.2011.02.006)] [Medline: [21521589](https://pubmed.ncbi.nlm.nih.gov/21521589/)]
57. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. *Lancet Infect Dis* 2014 Feb;14(2):160-168 [FREE Full text] [doi: [10.1016/s1473-3099\(13\)70244-5](https://doi.org/10.1016/s1473-3099(13)70244-5)]
58. Park HW, Park S, Chong M. Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea. *J Med Internet Res* 2020 May 05;22(5):e18897 [FREE Full text] [doi: [10.2196/18897](https://doi.org/10.2196/18897)] [Medline: [32325426](https://pubmed.ncbi.nlm.nih.gov/32325426/)]
59. Park A, Conway M. Tracking Health Related Discussions on Reddit for Public Health Applications. *AMIA Annu Symp Proc* 2017;2017:1362-1371 [FREE Full text] [Medline: [29854205](https://pubmed.ncbi.nlm.nih.gov/29854205/)]
60. Lamb A, Paul M, Dredze M. Investigating Twitter as a Source for Studying Behavioral Responses to Epidemics. *AAAI Fall Symposium - Technical Report*. 2012 Jan 01. URL: [https://www.researchgate.net/publication/266506521\\_Investigating\\_Twitter\\_as\\_a\\_Source\\_for\\_Studying\\_Behavioral\\_Responses\\_to\\_Epidemics](https://www.researchgate.net/publication/266506521_Investigating_Twitter_as_a_Source_for_Studying_Behavioral_Responses_to_Epidemics) [accessed 2020-09-24]
61. Lewis N. Information Seeking and Scanning. In: *The International Encyclopedia of Media Effects*. Hoboken, NJ: Wiley; Mar 08, 2017.
62. Lambert SD, Loiselle CG. Health information seeking behavior. *Qual Health Res* 2007 Oct;17(8):1006-1019. [doi: [10.1177/1049732307305199](https://doi.org/10.1177/1049732307305199)] [Medline: [17928475](https://pubmed.ncbi.nlm.nih.gov/17928475/)]
63. Walter D, Bohmer MM, Reiter S, Krause G, Wichmann O. Risk perception and information-seeking behaviour during the 2009/10 influenza A(H1N1)pdm09 pandemic in Germany. *Euro Surveill* 2012 Mar 29;17(13):pii=20131 [FREE Full text] [Medline: [22490383](https://pubmed.ncbi.nlm.nih.gov/22490383/)]
64. Pang NLS. Crisis-based information seeking: Monitoring versus blunting in the information seeking behaviour of working students during the Southeast Asian Haze Crisis. *Inform Res* 2014 Dec;19(4):online [FREE Full text]
65. Miles J. R - Squared, Adjusted R - Squared. In: *Everitt BS, Howell DC, editors. Encyclopedia of Statistics in Behavioral Science*. Hoboken, NJ: Wiley; 2005.
66. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *J Mach Learn Res* 2003 Mar;3:993-1022 [FREE Full text]
67. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature* 1999 Oct 21;401(6755):788-791. [doi: [10.1038/44565](https://doi.org/10.1038/44565)] [Medline: [10548103](https://pubmed.ncbi.nlm.nih.gov/10548103/)]
68. Blei DM, Lafferty JD. Dynamic topic models. In: *ICML '06: Proceedings of the 23rd International Conference on Machine Learning*. 2006 Jun Presented at: 23rd International Conference on Machine Learning; June 25-29, 2006; Pittsburgh, PA p. 113-120. [doi: [10.1145/1143844.1143859](https://doi.org/10.1145/1143844.1143859)]
69. Gobbo B, Balsamo D, Mauri M, Bajardi P, Panisson A, Ciuccarelli P. Topic Tomographies (TopTom): a visual approach to distill information from media streams. *Comput Graph Forum* 2019 Jul 10;38(3):609-621. [doi: [10.1111/cgf.13714](https://doi.org/10.1111/cgf.13714)]
70. Dou W, Yu L, Wang X, Ma Z, Ribarsky W. HierarchicalTopics: Visually Exploring Large Text Collections Using Topic Hierarchies. *IEEE Trans Visual Comput Graphics* 2013 Dec;19(12):2002-2011. [doi: [10.1109/tvcg.2013.162](https://doi.org/10.1109/tvcg.2013.162)]
71. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*. 2013 Dec Presented at: 26th International Conference on Neural Information Processing Systems; December 5-8, 2013; Lake Tahoe, NV p. 3111-3119.
72. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. *J Doc* 1972 Jan;28(1):11-21. [doi: [10.1108/eb026526](https://doi.org/10.1108/eb026526)]
73. Lin C. Projected gradient methods for nonnegative matrix factorization. *Neural Comput* 2007 Oct;19(10):2756-2779. [doi: [10.1162/neco.2007.19.10.2756](https://doi.org/10.1162/neco.2007.19.10.2756)] [Medline: [17716011](https://pubmed.ncbi.nlm.nih.gov/17716011/)]
74. Hoyer PO. Non-negative matrix factorization with sparseness constraints. *J Mach Learn Res* 2004 Dec;5:1457-1469. [doi: [10.1109/icmlc.2011.6016966](https://doi.org/10.1109/icmlc.2011.6016966)]
75. Chen Y, Skiena S. False-Friend Detection and Entity Matching via Unsupervised Transliteration. *ArXiv*. Preprint posted online on November 21, 2016 [FREE Full text]
76. Python Geocoder. URL: <https://geocoder.readthedocs.io> [accessed 2020-05-29]
77. Greene WH. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall; 2002.
78. Allen MP. Testing hypotheses in nested regression models. In: *Understanding Regression Analysis*. Boston, MA: Springer; 1997:113-117.
79. Zhao WX, Jiang J, Weng J, He J, Lim EP, Yan H, et al. Comparing Twitter and Traditional Media Using Topic Models. In: *Clough P, editor. Advances in Information Retrieval. ECIR 2011. Lecture Notes in Computer Science*, vol 6611. Berlin, Germany: Springer; 2011.
80. Diao Q, Jiang J, Zhu F, Lim EP. Finding bursty topics from microblogs. In: *ACL '12: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*. 2012 Jul Presented at: 50th Annual Meeting of the Association for Computational Linguistics; July 8-14, 2012; Jeju Island, Korea p. 536-544.



81. Bendavid E, Bendavid E, Mulaney B, Sood N, Shah S, Ling E, et al. COVID-19 Antibody Seroprevalence in Santa Clara County, California. medRxiv. Preprint posted online on April 04, 2020 . [doi: [10.1101/2020.04.24.20078824](https://doi.org/10.1101/2020.04.24.20078824)]
82. Bajak A, Howe J. A Study Said Covid Wasn't That Deadly. The Right Seized It. New York Times. 2020 May 14. URL: <https://www.nytimes.com/2020/05/14/opinion/coronavirus-research-misinformation.html> [accessed 2020-09-24]
83. Yong E. Why the Coronavirus Is So Confusing. The Atlantic. 2020 Apr 29. URL: <https://www.theatlantic.com/health/archive/2020/04/pandemic-confusing-uncertainty/610819> [accessed 2020-09-24]
84. Tausczik Y, Faasse K, Pennebaker JW, Petrie KJ. Public anxiety and information seeking following the H1N1 outbreak: blogs, newspaper articles, and Wikipedia visits. Health Commun 2012;27(2):179-185. [doi: [10.1080/10410236.2011.571759](https://doi.org/10.1080/10410236.2011.571759)] [Medline: [21827326](https://pubmed.ncbi.nlm.nih.gov/21827326/)]
85. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: [10.1371/journal.pone.0014118](https://doi.org/10.1371/journal.pone.0014118)] [Medline: [21124761](https://pubmed.ncbi.nlm.nih.gov/21124761/)]
86. Pruss D, Fujinuma Y, Daughton AR, Paul MJ, Arnot B, Albers Szafr D, et al. Zika discourse in the Americas: A multilingual topic analysis of Twitter. PLoS One 2019;14(5):e0216922 [FREE Full text] [doi: [10.1371/journal.pone.0216922](https://doi.org/10.1371/journal.pone.0216922)] [Medline: [31120935](https://pubmed.ncbi.nlm.nih.gov/31120935/)]
87. Smith M, Broniatowski DA, Paul MJ, Dredze M. Towards Real-Time Measurement of Public Epidemic Awareness: Monitoring Influenza Awareness through Twitter. 2015 Presented at: AAAI Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content; March 21-23, 2015; Stanford, CA.
88. Mollema L, Harmsen IA, Broekhuizen E, Clijnk R, De Melker H, Paulussen T, et al. Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013. J Med Internet Res 2015 May 26;17(5):e128 [FREE Full text] [doi: [10.2196/jmir.3863](https://doi.org/10.2196/jmir.3863)] [Medline: [26013683](https://pubmed.ncbi.nlm.nih.gov/26013683/)]
89. Wahlberg AAF, Sjöberg L. Risk perception and the media. J Risk Res 2000 Jan;3(1):31-50 [FREE Full text] [doi: [10.1080/136698700376699](https://doi.org/10.1080/136698700376699)]
90. Klemm C, Das E, Hartmann T. Swine flu and hype: a systematic review of media dramatization of the H1N1 influenza pandemic. J Risk Res 2014 Jun 20;19(1):1-20. [doi: [10.1080/13669877.2014.923029](https://doi.org/10.1080/13669877.2014.923029)]
91. Tchuente JM, Dube N, Bhunu CP, Smith RJ, Bauch CT. The impact of media coverage on the transmission dynamics of human influenza. BMC Public Health 2011 Feb 25;11(S5):online. [doi: [10.1186/1471-2458-11-s1-s5](https://doi.org/10.1186/1471-2458-11-s1-s5)]
92. Johnson BB. Explaining Americans' responses to dread epidemics: an illustration with Ebola in late 2014. J Risk Res 2016 Mar 03;20(10):1338-1357 [FREE Full text] [doi: [10.1080/13669877.2016.1153507](https://doi.org/10.1080/13669877.2016.1153507)]
93. Sell TK, Boddie C, McGinty EE, Pollack K, Smith KC, Burke TA, et al. Media Messages and Perception of Risk for Ebola Virus Infection, United States. Emerg Infect Dis 2017 Jan;23(1):108-111. [doi: [10.3201/eid2301.160589](https://doi.org/10.3201/eid2301.160589)]
94. Gozzi N, Perrotta D, Paolotti D, Perra N. Towards a data-driven characterization of behavioral changes induced by the seasonal flu. PLoS Comput Biol 2020 May;16(5):e1007879 [FREE Full text] [doi: [10.1371/journal.pcbi.1007879](https://doi.org/10.1371/journal.pcbi.1007879)] [Medline: [32401809](https://pubmed.ncbi.nlm.nih.gov/32401809/)]
95. Wise T, Zbozinek T, Micheleni G, Hagan C. Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States. PsyArXiv. Preprint posted online on March 19, 2020 . [doi: [10.31234/osf.io/dz428](https://doi.org/10.31234/osf.io/dz428)]
96. Wang W, Ahern L. Acting on surprise: emotional response, multiple-channel information seeking and vaccination in the H1N1 flu epidemic. Soc Infl 2015 Feb 23;10(3):137-148 [FREE Full text] [doi: [10.1080/15534510.2015.1011227](https://doi.org/10.1080/15534510.2015.1011227)]
97. Duggan M, Smith A. 6% of Online Adults are reddit Users. Pew Research Center. 2103 Jul 03. URL: <https://www.pewresearch.org/internet/2013/07/03/6-of-online-adults-are-reddit-users/> [accessed 2020-09-24]
98. Finlay SC. Age and Gender in Reddit Commenting and Success. J Inf Sci Theory Pract 2014 Sep 30;2(3):18-28. [doi: [10.1633/jistap.2014.2.3.2](https://doi.org/10.1633/jistap.2014.2.3.2)]
99. Singer P, Flock F, Meinhardt C, Zeitfogel E, Strohmaier M. Evolution of reddit: from the front page of the internet to a self-referential community? In: WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web. 2014 Apr Presented at: 23rd International Conference on World Wide Web; April 7-11, 2014; Seoul, Korea p. 517-522. [doi: [10.1145/2567948.2576943](https://doi.org/10.1145/2567948.2576943)]
100. van Deursen AJ, van Dijk JA. The digital divide shifts to differences in usage. New Media Soc 2013 Jun 07;16(3):507-526. [doi: [10.1177/1461444813487959](https://doi.org/10.1177/1461444813487959)]
101. van Deursen A, van Dijk J. Internet skills and the digital divide. New Media Soc 2010 Dec 06;13(6):893-911. [doi: [10.1177/1461444810386774](https://doi.org/10.1177/1461444810386774)]
102. Robinson L, Cotten SR, Ono H, Quan-Haase A, Mesch G, Chen W, et al. Digital inequalities and why they matter. Inf Commun Soc 2015 Mar 16;18(5):569-582 [FREE Full text] [doi: [10.1080/1369118x.2015.1012532](https://doi.org/10.1080/1369118x.2015.1012532)]
103. Guidry JP, Jin Y, Orr CA, Messner M, Meganck S. Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. Public Relat Rev 2017 Sep;43(3):477-486. [doi: [10.1016/j.pubrev.2017.04.009](https://doi.org/10.1016/j.pubrev.2017.04.009)]
104. Martínez-Rojas M, Pardo-Ferreira MDC, Rubio-Romero JC. Twitter as a tool for the management and analysis of emergency situations: A systematic literature review. Int J Inf Manag 2018 Dec;43:196-208. [doi: [10.1016/j.ijinfomgt.2018.07.008](https://doi.org/10.1016/j.ijinfomgt.2018.07.008)]
105. Ferrara E. #COVID-19 on Twitter: Bots, Conspiracies, and Social Media Activism. ArXiv. Preprint posted online on April 20, 2017 [FREE Full text]



106. Yang K, Torres-Lugo C, Menczer F. Prevalence of Low-Credibility Information on Twitter During the COVID-19 Outbreak. ArXiv. Preprint posted online on April 29, 2020 [[FREE Full text](#)]
107. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *J Med Internet Res* 2020 May 06;22(5):e19458 [[FREE Full text](#)] [doi: [10.2196/19458](#)] [Medline: [32352383](#)]
108. Guerrisi C, Turbelin C, Blanchon T, Hanslik T, Bonmarin I, Levy-Bruhl D, et al. Participatory Syndromic Surveillance of Influenza in Europe. *J Infect Dis* 2016 Dec 01;214(suppl\_4):S386-S392. [doi: [10.1093/infdis/jiw280](#)] [Medline: [28830105](#)]
109. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu Near You: Crowdsourced Symptom Reporting Spanning 2 Influenza Seasons. *Am J Public Health* 2015 Oct;105(10):2124-2130. [doi: [10.2105/ajph.2015.302696](#)]
110. Dalton C, Durrheim D, Fejsa J, Francis L, Carlson S, Tursan d'Espaignet E, et al. Flutracking: A weekly Australian community online survey of influenza-like illness in 2006, 2007 and 2008. *Commun Dis Intell* 2009 Dec 07;33(3):316-322 [[FREE Full text](#)]
111. Wójcik OP, Brownstein JS, Chunara R, Johansson MA. Public health for the people: participatory infectious disease surveillance in the digital age. *Emerg Themes Epidemiol* 2014;11:7 [[FREE Full text](#)] [doi: [10.1186/1742-7622-11-7](#)] [Medline: [24991229](#)]
112. COVID-19 News Tracker. Scops. URL: <https://covid19.scops.ai/scops/home/> [accessed 2020-09-24]

## Abbreviations

**API:** application programming interface  
**CDC:** US Centers for Disease Control and Prevention  
**LDA:** latent Dirichlet allocation  
**NMF:** nonnegative matrix factorization  
**TF-IDF:** term frequency-inverse document frequency  
**WHO:** World Health Organization

*Edited by G Eysenbach, R Kukafka; submitted 19.06.20; peer-reviewed by D Garcia, J Barros; comments to author 18.07.20; revised version received 31.07.20; accepted 09.09.20; published 12.10.20.*

### *Please cite as:*

Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, Perra N  
*Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis*  
*J Med Internet Res* 2020;22(10):e21597  
URL: <http://www.jmir.org/2020/10/e21597/>  
doi: [10.2196/21597](#)  
PMID: [32960775](#)

©Nicolò Gozzi, Michele Tizzani, Michele Starnini, Fabio Ciulla, Daniela Paolotti, André Panisson, Nicola Perra. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 12.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Quantifying Public Interest in Police Reforms by Mining Internet Search Data Following George Floyd's Death

John W Ayers<sup>1</sup>, PhD; Benjamin M Althouse<sup>2</sup>, PhD; Adam Poliak<sup>3</sup>, PhD; Eric C Leas<sup>1</sup>, PhD; Alicia L Nobles<sup>1</sup>, PhD; Mark Dredze<sup>4</sup>, PhD; Davey Smith<sup>1</sup>, MD

<sup>1</sup>University of California San Diego, La Jolla, CA, United States

<sup>2</sup>Institute for Disease Modeling, Bill and Melinda Gates Foundation, Seattle, WA, United States

<sup>3</sup>Barnard College, Columbia University, New York, NY, United States

<sup>4</sup>Johns Hopkins University, Baltimore, MD, United States

**Corresponding Author:**

John W Ayers, PhD

University of California San Diego

9500 Gilman Drive

#333 CRSF

La Jolla, CA, 92093

United States

Phone: 1 619 371 1846

Email: [ayers.john.w@gmail.com](mailto:ayers.john.w@gmail.com)

## Abstract

**Background:** The death of George Floyd while in police custody has resurfaced serious questions about police conduct that result in the deaths of unarmed persons.

**Objective:** Data-driven strategies that identify and prioritize the public's needs may engender a public health response to improve policing. We assessed how internet searches indicative of interest in police reform changed after Mr Floyd's death.

**Methods:** We monitored daily Google searches (per 10 million total searches) that included the terms "police" and "reform(s)" (eg, "reform the police," "best police reforms," etc) originating from the United States between January 1, 2010, through July 5, 2020. We also monitored searches containing the term "police" with "training," "union(s)," "militarization," or "immunity" as markers of interest in the corresponding reform topics.

**Results:** The 41 days following Mr Floyd's death corresponded with the greatest number of police "reform(s)" searches ever recorded, with 1,350,000 total searches nationally. Searches increased significantly in all 50 states and Washington DC. By reform topic, nationally there were 1,220,000 total searches for "police" and "union(s)"; 820,000 for "training"; 360,000 for "immunity"; and 72,000 for "militarization." In terms of searches for all policy topics by state, 33 states searched the most for "training," 16 for "union(s)," and 2 for "immunity." States typically in the southeast had fewer queries related to any police reform topic than other states. States that had a greater percentage of votes for President Donald Trump during the 2016 election searched more often for police "union(s)" while states favoring Secretary Hillary Clinton searched more for police "training."

**Conclusions:** The United States is at a historical juncture, with record interest in topics related to police reform with variability in search terms across states. Policy makers can respond to searches by considering the policies their constituencies are searching for online, notably police training and unions. Public health leaders can respond by engaging in the subject of policing and advocating for evidence-based policy reforms.

(*J Med Internet Res* 2020;22(10):e22574) doi:[10.2196/22574](https://doi.org/10.2196/22574)

**KEYWORDS**

policing; digital health; bioinformatics; public health; public interest; data mining; internet; search; trend; Google Trends

## Introduction

The death of George Floyd at the hands of police has resurfaced serious questions about police conduct in the United States that

has resulted in 1001 deaths of unarmed persons from 2013 through 2019 [1]. Despite widespread protests, public interest in police reform following George Floyd's death has not been quantified.

Data-driven strategies that identify and prioritize the needs of the public may engender a public health response to improve policing [2,3]. We argue tracking changes in aggregate internet searches is one such strategy, where the content of searches reflects the thoughts and behaviors of the public, the volume of searches reflect their priority, and the location of the search reflects the community [4-6]. This strategy has informed rapid public health responses across many domains, including responses to demand for unproven therapies during the COVID-19 pandemic [7], demand for HIV testing following Charlie Sheen's public HIV+ disclosure [8], interest in sexual harassment prevention and training following the #MeToo movement [9], and increased suicidal ideation during the initial airing of Netflix's 13 Reasons Why drama [10], to name a few. Moreover, once traditional data became available—in some cases years later—these example findings were confirmed [11,12], underscoring the validity of mining search histories. Similarly, public health leaders can survey internet search trends to rapidly detect public interest in and priorities for police reform.

Herein, we assess public interest in police reforms by mining aggregate internet searches and discovering the types of reforms being searched for by the public, both on the national and US state level (including Washington DC). Such reforms include police training, police unions, police militarization, and the qualified immunity doctrine that reduces the liability police officers face for potential misconduct [13]. To translate these trends into potential policy making, we quantified the police reforms most searched for by each state (including DC) and how each state's political culture reflected by results from the 2016 presidential election explained differences in searches.

## Methods

We monitored the fraction of daily Google searches (per 10 million searches) that included the terms “police” and “reform(s)” originating from the United States between January 1, 2010, through July 5, 2020. This strategy would capture “reform the police,” “best police reforms,” etc as markers for general interest in policing reforms. We also monitored searches containing the term “police” with “training,” “union(s),” “militarization,” or “immunity.” These searches were selected because they have been frequently cited by experts as avenues for reforms [14-16], and represent public prioritization of specific policing reform topics. Search rates were monitored both at the national and state (including DC) level.

### Quantifying the Impact of Mr Floyd's Death

Using historical search rates for each outcome (“police,” “reform[s],” “training,” “union[s],” “militarization,” or “immunity”), we forecasted a counterfactual scenario of expected search rates had Mr Floyd's death not occurred on May 25, 2020. The expected search rates were forecasted using an autoregressive integrated moving average (ARIMA) model,

selected by Hyndman and Khandakar's algorithm [17], fit to the historical search rates using all available data January 1, 2010, through May 24, 2020. The forecasted search rates were compared to observed search rates from May 25, 2020, through July 5, 2020, including computing the ratio of observed and expected search rates with bootstrap confidence intervals by day and cumulatively for the entire post period.

### Do Search Rates for Specific Police Reform Topics Differ Across States?

First, to assess which states had the greatest number of searches for any policy topic, we estimated the cumulative search rates by state for each policy topic, by summing the fraction of searches per 10 million for the entire post period (May 25, 2020, through July 5, 2020) and ranking states accordingly. Second, to identify which specific policy topics had the greater interest within each state, we estimated each state's total searches for any reform topic, by scaling the cumulative search volumes for each specific policy topic by the total cumulative search volume for all policy topics and ranking states accordingly. For each search query outcome, we contrasted any single state with the mean for all states and compared any two single or group of states by these percentages.

### Does Political Culture Impact Searches for Police Reform Policy Topics?

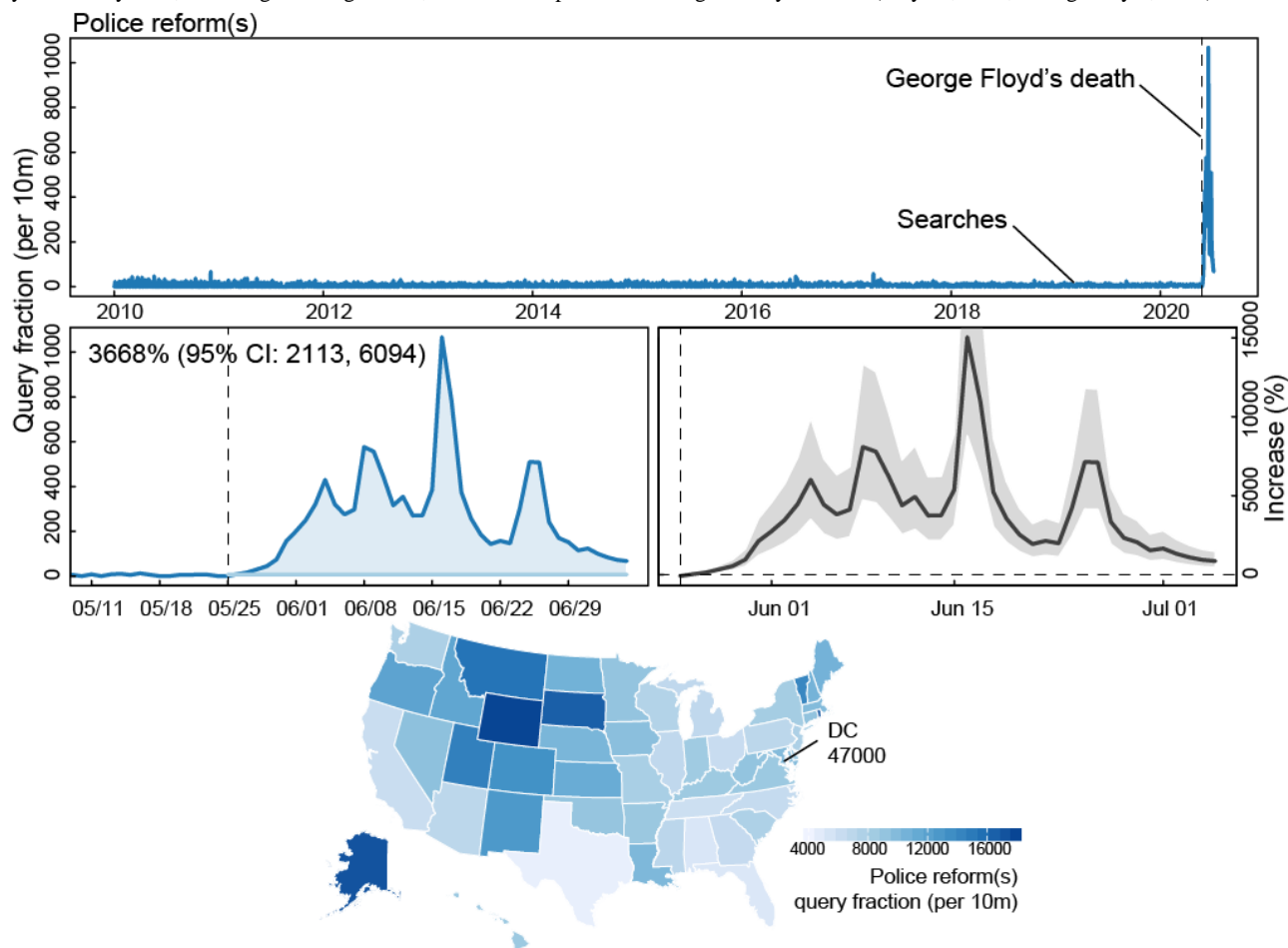
Given the political implications in policy making, we evaluated how political culture impacted searches for specific policy topics, by comparing cumulative search rates by state for each policy topic against their percent vote share for President Donald Trump during the 2016 election. While crude, the presidential percent vote share is often used as an indicator of the political culture of states, reflecting the strength of support or opposition to conservative or liberal policies [18]. For each, we estimated the breakdown of searches for police reform policy topics by states voting for Hillary Clinton or Donald Trump to evaluate any absolute relationships. Secondarily, we evaluated how more support of Trump predicts searches for police reform policy topics by plotting states by their search volume post Mr Floyd's death and percent vote share for Trump, and estimated the best fitting line including the slope of this line. All analyses were performed using R, version 3.5.1 (The R Project for Statistical Computing).

## Results

### Internet Searches for Police “Reform(s)”

The 41 days following Mr Floyd's death corresponded with the greatest number of police “reform(s)” searches ever recorded, with the highest daily fraction of all searches for “reform(s)” eclipsing the past high by over 150-fold (Figure 1). In absolute terms, this translated into about 1,350,000 total searches for police “reform(s).”

**Figure 1.** Internet searches for “police” and “reform(s)” following the death of George Floyd. Queries included all searches with the terms “police” and “reform(s).” Top panel shows the historic trends of reform searches in comparison to after Mr Floyd’s death. Middle panels show the percent increases in searches after Mr Floyd’s death (vertical dashed line is the day of the killing, May 25, 2020), and the bottom map shows the cumulative query fraction by state, including Washington DC, for the entire period following Mr Floyd’s death (May 25, 2020, through July 5, 2020).



Police “reform(s)” search rates cumulatively increased 3668% (95% CI 2113-6094) above expected levels and began increasing on May 25 with queries first eclipsing the past all-time high 5 days later (May 30) and ultimately setting the new record on June 16. During this time, local protests erupted (May 26), police officer Derek Chauvin was charged with third-degree murder and second-degree manslaughter over Mr Floyd’s death (May 29), protests went national (May 31), Mr Floyd’s brother testified at a House Judiciary Committee (June 10), Mr Rayshard Brooks’ death was recorded by video while struggling with police (June 12), and protests reached their peak (June 14) [19]. While queries declined, they remained significantly above expected levels through the last day of observation (July 5).

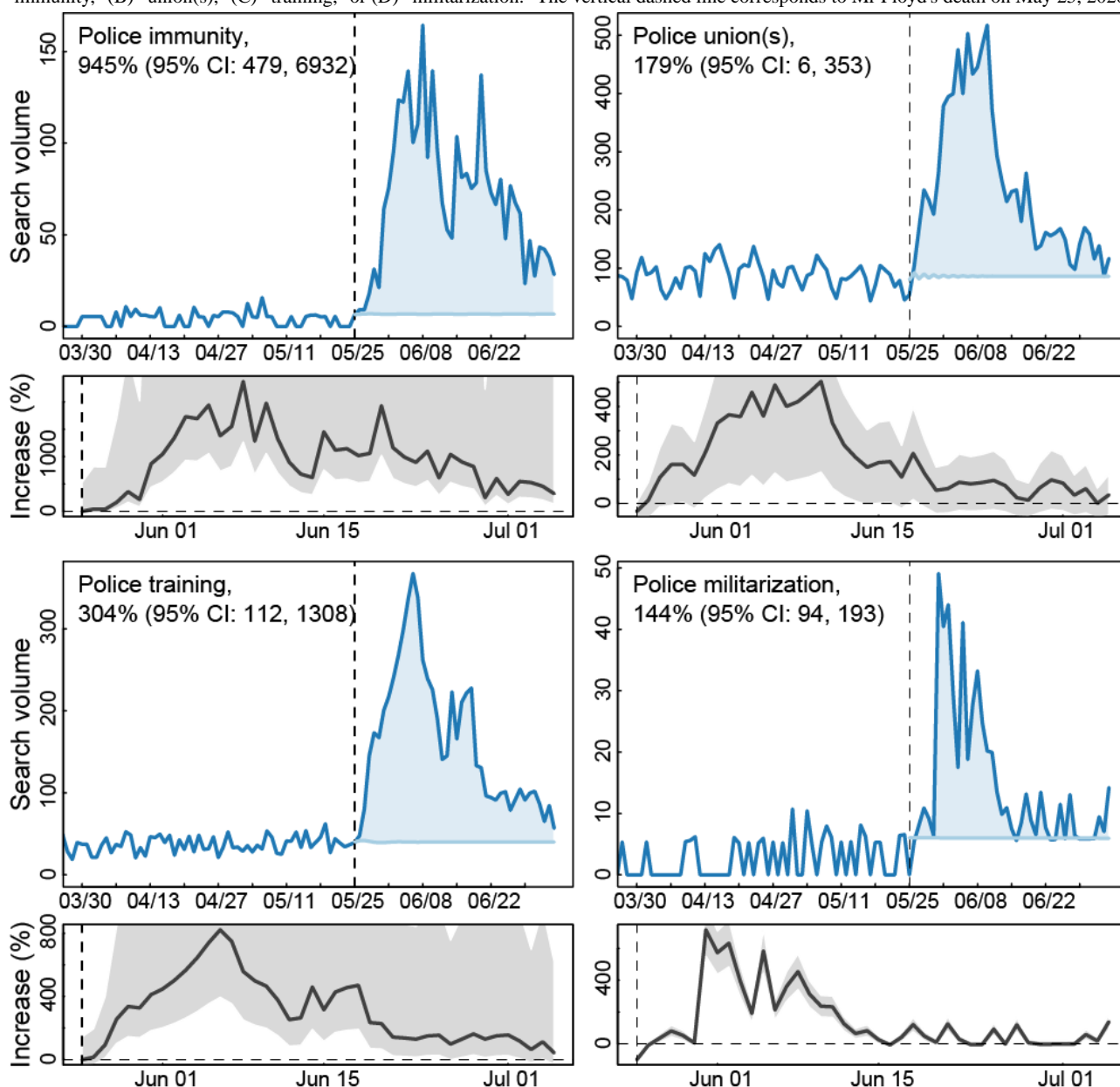
All states and Washington DC had statistically significant increases in search rates for police “reform(s).” By total volume, DC led all states with a total query fraction [QF] of 7300 per 10 million searches during the entire period after Mr Floyd’s death, more than 4.2-fold greater than the average for the states.

DC was followed by Alaska (QF=2760, 1.7-fold higher), Vermont (QF=2670, 1.65-fold higher), and Delaware (QF=2640, 1.64-fold higher). The states with the lowest cumulative search rates included Arkansas (QF=783, 55% below the mean), Maine (QF=1011, 41% below the mean), Kentucky (QF=1040, 40% below the mean), and Louisiana (1160, 33% below the mean).

### Internet Searches for Specific Police Reform Topics

Searches for specific reform topics also set new national benchmarks (Figure 2). Searches for “police” and “union(s)” eclipsed the past all-time highs by 4.5-fold, “training” by 4.8-fold, “immunity” by 53-fold, and “militarization” by 34-fold. Statistically, police “immunity” search rates were cumulatively 945% (95% CI 480-6932) higher than expected during the entire post period, followed by “training” (305%; 95% CI 112-1307), “union(s)” (179%; 95% CI 5.8-353), and “militarization” (144%; 95% CI 94-193). This translates into about 1,220,000 total searches for “union(s),” 820,000 for “training,” 360,000 for “immunity,” and 72,000 for “militarization.”

**Figure 2.** Internet searches for specific police reforms following the death of George Floyd. Queries included all searches with the terms “police” and (A) “immunity,” (B) “union(s),” (C) “training,” or (D) “militarization.” The vertical dashed line corresponds to Mr Floyd’s death on May 25, 2020.



By state, all states including DC had statistically significant increases in searches for specific reform topics after the killing of Mr Floyd, including police “union(s)” (led by Maine, Rhode Island, Wyoming, and West Virginia), “training” (led by Pennsylvania, DC, Minnesota, and Delaware), “immunity” (led by DC, North Dakota, New Mexico, and New Hampshire), and “militarization” (led by DC, New Hampshire, Nevada, and Colorado) (Figure 3). Search rates were somewhat correlated

by state across categories (overall intraclass correlation coefficient=0.31) and the 5 states with the lowest mean search rate across all outcomes after Mr Floyd’s death were Mississippi, Alabama, Tennessee, Texas, and Arkansas (southern states), and 72% (95% CI 64-84) lower than the 5 states with the highest search rate (DC, Pennsylvania, Minnesota, Maine, and Delaware).



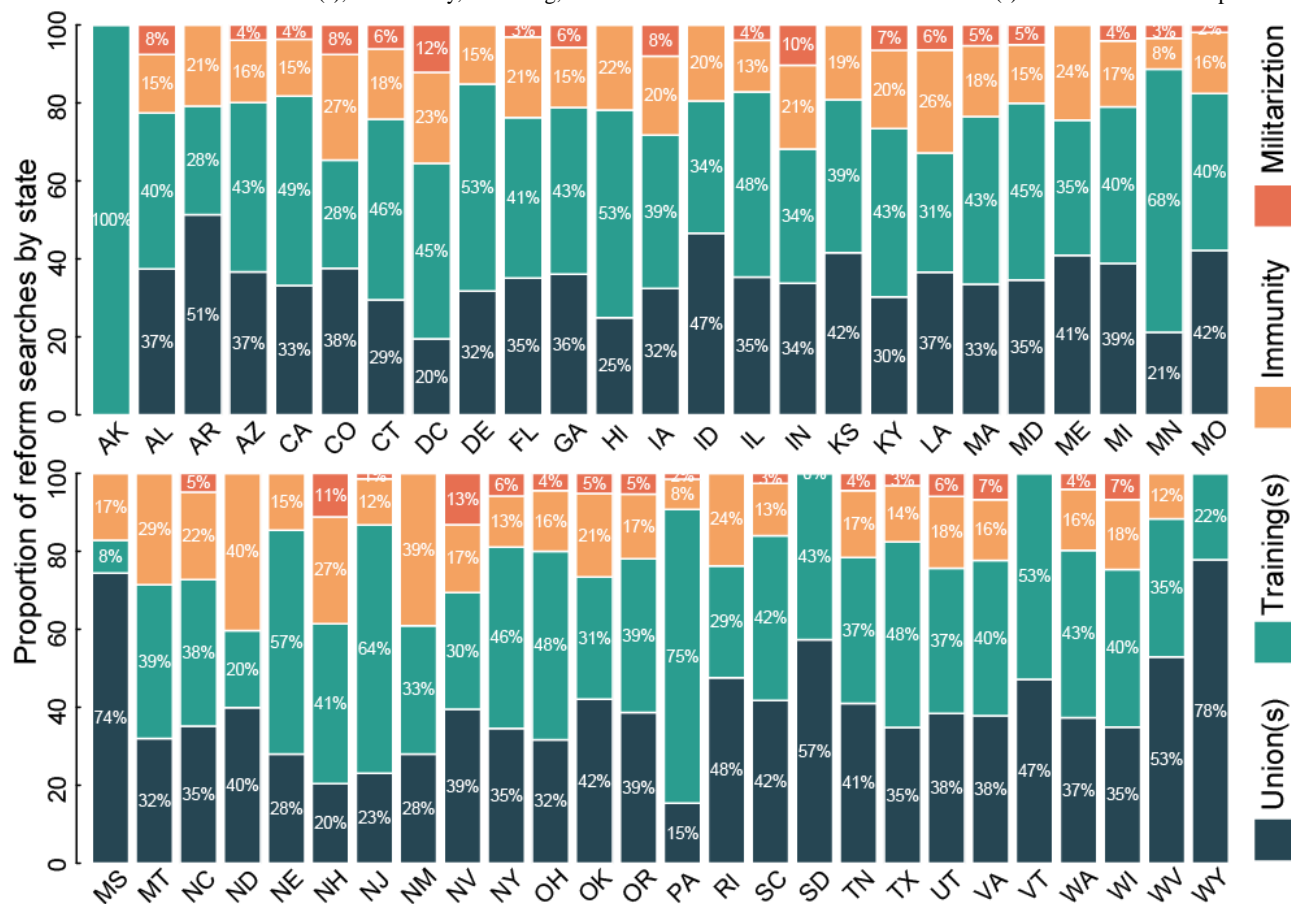
**Figure 3.** Internet searches for specific police reforms following the death of George Floyd, by US state (including Washington DC). Queries included all searches with the terms “police” and (A) “immunity,” (B) “union(s),” (C) “training,” or (D) “militarization.” The value presented is the cumulative query fraction of searches for the focal terms relative to all searches per 10 million for the period after Mr Floyd’s death (May 25, 2020, through July 5, 2020). States are ordered by cumulative query fraction for “training.”



Within states, there were typically more searches for police “training” than any other reform topic, followed by police “union(s),” police “immunity,” and police “militarization” (Figure 4). Still, informative differences emerged in how states favored searching for one specific reform topic over another. In total, 33 states searched more often for police “training” than any other policy, including Alaska, Pennsylvania, Minnesota,

and New Jersey, averaging 76% of the total search volume for all 4 reform search topics. In 16 states, police “union(s)” was searched more often than any other topic, including Wyoming, Mississippi, South Dakota, and West Virginia, averaging 66%. Only 2 states searched more for police “immunity” (North Dakota and New Mexico), and no states showed increased searches for police “militarization” compared to all other topics.

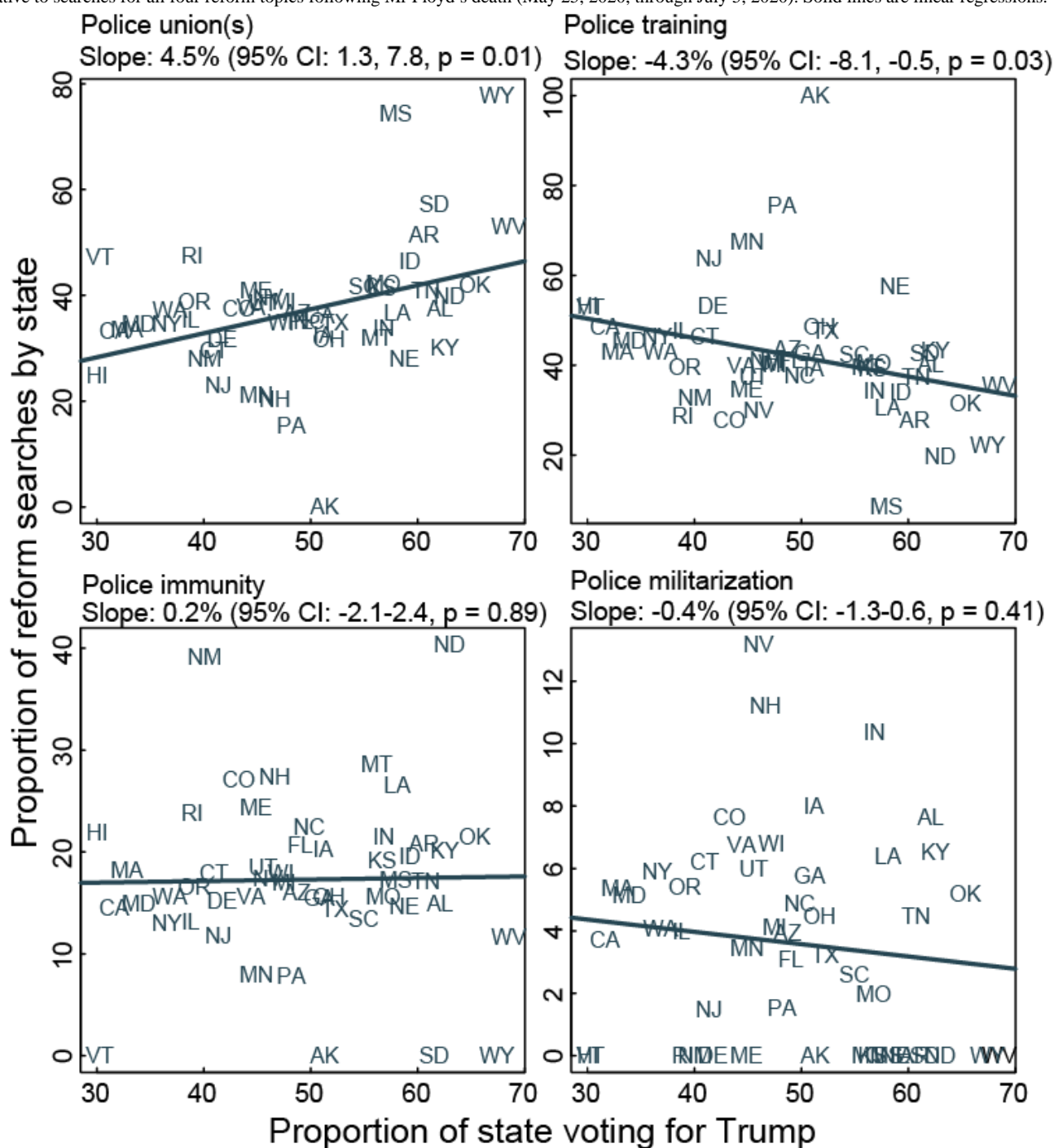
**Figure 4.** Internet searches for specific police reforms following the death of George Floyd, by US state (including Washington DC). Queries included all searches with the terms “police” and (A) “immunity,” (B) “union(s),” (C) “training,” or (D) “militarization,” normalized to the total volume of searches for all four topics by state following Mr Floyd’s death (May 25, 2020, through July 5, 2020). For instance, 37% for police “union(s)” in Alaska means that 37% of searches for “union(s),” “immunity,” “training,” and “militarization” combined were for “union(s).” States are ordered alphabetically.



Of the states won by President Trump during the 2016 presidential election, 57% searched more for police “union(s),” while 81% of states won by Secretary Clinton searched more for “training.” Increased vote share for Trump predicted a dose-response relationship with searches for police “union(s)” or “training” (Figure 5). A 10% increase in vote share for Trump predicted a 4.5% (95% CI 1.3-7.8) increase in cumulative

searches for police “union(s)” following Mr Floyd’s death. A similar increase in Trump vote share predicted a 4.3% (95% CI –8.1 to –0.5) decrease in cumulative searches for police “training.” Regardless of a vote preference for Trump or Clinton, states were no more or less likely to search more for police and “immunity” or “militarization” both absolutely and for a marginal change in vote share for Trump.

**Figure 5.** Internet searches for specific police reforms and electoral support for President Donald Trump. Figure shows the proportion of a state voting for Trump in the 2016 election against the proportion of searches for “police” and (A) “immunity,” (B) “union(s),” (C) “training,” or (D) “militarization,” relative to searches for all four reform topics following Mr Floyd’s death (May 25, 2020, through July 5, 2020). Solid lines are linear regressions.



## Discussion

The United States is at an acute historical juncture with record demand for police reforms sweeping the entire nation. Our results quantify the increasing public interest in police reforms, led by an interest in reforming police training followed by unions, immunity, and militarization nationally. In some states, there is a greater interest in reforming unions or immunity doctrines.

Determining what aspects of police reform to prioritize based on public interest is a challenging problem for elected officials.

We argue that monitoring the search queries of a population might clarify the needs and desires of a constituency in real time, especially when little other data is available. The increase in searches related to police reform indicates that the public is actively interested in police reforms, while searches for specific reform topics indicate a desire for leaders to explore these topics. Indeed, many of the reform topics searched for are backed by strong evidence. Restrictions on police unions that potentially advocate for the status quo and hinder reforms and curtailing police immunity that limit criminal and civil liability for police are both effective [20]. On the other hand, some reforms that may be inclusive of searches for “training” are more complex.

For example, reforms that implement training on procedural justice or the use of force is associated with less police violence [14], but racial sensitivity and implicit bias training are not associated with less police violence [21,22], even though unarmed Black individuals are more likely to be shot than unarmed White individuals [23]. In these cases, a search for training can serve as the basis for leaders to educate the public and seek additional feedback.

The latter raises a serious limitation of our approach; we cannot link search behavior to a policy preference or identify from whom queries originated. Regardless, searches can predict policy priorities on similar public safety issues, and they overcome many of the limitations that often delay traditional data gathering until the point that the data are no longer actionable [24,25]. Moreover, our focus is restricted to a few reform topics. Additional reform topics, such as body cameras (which have mixed empirical effect) [26] or limiting police funding (with unknown impacts and unanticipated effects) [27] can be studied using our approach. It is important that legislative leaders and policy makers listen to the voices of millions of constituents anonymously searching for solutions and consider policies inclusive of their reform interests as a starting point. Notably, the most high-profile police reform entitled “The Justice Act” proposed by Senator Tim Scott addressed training but did not address police unions, immunity, or militarization. Perhaps amendments to this proposed legislation might consider reforms targeting police unions, given that even in Mr Scott’s home state of South Carolina there were more searches related to police unions than training. Since most policy making occurs among states, decision makers there can rely on our results to identify policies that align with the interests of the public they serve. For instance, in North Dakota or New Mexico, policy makers might consider reforms to police immunity since that reform topic was searched more often than other topics. Extending our work to metropolitan areas can likewise inform more local responses.

While both the American Medical Association [28] and the American Public Health Association [29] have spoken out on police violence, the latter calling it a “public health crisis,” health researchers have done little to advance evidence-based police reforms. As of our writing, only 7 studies on PubMed

mention “police” and “misconduct” [30] and 10 mention “police” and “reform(s)” [31] in their title or abstract. This is not to say there is little scientific work on police reform (there is a tremendous literature [32–34]), but little of it has been conducted from a public health perspective or with the benefit of considering strategies popularized by other public health policy successes. For instance, public health advocates have achieved substantial policy reforms to curb deaths attributable to tobacco use [35,36].

Public health scientists, physicians, and other health professionals must do more. Research on policing is essential to public health practice and aligns with public health’s core mission: prevention. The health community has extensive experience studying relevant areas, including unintended medical errors and systematic bias, and can bring those insights to the topic of policing. Moreover, now is the time for health leaders to insist on evidence-based reforms and support science on police reforms that are responsive to the public’s needs.

One way to rapidly leverage science to inform policing is to turn to free and abundant big media data. Researchers could further mine internet search data to answer questions related to police practices. For instance, victims of nonfatal instances of police misconduct might search online for help and these digital tracings could be used to evaluate trends and the geographic distribution of misconduct in almost real time. A similar method is already in practice among health researchers who study drug or device adverse events [37,38] and HIV/AIDS, malaria, and tuberculosis in Africa [39]. Expanding the study of policing to other forms of big media data, including social media, online forums, and news media can similarly provide actionable insights.

Policy changes and public health research in police reform lags while public interest in policing reform is at an all-time high. Research that can help inform policies to improve policing should become a priority among the public health community, with our study serving as a demonstration and call to action. By offering to provide policy makers contemporary insights into policing reform topics their constituents have sought online, we hope that policy makers and elected officials will be inspired to act to prevent unnecessary deaths, like Mr Floyd’s.

## Acknowledgments

This work benefited from intramural support from the Division of Infectious Diseases and Global Public Health and the Center for Data Driven Health at the Qualcomm Institute, both with the University of California San Diego. This work is also supported by a grant from the National Institutes of Health (AI036214).

The funder/sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of their respective employers or funders.

## Conflicts of Interest

MD has received consulting fees from Directing Medicine and Good Analytics, which are companies that advise on the use of digital data for public health surveillance. He has also received payments from Sickweather, which uses social media for infectious

disease monitoring, and Bloomberg LP. JWA owns equity positions in Directing Medicine, Health Watcher, and Good Analytics, which are companies that advise on the use of digital data for public health surveillance. The remaining authors report no conflicts.

## References

1. Mapping Police Violence. URL: <https://mappingpoliceviolence.org/> [accessed 2020-06-14]
2. Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? JAMA 2014 Apr 09;311(14):1399-1400 [FREE Full text] [doi: [10.1001/jama.2014.1505](https://doi.org/10.1001/jama.2014.1505)] [Medline: [24577162](https://pubmed.ncbi.nlm.nih.gov/24577162/)]
3. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
4. Althouse B, Allem J, Childers M, Dredze M, Ayers J. Population health concerns during the United States' Great Recession. Am J Prev Med 2014 Feb;46(2):166-170. [doi: [10.1016/j.amepre.2013.10.008](https://doi.org/10.1016/j.amepre.2013.10.008)] [Medline: [24439350](https://pubmed.ncbi.nlm.nih.gov/24439350/)]
5. Leas EC, Althouse BM, Dredze M, Obradovich N, Fowler JH, Noar SM, et al. Big Data Sensors of Organic Advocacy: The Case of Leonardo DiCaprio and Climate Change. PLoS One 2016;11(8):e0159885 [FREE Full text] [doi: [10.1371/journal.pone.0159885](https://doi.org/10.1371/journal.pone.0159885)] [Medline: [27482907](https://pubmed.ncbi.nlm.nih.gov/27482907/)]
6. Eysenbach G, Köhler C. Health-related searches on the Internet. JAMA 2004 Jun 23;291(24):2946. [doi: [10.1001/jama.291.24.2946](https://doi.org/10.1001/jama.291.24.2946)] [Medline: [15213205](https://pubmed.ncbi.nlm.nih.gov/15213205/)]
7. Liu M, Caputi TL, Dredze M, Kesselheim AS, Ayers JW. Internet Searches for Unproven COVID-19 Therapies in the United States. JAMA Intern Med 2020 Aug 01;180(8):1116-1118 [FREE Full text] [doi: [10.1001/jamainternmed.2020.1764](https://doi.org/10.1001/jamainternmed.2020.1764)] [Medline: [32347895](https://pubmed.ncbi.nlm.nih.gov/32347895/)]
8. Ayers J, Althouse B, Dredze M, Leas E, Noar S. News and Internet Searches About Human Immunodeficiency Virus After Charlie Sheen's Disclosure. JAMA Intern Med 2016 Apr;176(4):552-554. [doi: [10.1001/jamainternmed.2016.0003](https://doi.org/10.1001/jamainternmed.2016.0003)] [Medline: [26902971](https://pubmed.ncbi.nlm.nih.gov/26902971/)]
9. Caputi T, Nobles A, Ayers J. Internet Searches for Sexual Harassment and Assault, Reporting, and Training Since the #MeToo Movement. JAMA Intern Med 2019 Feb 01;179(2):258-259 [FREE Full text] [doi: [10.1001/jamainternmed.2018.5094](https://doi.org/10.1001/jamainternmed.2018.5094)] [Medline: [30575847](https://pubmed.ncbi.nlm.nih.gov/30575847/)]
10. Ayers J, Althouse B, Leas E, Dredze M, Allem J. Internet Searches for Suicide Following the Release of 13 Reasons Why. JAMA Intern Med 2017 Oct 01;177(10):1527-1529 [FREE Full text] [doi: [10.1001/jamainternmed.2017.3333](https://doi.org/10.1001/jamainternmed.2017.3333)] [Medline: [28759671](https://pubmed.ncbi.nlm.nih.gov/28759671/)]
11. Allem J, Leas EC, Caputi TL, Dredze M, Althouse BM, Noar SM, et al. The Charlie Sheen Effect on Rapid In-home Human Immunodeficiency Virus Test Sales. Prev Sci 2017 Jul;18(5):541-544. [doi: [10.1007/s11121-017-0792-2](https://doi.org/10.1007/s11121-017-0792-2)] [Medline: [28516308](https://pubmed.ncbi.nlm.nih.gov/28516308/)]
12. Leas E, Dredze M, Ayers J. Ignoring Data Delays Our Reaction to Emerging Public Health Tragedies Like 13 Reasons Why. JAMA Psychiatry 2019 Sep 25;77(1):103. [doi: [10.1001/jamapsychiatry.2019.2755](https://doi.org/10.1001/jamapsychiatry.2019.2755)] [Medline: [31553419](https://pubmed.ncbi.nlm.nih.gov/31553419/)]
13. Campaign Zero. URL: <https://www.joincampaignzero.org/#vision> [accessed 2020-10-08]
14. Wood G, Tyler TR, Papachristos AV. Procedural justice training reduces police use of force and complaints against officers. Proc Natl Acad Sci U S A 2020 May 05;117(18):9815-9821. [doi: [10.1073/pnas.1920671117](https://doi.org/10.1073/pnas.1920671117)] [Medline: [32312803](https://pubmed.ncbi.nlm.nih.gov/32312803/)]
15. Delehanty C, Mewhirter J, Welch R, Wilks J. Militarization and police violence: The case of the 1033 program. 2017 Jun 14;4(2):205316801771288. [doi: [10.1177/2053168017712885](https://doi.org/10.1177/2053168017712885)]
16. Rushin S. Police Disciplinary Appeals. University of Pennsylvania Law Review 2019;176(3):545 [FREE Full text]
17. Hyndman R, Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R. J Stat Soft 2008;27(3):1-22. [doi: [10.18637/jss.v027.i03](https://doi.org/10.18637/jss.v027.i03)]
18. Fiorina M, Abrams S, Pope J. Culture War? The Myth of a Polarized America. New York, NY: Pearson Longman; 2005.
19. Taylor DB. George Floyd Protests: A Timeline. NY Times. 2020 Jul 10. URL: <https://www.nytimes.com/article/george-floyd-protests-timeline.html> [accessed 2020-10-08]
20. Dharmapala D, McAdams RH, Rappaport J. Collective Bargaining and Police Misconduct: Evidence from Florida. Research Briefs in Economic Policy No. 171 / CATO Institute. 2019 Jul 10. URL: <https://www.cato.org/publications/research-briefs-economic-policy/collective-bargaining-police-misconduct-evidence> [accessed 2020-10-08]
21. Sherman LW. Reducing Fatal Police Shootings as System Crashes: Research, Theory, and Practice. Annu. Rev. Criminol 2018 Jan 13;1(1):421-449. [doi: [10.1146/annurev-criminol-032317-092409](https://doi.org/10.1146/annurev-criminol-032317-092409)]
22. Correll J, Park B, Judd CM, Wittenbrink B. The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. Journal of Personality and Social Psychology 2002;83(6):1314-1329. [doi: [10.1037/0022-3514.83.6.1314](https://doi.org/10.1037/0022-3514.83.6.1314)]
23. Ross CT. A Multi-Level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States, 2011-2014. PLoS One 2015;10(11):e0141854 [FREE Full text] [doi: [10.1371/journal.pone.0141854](https://doi.org/10.1371/journal.pone.0141854)] [Medline: [26540108](https://pubmed.ncbi.nlm.nih.gov/26540108/)]
24. Ayers J, Althouse B, Leas E, Alcorn T, Dredze M. Can Big Media Data Revolutionize Gun Violence Prevention? In: Bloomberg Data for Good Exchange Conference. 2016 Presented at: Bloomberg Data for Good Exchange Conference; Sept 25, 2016; New York City, NY p. 25 URL: [https://www.researchgate.net/publication/309663843\\_Can\\_Big\\_Media\\_Data\\_Revolutionize\\_Gun\\_Violence\\_Prevention](https://www.researchgate.net/publication/309663843_Can_Big_Media_Data_Revolutionize_Gun_Violence_Prevention)



25. Caputi TL, Ayers JW, Dredze M, Suplina N, Burd-Sharps S. Collateral Crises of Gun Preparation and the COVID-19 Pandemic: Infodemiology Study. *JMIR Public Health Surveill* 2020 May 28;6(2):e19369 [FREE Full text] [doi: [10.2196/19369](https://doi.org/10.2196/19369)] [Medline: [32437329](https://pubmed.ncbi.nlm.nih.gov/32437329/)]
26. Yokum D, Ravishankar A, Coppock A. A randomized control trial evaluating the effects of police body-worn cameras. *Proc Natl Acad Sci U S A* 2019 May 21;116(21):10329-10332. [doi: [10.1073/pnas.1814773116](https://doi.org/10.1073/pnas.1814773116)] [Medline: [31064877](https://pubmed.ncbi.nlm.nih.gov/31064877/)]
27. Weichselbaum S, Lews N. Support For Defunding The Police Department Is Growing. Here's Why It's Not A Silver Bullet. The Marshall Project. 2020 Jun 9. URL: <https://www.themarshallproject.org/2020/06/09/support-for-defunding-the-police-department-is-growing-here-s-why-it-s-not-a-silver-bullet> [accessed 2020-10-08]
28. Ehrenfeld J, Harris P. Police Brutality Must Stop. American Medical Association. 2020 May 29. URL: <https://www.ama-assn.org/about/leadership/police-brutality-must-stop> [accessed 2020-10-08]
29. APHA calls out police violence as a public health crisis. American Public Health Association. 2020 Jun 4. URL: <https://www.apha.org/news-and-media/news-releases/apha-news-releases/2020/apha-calls-out-police-violence> [accessed 2020-10-08]
30. ("police"[Title/Abstract] AND "misconduct"[Title/Abstract]). PubMed. 2020. URL: [https://pubmed.ncbi.nlm.nih.gov/?term=\(%22police%22\[Title/Abstract\]+AND+%22misconduct%22\[Title/Abstract\]\)+&filter=simsearch3.fff&filter=ds1.y\\_1&sort=date](https://pubmed.ncbi.nlm.nih.gov/?term=(%22police%22[Title/Abstract]+AND+%22misconduct%22[Title/Abstract])+&filter=simsearch3.fff&filter=ds1.y_1&sort=date) [accessed 2020-10-08]
31. ("police"[Title/Abstract] AND "reform"[Title/Abstract]). PubMed. 2020. URL: [https://pubmed.ncbi.nlm.nih.gov/?term=%28%22police%22%5BTitle%2FAbstract%5D+AND+%22reform%22%5BTitle%2FAbstract%5D%29&filter=simsearch3.fff&filter=ds1.y\\_1&sort=date](https://pubmed.ncbi.nlm.nih.gov/?term=%28%22police%22%5BTitle%2FAbstract%5D+AND+%22reform%22%5BTitle%2FAbstract%5D%29&filter=simsearch3.fff&filter=ds1.y_1&sort=date) [accessed 2020-10-08]
32. Pierson E, Simoiu C, Overgoor J, Corbett-Davies S, Jensen D, Shoemaker A, et al. A large-scale analysis of racial disparities in police stops across the United States. *Nat Hum Behav* 2020 Jul 4;4(7):736-745. [doi: [10.1038/s41562-020-0858-1](https://doi.org/10.1038/s41562-020-0858-1)] [Medline: [32367028](https://pubmed.ncbi.nlm.nih.gov/32367028/)]
33. Mummolo J. Militarization fails to enhance police safety or reduce crime but may harm police reputation. *Proc Natl Acad Sci U S A* 2018 Sep 11;115(37):9181-9186. [doi: [10.1073/pnas.1805161115](https://doi.org/10.1073/pnas.1805161115)] [Medline: [30126997](https://pubmed.ncbi.nlm.nih.gov/30126997/)]
34. Knox D, Mummolo J. Making inferences about racial disparities in police violence. *Proc Natl Acad Sci USA* 2020 Jan 21;117(3):1261-1262. [doi: [10.1073/pnas.1919418117](https://doi.org/10.1073/pnas.1919418117)]
35. Heaton C. The Tobacco Master Settlement Agreement — Strategic Lessons for Addressing Public Health Problems. *N Engl J Med* 2018 Sep 13;379(11):997-1000. [doi: [10.1056/nejmp1802633](https://doi.org/10.1056/nejmp1802633)]
36. Frieden TR. Tobacco control progress and potential. *JAMA* 2014 Jan 08;311(2):133-134. [doi: [10.1001/jama.2013.284534](https://doi.org/10.1001/jama.2013.284534)] [Medline: [24399547](https://pubmed.ncbi.nlm.nih.gov/24399547/)]
37. Yom-Tov E, Lev-Ran S. Adverse Reactions Associated With Cannabis Consumption as Evident From Search Engine Queries. *JMIR Public Health Surveill* 2017 Oct 26;3(4):e77 [FREE Full text] [doi: [10.2196/publichealth.8391](https://doi.org/10.2196/publichealth.8391)] [Medline: [29074469](https://pubmed.ncbi.nlm.nih.gov/29074469/)]
38. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. *J Med Internet Res* 2013 Jun 18;15(6):e124 [FREE Full text] [doi: [10.2196/jmir.2614](https://doi.org/10.2196/jmir.2614)] [Medline: [23778053](https://pubmed.ncbi.nlm.nih.gov/23778053/)]
39. Abebe R, Hill S, Vaughan JW, Small PM, Schwartz HA. Using Search Queries to Understand Health Information Needs in Africa. *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media* 2019;13:3-14 [FREE Full text]

## Abbreviations

**ARIMA:** autoregressive integrated moving average

**QF:** query fraction

*Edited by G Eysenbach; submitted 16.07.20; peer-reviewed by YH Lin, JP Allem, A Hamed; comments to author 04.08.20; revised version received 21.08.20; accepted 21.09.20; published 21.10.20.*

### *Please cite as:*

Ayers JW, Althouse BM, Poliak A, Leas EC, Nobles AL, Dredze M, Smith D  
Quantifying Public Interest in Police Reforms by Mining Internet Search Data Following George Floyd's Death  
*J Med Internet Res* 2020;22(10):e22574  
URL: <http://www.jmir.org/2020/10/e22574/>  
doi: [10.2196/22574](https://doi.org/10.2196/22574)  
PMID: [33084578](https://pubmed.ncbi.nlm.nih.gov/33084578/)

©John W Ayers, Benjamin M Althouse, Adam Poliak, Eric C Leas, Alicia L Nobles, Mark Dredze, Davey Smith. Originally published in the *Journal of Medical Internet Research* (<http://www.jmir.org>), 21.10.2020. This is an open-access article distributed

under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Public Interest in Acne on the Internet: Comparison of Search Information From Google Trends and Naver

Tae Heum Park<sup>1</sup>, MD; Woo Il Kim<sup>1</sup>, MD; Suyeon Park<sup>2</sup>, PhD; Jaeouk Ahn<sup>3</sup>, PhD, MD; Moon Kyun Cho<sup>1</sup>, PhD, MD; Sooyoung Kim<sup>1</sup>, MD

<sup>1</sup>Department of Dermatology, Soonchunhyang University Hospital, Seoul, Republic of Korea

<sup>2</sup>Department of Biostatistics, Soonchunhyang University Hospital, Seoul, Republic of Korea

<sup>3</sup>Department of Medical IT Engineering, Soonchunhyang University, Asan, Republic of Korea

**Corresponding Author:**

Sooyoung Kim, MD

Department of Dermatology

Soonchunhyang University Hospital

59, Daesagwan-ro

Yongsan-gu

Seoul,

Republic of Korea

Phone: 82 2 709 9368

Fax: 82 2 709 9374

Email: [skim1@schmc.ac.kr](mailto:skim1@schmc.ac.kr)

## Abstract

**Background:** Acne vulgaris is a common skin disease primarily affecting young adults. Given that the internet has become a major source of health information, especially among the young, the internet is a powerful tool of communication and has a significant influence on patients.

**Objective:** This study aimed to clarify the features of patients' interest in and evaluate the quality of information about acne vulgaris on the internet.

**Methods:** We compared the search volumes on acne vulgaris with those of other dermatological diseases using Google Trends from January 2004 to August 2019. We also determined the search volumes for relevant keywords of acne vulgaris on Google and Naver and evaluated the quality of answers to the queries in KnowledgeiN.

**Results:** The regression analysis of Google Trends data demonstrated that the patients' interest in acne vulgaris was higher than that for other dermatological diseases, such as atopic dermatitis ( $\beta=-20.33$ , 95% CI  $-22.27$  to  $-18.39$ ,  $P<.001$ ) and urticaria ( $\beta=-27.09$ , 95% CI  $-29.03$  to  $-25.15$ ,  $P<.001$ ) and has increased yearly ( $\beta=2.38$ , 95% CI  $2.05$  to  $2.71$ ,  $P<.001$ ). The search volume for acne vulgaris was significantly higher in the summer than in the spring ( $\beta=-5.04$ , 95% CI  $-9.21$  to  $-0.88$ ,  $P=.018$ ) and on weekends than on weekdays ( $\beta=-6.68$ , 95% CI  $-13.18$  to  $-0.18$ ,  $P=.044$ ). The most frequently searched relevant keywords with "acne vulgaris" and "cause" were "stress," "food," and "cosmetics." Among food, the 2 highest acne vulgaris-related keywords were milk and wheat in Naver and coffee and ramen in Google. The queries in Naver KnowledgeiN were mostly answered by a Korean traditional medicine doctor (53.4%) or the public (33.6%), but only 12.0% by dermatologists.

**Conclusions:** Physicians should be aware of patients' interest in and beliefs about acne vulgaris to provide the best patient education and care, both online and in the clinic.

(*J Med Internet Res* 2020;22(10):e19427) doi:[10.2196/19427](https://doi.org/10.2196/19427)

**KEYWORDS**

acne vulgaris; internet; infodemiology; infoveillance; cosmetics; diet; dermatology; Google

## Introduction

### Background

The internet has spread widely since the 1980s and has become a powerful tool and route of communication. People increasingly use it to find information about health problems. According to the *Measuring the Information Society Report* from 2018 by the International Telecommunication Union, the internet usage rate in South Korea was 95.10% in 2018 [1]. According to 2018 internet usage statistics by the Ministry of Science and Information and Communications Technology of South Korea, the internet usage rate was 91.5% (46,124,694 persons), and the average daily internet usage time in South Korea was 2 hours 15 minutes [2]. As internet usage increases, the influence of the internet is also growing. According to the 2018 internet usage statistics by the Ministry of Science and Information and Communications Technology of South Korea, we use it for viewing movies, videos, and images (87.0%); shopping (62.0%); banking (63.7%); and communication through messengers (95.9%) [2]. In the face of the rapid expansion of smartphones, smartphones have made the internet more ubiquitous; we can get any information wherever we are, which contributes to the percentage of people using the mobile internet (90.4%) in 2018 in South Korea. In particular, young people actively use the internet to exchange their thoughts and get information, including medical and health care information. Among people aged in their 10s, 20s, and 30s, the ratios of people using the internet were 99.9%, 99.9%, and 99.9%, respectively, and those of people using the mobile internet were 98.7%, 99.9%, and 99.9%, respectively [2].

Acne vulgaris is a very common dermatological disease and the 8th most common disease worldwide; it is estimated to affect 633 million people globally [3,4]. It primarily affects young adults, mostly in their teens, 20s, and 30s [5]. Four mechanisms are considered important in the pathogenesis of acne vulgaris: (1) sebaceous hypersecretion, (2) hair follicle hyperkeratosis, (3) colonization of *Cutibacterium acnes*, and (4) inflammatory reactions [6]. External factors such as the environment are also known as important risk factors. Mechanical stimulation like friction, psychological stress [7], and excessive sweating have also been suggested as causes of acne vulgaris. Food is also a possible cause, but the relationship is still unclear, even though there are studies that show a relationship between acne vulgaris and food. The most commonly involved site of acne vulgaris is the face, followed by the neck, back, and chest. Clinically, various kinds of skin lesions are possible: comedon, papule, pustule, and nodule. Deep skin lesions can result in scars, and it is estimated that 95% of acne vulgaris patients have acne scars [8], which can induce serious psychosocial problems [4].

### Goal of This Study

Acne vulgaris is prevalent. It is observed in 80%-90% of adolescents and young adults, and it has high social interest because of its cosmetic sequelae and psychosocial problems. The purpose of this study was to clarify the features of patients' interest in and evaluate the quality of information about acne vulgaris on the internet. Therefore, we aimed to investigate public interest in acne vulgaris by assessing the search volume

of acne vulgaris-related keywords through the major search sites (Naver, Google) that are widely used in South Korea. In addition, we analyzed who answered questions about acne vulgaris on the internet and evaluated the answers posted on internet sites to investigate the accuracy of the information on the internet for public health.

## Methods

### Study Design

We used the search portal sites Naver and Google, which are among the most commonly used search sites in South Korea.

First, we compared the interest in acne vulgaris and other dermatological diseases by using Google Trends, a website that analyzes the popularity of top search queries in Google Search across various regions and languages. Google Trends provides a search volume index (0 to 100), which is the search volume relative to the total number of searches performed on Google. We compared the public interest in acne vulgaris to that of other dermatological disorders, such as atopic dermatitis or urticaria, as well as the yearly and seasonal trends in interest, using regression analysis of 15 years' worth of Google Trends data. From the regression analysis of Google Trends 90-day data, we looked at whether there is a difference between interest on weekends and on weekdays. To determine whether there is a correlation between Google Trends geographic data and personal income, private consumption, or composition of the population, we performed correlation analyses. Keywords used for the search were "acne vulgaris," "urticaria" (urticae or urtica were not included), and "atopic dermatitis" (atopic eczema was also included) in Korean.

Next, we compared search volumes using keywords for "acne vulgaris, treatment," "acne vulgaris, laser," "acne vulgaris, cause," "acne vulgaris, diet," and "acne vulgaris, cosmetics" in Korean in Google and Naver KnowledgeiN. We also sorted 500 answers by responders (dermatologist, general practitioner or specialist medical doctor, Korean medicine doctor, or general public) in Naver KnowledgeiN.

### Statistical Analysis

We used linear regression analysis to observe yearly trends and seasonal trends; correlation analysis to determine correlations between Google Trends geographic data and personal income, private consumption, or composition of population; and text mining to express the search volume of acne vulgaris-related keywords. The statistics program R 3.6.2 was used to analyze data, and  $P < .05$  was considered statistically significant.

## Results

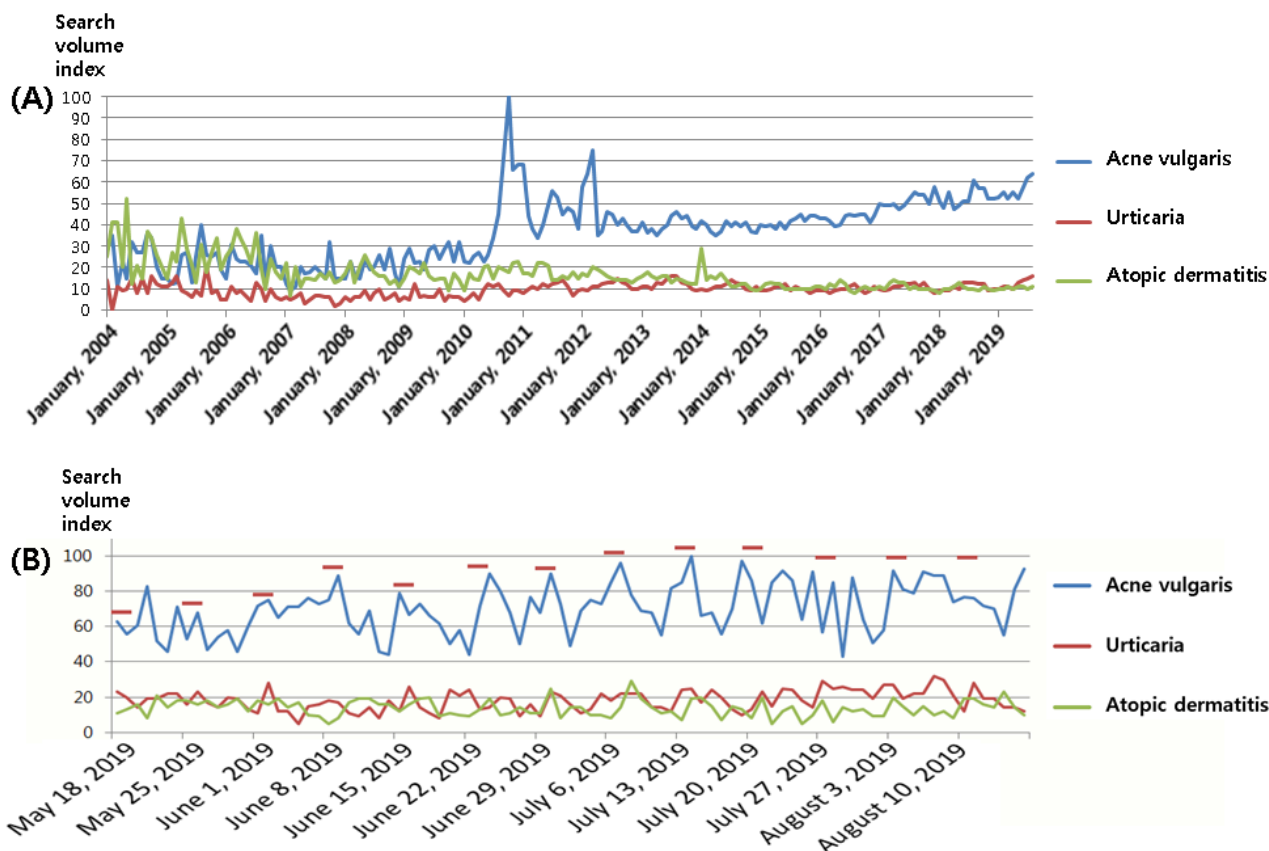
### Acne Vulgaris is One of the Most Popular Dermatological Diseases on the Internet

The regression analysis of Google Trends data for 15 years demonstrated that patients' interest in acne vulgaris was higher than the interest in atopic dermatitis ( $\beta = -20.33$ , 95% CI  $-22.27$  to  $-18.39$ ,  $P < .001$ ) and urticaria ( $\beta = -27.09$ , 95% CI  $-29.03$  to  $-25.15$ ,  $P < .001$ ). Public interest in acne vulgaris has increased yearly ( $\beta = 2.38$ , 95% CI  $2.05$  to  $2.71$ ,  $P < .001$ ) and was higher

in summer than in spring ( $\beta=-5.04$ , 95% CI  $-9.21$  to  $-0.88$ ,  $P=.018$ ). But there was no significant difference between other seasons. For the temporal trends between weekends and

weekdays, the interest in acne vulgaris on weekends was higher than on weekdays ( $\beta=-6.68$ , 95% CI  $-13.18$  to  $-0.18$ ,  $P=.044$ ; Figure 1).

**Figure 1.** Temporal trends in searches for acne: (A) Google Trends time data (15 years). The search volume index (0 to 100) represents the search volumes relative to the total number of Google searches. (B) Google Trends time data (90 days). The horizontal bars indicate weekends.



### Seoul Has the Greatest Interest in Acne Vulgaris on the Internet Among Cities and Provinces

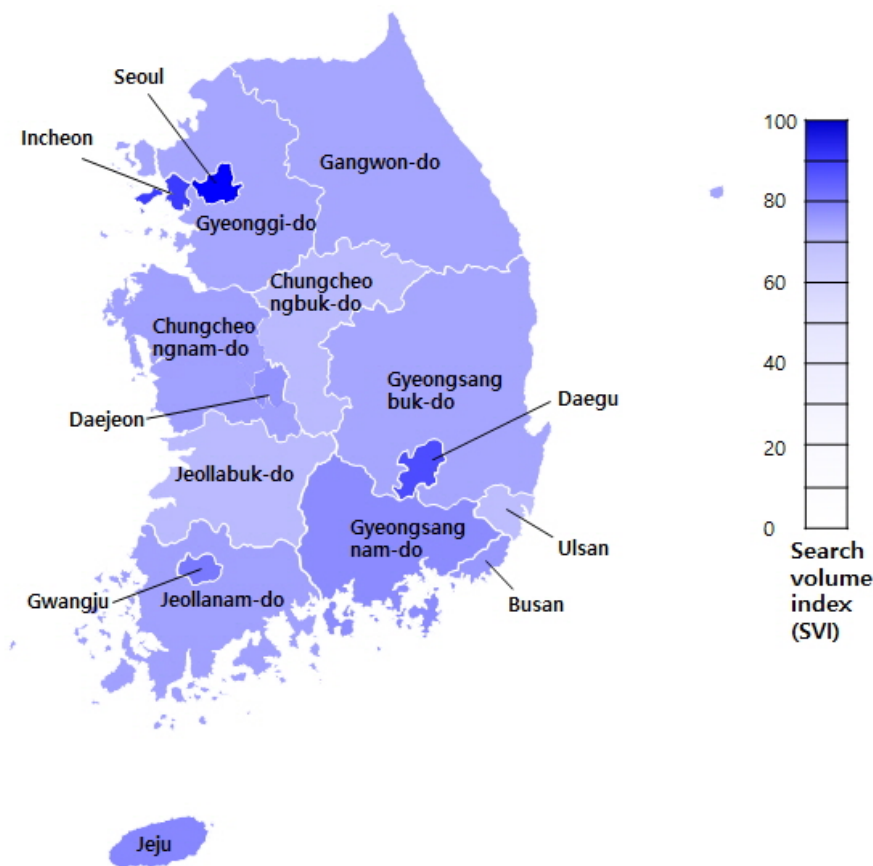
We also determined interest in acne vulgaris among cities and provinces through Google Trends geographic data. Seoul had the highest interest in acne vulgaris among cities and provinces, followed by Incheon and Daegu. Jeollabuk-do and

Chungcheongbuk-do had the lowest interest. There was no significant correlation between search volumes and personal income (correlation coefficient=0.30,  $P=.27$ ) or between search volumes and private consumption (correlation coefficient=0.37,  $P=.16$ ; Table 1; Figure 2). There was also no significant correlation between search volume and age (Table 2).



**Table 1.** Acne vulgaris search volume and personal income in cities and provinces.

Cities and provinces	Search volume	Personal income (annual, KRW)	Private consumption (annual, KRW)
Seoul	100	21,429,000	20,211,000
Incheon	91	17,550,000	14,486,000
Daegu	88	17,568,000	15,682,000
Gwangju	81	17,343,000	16,122,000
Jeju	79	17,464,000	15,107,000
Gyeongsangnam-do	78	16,864,000	14,735,000
Daejeon	77	18,454,000	16,286,000
Busan	76	18,332,000	16,208,000
Chungcheongnam-do	75	17,613,000	14,047,000
Jeollanam-do	75	15,938,000	14,112,000
Gyeonggi-do	74	18,580,000	15,786,000
Gyeongsangbuk-do	74	16,504,000	14,395,000
Gangwon-do	74	16,583,000	14,957,000
Ulsan	73	19,912,000	16,494,000
Jeollabuk-do	71	16,848,000	14,194,000
Chungcheongbuk-do	71	17,030,000	14,381,000

**Figure 2.** Regional trends in internet searches for acne.

**Table 2.** Acne vulgaris search volume and population composition in cities and provinces.

Cities and provinces	Search volume			Ratio of the population in the age range, %					
	2018	2017	2016	10s to 20s (2018)	10s (2018)	10s to 20s (2017)	10s (2017)	10s to 20s (2016)	10s (2016)
Seoul	100	100	89	23.60	8.76	23.76	9.08	23.94	9.41
Incheon	83	81	95	23.76	9.89	24.13	10.23	24.40	10.58
Daegu	88	80	75	23.58	10.14	24.00	10.58	24.31	11.03
Gwangju	73	82	69	25.88	11.78	26.19	12.28	26.48	12.77
Jeju	90	78	100	23.30	11.05	23.53	11.39	23.74	11.81
Gyeongsangnam-do	77	80	80	21.93	10.25	22.31	10.56	22.68	10.90
Daejeon	82	69	99	25.13	10.77	25.43	11.18	25.67	11.62
Busan	73	70	65	21.53	8.61	21.95	8.93	22.31	9.30
Chungcheongnam-do	70	70	88	21.73	10.04	22.05	10.32	22.35	10.61
Jeollanam-do	76	76	81	20.73	9.75	21.04	10.11	21.34	10.50
Gyeonggi-do	77	72	76	23.97	10.59	24.33	10.96	24.59	11.32
Gyeongsangbuk-do	81	69	71	20.37	9.16	20.80	9.45	21.22	9.80
Gangwon-do	80	66	73	21.58	9.72	22.12	10.12	22.46	10.54
Ulsan	63	73	68	23.79	10.52	24.36	10.89	24.86	11.32
Jeollabuk-do	74	64	72	22.23	10.39	22.64	10.78	22.90	11.13
Chungcheongbuk-do	85	73	58	22.55	10.27	22.92	10.40	23.24	10.79

### Most Frequently Searched Relevant Keywords Were Cosmetics, Food, Laser, and Hormone

We searched Naver KnowledgeiN and Google for topics combined with “acne vulgaris” in 5 areas: treatment, cause, cosmetics, laser, and food (Table 3).

When comparing the number of search results from Naver KnowledgeiN with that from Google, Google had a greater number of search results: “acne vulgaris and cause,”

1,078,537/12,797,237, 8.4%, Naver and 11,718,700/12,797,237, 91.6%, Google; “acne vulgaris and food,” 257,137/5,740,137, 4.5%, Naver and 5,483,000/5,740,137, 95.5%, Google; “acne vulgaris and cosmetics,” 1,250,879/14,097,879, 8.9%, Naver and 12,847,000/14,097,879, 91.1%, Google; “acne vulgaris and treatment,” 277,918/2,035,700, 13.7%, Naver and 1,757,782/2,035,700, 86.3%, Google; and “acne vulgaris and laser,” 159,158/1,908,158, 8.3%, Naver and 1,749,000/1,908,158, 91.7%, Google.

**Table 3.** Frequently searched relevant keywords determined by text mining.

Topic	Naver	Google
<b>Acne and cause</b>		
Stress	371,735	1,520,000
Food	265,991	1,890,000
Cosmetics	29,249	6,630,000
<b>Acne and food</b>		
Milk	57,135	478,000
Flour	57,161	167,000
Coffee	27,005	1,460,000
Ramen	4285	1,370,000
<b>Acne and cosmetics</b>		
Toner	334,502	6,020,000
Lotion	332,153	814,000
Cleansing foams	57,135	472,000
Essences	68,667	1,250,000
<b>Acne and treatment</b>		
Lasers	168,750	1,250,000
Antibiotics	54,363	184,000
<b>Acne and laser</b>		
Fraxel laser	40,542	154,000
Photodynamic therapy	29,933	125,000
Toning laser	18,613	428,000
Intense pulsed light	16,379	280,000
LED	1378	316,000

### Information via the Internet Has Been Provided Mostly by Nonexperts

We sorted 500 answers by responders in Naver KnowledgeiN. The queries in Naver KnowledgeiN were mostly answered by Korean traditional medical doctors (267/500, 53.4%), but only 12.0% (60/500) by dermatologists, 33.6% (168/500) by the public, and 1.0% (5/500) by a general practitioner or a specialist medical doctor. The most common question about acne vulgaris was treatment (393/500, 78.6%), followed by cause (61/500, 12.2%), cosmetics (22/500, 4.4%), and laser (9/500, 1.8%).

## Discussion

### Principal Findings

People have greater interest in acne vulgaris than in other dermatological disorders, and patients' interest in acne vulgaris has significantly increased yearly. There were also significant differences in interest between cities and provinces. But there were no significant correlations between public interest and personal income, private consumption, or composition of the population. We selected acne-related topics that people would be curious about: cause, food, cosmetics, treatment, and laser. Among the foods, there were many searches for milk, wheat,

and coffee. Regarding lasers, Fraxel laser, photodynamic therapy (Naver), toning laser, and LED (Google) were of high interest. Answers to the questions on the internet were mostly provided by Korean traditional medical doctors (53.4%) or the public without expertise (33.6%) and only 12.2% by dermatologists.

Public interest in beauty has been increasing in South Korea. According to the International Society of Aesthetic Plastic Surgery, the size of the Korean plastic surgery market was estimated at about 5 trillion won in 2017, which was about a quarter of the world market, and the annual number of plastic surgeries per 1000 people was estimated at 13.5, making it first in the world [9]. The market size of the cosmetics and plastic surgery industries in South Korea has been growing [9,10], which could explain why the interest in acne vulgaris is high and increased much more abruptly than other dermatological disorders, such as atopic dermatitis. Another reason for this high interest may be that acne vulgaris mostly affects young people, who use the internet much more frequently.

People's lifestyles and thoughts are related to internet search volumes. Because acne vulgaris frequently affects young people who work or go to school on weekdays [5], the search volume for acne vulgaris was significantly higher on weekends than on weekdays. Conventionally, people might think that the higher their income, the more they care about their appearance, but

there was no significant correlation between public interest and personal income or private consumption. From this, we can see that people's interest in beauty is universal regardless of economic condition. Related to this finding, some studies have revealed that concern about physical appearance has no association with socioeconomic status [11].

The internet can probably help to identify etiology and epidemiology of disease [12]. It is known that acne vulgaris tends to get worse in the summer [13,14]. High temperatures can change sebum excretion [15], which can worsen acne vulgaris. Consistently, we found that the search volume for acne vulgaris was significantly higher in summer than in spring. Several studies have suggested that dairy products, chocolate, and hyperglycemic foods, such as ramen and flour, can exacerbate acne vulgaris [16-22]. Consistent with these studies, we found that people frequently searched for ramen, flour, chocolate, or milk along with acne vulgaris. Stress and hormones, which were frequently searched by the public on the internet, are also well known to exacerbate acne vulgaris [7,23-25]. Similarly, there are many other studies attempting to estimate epidemiology or track outbreaks of various kinds of diseases, such as influenza, cellulitis, cancer, and tuberculosis, by using the internet [12,26,27].

The power of mass media is extremely strong, even in the medical field [28,29]. In Figure 1A, there was a peak in search volume around the second week of October 2010, probably caused by news articles reported on October 13, 2010, from Korean major broadcasters such as MBC, SBS, or MBN that emphasized the relevance of food and acne vulgaris. In addition to conventional mass media, the internet, smartphones, and social media are also emerging as influential tools [30-32]. In 2010, there was an abrupt spread of smartphones in South Korea, which enabled people to get health information wherever they are and contributed to the peak of search volumes in 2010.

Physicians can use the powerful internet in the medical field [33]. First, through the internet, information about which the public is curious can be easily identified. Second, it can be a helpful tool for patient education because the internet is easily accessible and interactive and can provide media-based material, such as photos and videos. About this, especially in the field of skin cancer, several studies have shown that dermatologists can even effectively change people's skin-related health behavior through an internet intervention and can thus reduce the prevalence of skin diseases [34-36]. Through internet-based deep learning, diagnostic tools for skin cancer are even currently under development [37,38].

However, there is also the probability that false information could be provided, which can also be a major health risk. Much of the information was provided by nonexperts, such as the public or Korean traditional medical doctors, rather than by experts, that is, dermatologists, so the risk is high that false information might spread and endanger patients. Moreover, during data collection, we found manipulation for the purpose of promotion; in Naver KnowledgeiN, someone posted the same questions and adopted only one user as a best answer user. In this way, some people, if they have dishonest intentions, can spread misleading information very easily through the internet. Newspapers and TV news reports also have reported cases of adverse health effects because of incorrect medical information on the internet (Table 4). The Naver online café, Anaki, has greatly restricted the use of medicines for children and has even held chicken pox parties, causing a great deal of controversy. In addition, in the United States, conspiracy theories about vaccination spread on the internet, and the measles vaccination rate dropped significantly, which caused another measles pandemic [39]. As such, in the medical field, the internet has both aspects, in that helpful information can be provided to patients and yet wrong information can be provided and cause serious harm to patients.

**Table 4.** Risk cases for dissemination of unverified facts.

Source or harm	Example
Anaki	The Naver Café, with a membership as high as 60,000 people, caused problems with the dissemination of unfounded medical information, such as neglecting children's illness, refusing vaccinations, throwing chicken pox parties, and selling untested medicines.
Anyemo	This antivaccine movement website refused vaccinations without any reasonable evidence, which could cause pandemics of infectious diseases.
Measles epidemic in the United States	The measles epidemic in the United States has greatly increased because of the decrease in vaccination resulting from unfounded stories about vaccination spread through social media platforms such as Facebook.

## Limitations

There are several limitations to this study. First, Google Trends provides only a relative search volume index, not the absolute search volume, and does not provide a way to calculate the search volume index. Second, we could not find statistically significant factors affecting geographic differences. Personal income, personal consumption, and composition of the population had only weak positive correlations with acne vulgaris. Third, there are many advertisements about acne vulgaris, which might not represent the real interest of the public. Fourth, we could not analyze the exact accuracy of

information, but only estimate its quality through the ratio of who answered the questions in Naver KnowledgeiN. Last, it is not an interventional study. Further interventional studies are needed to measure the real effect of the internet in the real world.

## Conclusions

As part of patient education, we dermatologists need to correct wrong information. In addition, we can use the internet for patient education by launching official websites that provide accurate information about skin diseases or by operating dermatologists' Q&A sites. As such, we need to try both in and outside the office to understand people's interests and beliefs

in the internet space and try to intervene through the internet for patients with skin diseases. and smartphones to provide the best treatment and education

## Acknowledgments

This work was supported by the Soonchunhyang University Research Fund (number 20191261).

## Conflicts of Interest

None declared.

## References

- Measuring the Information Society Report, Volume 1. International Telecommunication Union. 2018. URL: <https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2018/MISR-2018-Vol-1-E.pdf> [accessed 2020-09-22]
- No M, Heo Y, Choi Y, Lee H. 2018 Survey on the Internet Usage. Korean Ministry of Science and ICT. URL: <https://tinyurl.com/yyefhjzk> [accessed 2020-09-22]
- GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* 2016 Oct 8;388(10053):1545-1602 [FREE Full text] [doi: [10.1016/S0140-6736\(16\)31678-6](https://doi.org/10.1016/S0140-6736(16)31678-6)] [Medline: [27733282](https://pubmed.ncbi.nlm.nih.gov/27733282/)]
- Dawson AL, Dellavalle RP. Acne vulgaris. *BMJ* 2013 May 08;346:f2634. [doi: [10.1136/bmj.f2634](https://doi.org/10.1136/bmj.f2634)] [Medline: [23657180](https://pubmed.ncbi.nlm.nih.gov/23657180/)]
- Bhate K, Williams HC. Epidemiology of acne vulgaris. *Br J Dermatol* 2013 Mar;168(3):474-485. [doi: [10.1111/bjd.12149](https://doi.org/10.1111/bjd.12149)] [Medline: [23210645](https://pubmed.ncbi.nlm.nih.gov/23210645/)]
- Simonart T. Immunotherapy for acne vulgaris: current status and future directions. *Am J Clin Dermatol* 2013 Dec;14(6):429-435. [doi: [10.1007/s40257-013-0042-8](https://doi.org/10.1007/s40257-013-0042-8)] [Medline: [24019180](https://pubmed.ncbi.nlm.nih.gov/24019180/)]
- Orion E, Wolf R. Psychologic factors in the development of facial dermatoses. *Clin Dermatol* 2014;32(6):763-766. [doi: [10.1016/j.clindermatol.2014.02.015](https://doi.org/10.1016/j.clindermatol.2014.02.015)] [Medline: [25441469](https://pubmed.ncbi.nlm.nih.gov/25441469/)]
- Fife D. Evaluation of Acne Scars: How to Assess Them and What to Tell the Patient. *Dermatol Clin* 2016 Apr;34(2):207-213. [doi: [10.1016/j.det.2015.11.009](https://doi.org/10.1016/j.det.2015.11.009)] [Medline: [27015781](https://pubmed.ncbi.nlm.nih.gov/27015781/)]
- Plastic Makes Perfect. *The Economist*. 2013 Jan 30. URL: <https://www.economist.com/graphic-detail/2013/01/30/plastic-makes-perfect> [accessed 2020-09-22]
- The International Trade Administration (ITA). South Korea Country Commercial Guide. US Department of Commerce. 2020 Oct 05. URL: <https://www.trade.gov/country-commercial-guides/south-korea-market-overview> [accessed 2020-10-15]
- Harris DL, Carr AT. Prevalence of concern about physical appearance in the general population. *British Journal of Plastic Surgery* 2001 May;54(3):223-226. [doi: [10.1054/bjps.2001.3550](https://doi.org/10.1054/bjps.2001.3550)] [Medline: [11254414](https://pubmed.ncbi.nlm.nih.gov/11254414/)]
- Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health* 2017 Sep;7(3):185-189. [doi: [10.1016/j.jegh.2017.06.001](https://doi.org/10.1016/j.jegh.2017.06.001)] [Medline: [28756828](https://pubmed.ncbi.nlm.nih.gov/28756828/)]
- Celis-Morales C, Livingstone KM, Petermann-Rocha F, Navas-Carretero S, San-Cristobal R, O'Donovan CB, Food4Me Study. Frequent Nutritional Feedback, Personalized Advice, and Behavioral Changes: Findings from the European Food4Me Internet-Based RCT. *Am J Prev Med* 2019 Aug;57(2):209-219. [doi: [10.1016/j.amepre.2019.03.024](https://doi.org/10.1016/j.amepre.2019.03.024)] [Medline: [31248745](https://pubmed.ncbi.nlm.nih.gov/31248745/)]
- Narang I, Sardana K, Bajpai R, Garg VK. Seasonal aggravation of acne in summers and the effect of temperature and humidity in a study in a tropical setting. *J Cosmet Dermatol* 2019 Aug;18(4):1098-1104. [doi: [10.1111/jocd.12777](https://doi.org/10.1111/jocd.12777)] [Medline: [30238598](https://pubmed.ncbi.nlm.nih.gov/30238598/)]
- Williams M, Cunliffe WJ, Williamson B, Forster RA, Cotterill JA, Edwards JC. The effect of local temperature changes on sebum excretion rate and forehead surface lipid composition. *Br J Dermatol* 1973 Mar;88(3):257-262. [doi: [10.1111/j.1365-2133.1973.tb07544.x](https://doi.org/10.1111/j.1365-2133.1973.tb07544.x)] [Medline: [4270005](https://pubmed.ncbi.nlm.nih.gov/4270005/)]
- Adebamowo CA, Spiegelman D, Berkey CS, Danby FW, Rockett HH, Colditz GA, et al. Milk consumption and acne in adolescent girls. *Dermatol Online J* 2006 May 30;12(4):1. [Medline: [17083856](https://pubmed.ncbi.nlm.nih.gov/17083856/)]
- Adebamowo CA, Spiegelman D, Berkey CS, Danby FW, Rockett HH, Colditz GA, et al. Milk consumption and acne in teenaged boys. *J Am Acad Dermatol* 2008 May;58(5):787-793 [FREE Full text] [doi: [10.1016/j.jaad.2007.08.049](https://doi.org/10.1016/j.jaad.2007.08.049)] [Medline: [18194824](https://pubmed.ncbi.nlm.nih.gov/18194824/)]
- Adebamowo CA, Spiegelman D, Danby FW, Frazier AL, Willett WC, Holmes MD. High school dietary dairy intake and teenage acne. *J Am Acad Dermatol* 2005 Feb;52(2):207-214. [doi: [10.1016/j.jaad.2004.08.007](https://doi.org/10.1016/j.jaad.2004.08.007)] [Medline: [15692464](https://pubmed.ncbi.nlm.nih.gov/15692464/)]
- Berra B, Rizzo AM. Glycemic index, glycemic load: new evidence for a link with acne. *J Am Coll Nutr* 2009 Aug;28 Suppl:450S-454S. [doi: [10.1080/07315724.2009.10718111](https://doi.org/10.1080/07315724.2009.10718111)] [Medline: [20234032](https://pubmed.ncbi.nlm.nih.gov/20234032/)]
- Delost GR, Delost ME, Lloyd J. The impact of chocolate consumption on acne vulgaris in college students: A randomized crossover study. *J Am Acad Dermatol* 2016 Jul;75(1):220-222. [doi: [10.1016/j.jaad.2016.02.1159](https://doi.org/10.1016/j.jaad.2016.02.1159)] [Medline: [27317522](https://pubmed.ncbi.nlm.nih.gov/27317522/)]



21. Smith RN, Mann NJ, Braue A, Mäkeläinen H, Varigos GA. A low-glycemic-load diet improves symptoms in acne vulgaris patients: a randomized controlled trial. *Am J Clin Nutr* 2007 Jul;86(1):107-115. [doi: [10.1093/ajcn/86.1.107](https://doi.org/10.1093/ajcn/86.1.107)] [Medline: [17616769](https://pubmed.ncbi.nlm.nih.gov/17616769/)]
22. Claudel JP, Auffret N, Leccia MT, Poli F, Dréno B. Acne and nutrition: hypotheses, myths and facts. *J Eur Acad Dermatol Venereol* 2018 Oct;32(10):1631-1637. [doi: [10.1111/jdv.14998](https://doi.org/10.1111/jdv.14998)] [Medline: [29633388](https://pubmed.ncbi.nlm.nih.gov/29633388/)]
23. Borrel V, Thomas P, Catovic C, Racine P, Konto-Ghiorgi Y, Lefeuvre L, et al. Acne and Stress: Impact of Catecholamines on *Cutibacterium acnes*. *Front Med (Lausanne)* 2019;6:155 [FREE Full text] [doi: [10.3389/fmed.2019.00155](https://doi.org/10.3389/fmed.2019.00155)] [Medline: [31355200](https://pubmed.ncbi.nlm.nih.gov/31355200/)]
24. Ramsay B, Alaghband-Zadeh J, Carter G, Wheeler MJ, Cream JJ. Raised serum androgens and increased responsiveness to luteinizing hormone in men with acne vulgaris. *Acta Derm Venereol* 1995 Jul;75(4):293-296 [FREE Full text] [doi: [10.2340/0001555575293296](https://doi.org/10.2340/0001555575293296)] [Medline: [8578952](https://pubmed.ncbi.nlm.nih.gov/8578952/)]
25. Aizawa H, Niimura M. Serum hormone levels in men with severe acne. *J Dermatol* 1992 Jul;19(7):404-407. [doi: [10.1111/j.1346-8138.1992.tb03249.x](https://doi.org/10.1111/j.1346-8138.1992.tb03249.x)] [Medline: [1401498](https://pubmed.ncbi.nlm.nih.gov/1401498/)]
26. Wiwanitkit V. Google Trends (GT) related to influenza. *Cad Saude Publica* 2015 Jun;31(6):1334. [doi: [10.1590/0102-311XCA020615](https://doi.org/10.1590/0102-311XCA020615)] [Medline: [26200381](https://pubmed.ncbi.nlm.nih.gov/26200381/)]
27. Zhang Y, Bambrick H, Mengersen K, Tong S, Hu W. Using Google Trends and ambient temperature to predict seasonal influenza outbreaks. *Environ Int* 2018 Aug;117:284-291. [doi: [10.1016/j.envint.2018.05.016](https://doi.org/10.1016/j.envint.2018.05.016)] [Medline: [29778013](https://pubmed.ncbi.nlm.nih.gov/29778013/)]
28. McCartney M. Medicine and the Media. Panic about nuclear apocalypse overshadows Japan's real plight. *BMJ* 2011 Mar 22;342:d1845. [doi: [10.1136/bmj.d1845](https://doi.org/10.1136/bmj.d1845)] [Medline: [21427033](https://pubmed.ncbi.nlm.nih.gov/21427033/)]
29. Lipworth W, Kerridge I, Morrell B, Bonfiglioli C, Forsyth R. Medicine, the media and political interests. *J Med Ethics* 2012 Dec;38(12):768-770. [doi: [10.1136/medethics-2012-100875](https://doi.org/10.1136/medethics-2012-100875)] [Medline: [23001949](https://pubmed.ncbi.nlm.nih.gov/23001949/)]
30. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. *J Med Internet Res* 2012 Sep 27;14(5):e128 [FREE Full text] [doi: [10.2196/jmir.1994](https://doi.org/10.2196/jmir.1994)] [Medline: [23017375](https://pubmed.ncbi.nlm.nih.gov/23017375/)]
31. Jia JL, Polin DJ, Sarin KY. Emerging technologies for health information in dermatology: opportunities and drawbacks of web-based searches, social media, mobile applications, and direct-to-consumer genetic testing in patient care. *Semin Cutan Med Surg* 2019 Mar 01;38(1):E57-E53. [doi: [10.12788/j.sder.2019.002](https://doi.org/10.12788/j.sder.2019.002)] [Medline: [31051025](https://pubmed.ncbi.nlm.nih.gov/31051025/)]
32. Croley JAA. Emerging roles of social media in dermatology. *Cutis* 2019 Jan;103(1):E13-E14. [Medline: [30758345](https://pubmed.ncbi.nlm.nih.gov/30758345/)]
33. Nyongesa H, Munguti C, Omondi C, Mokuwa W. Harnessing the power of social media in optimizing health outcomes. *Pan Afr Med J* 2014;18:290 [FREE Full text] [doi: [10.11604/pamj.2014.18.290.4634](https://doi.org/10.11604/pamj.2014.18.290.4634)] [Medline: [25469184](https://pubmed.ncbi.nlm.nih.gov/25469184/)]
34. Heckman CJ, Darlow SD, Ritterband LM, Handorf EA, Manne SL. Efficacy of an Intervention to Alter Skin Cancer Risk Behaviors in Young Adults. *Am J Prev Med* 2016 Jul;51(1):1-11 [FREE Full text] [doi: [10.1016/j.amepre.2015.11.008](https://doi.org/10.1016/j.amepre.2015.11.008)] [Medline: [26810358](https://pubmed.ncbi.nlm.nih.gov/26810358/)]
35. Robinson JK, Alam M, Ashourian N, Khan M, Kundu R, Laumann AE, et al. Skin cancer prevention education for kidney transplant recipients: a systematic evaluation of Internet sites. *Prog Transplant* 2010 Dec;20(4):344-349. [doi: [10.7182/prtr.20.4.9877500752888660](https://doi.org/10.7182/prtr.20.4.9877500752888660)] [Medline: [21265287](https://pubmed.ncbi.nlm.nih.gov/21265287/)]
36. Robinson JK, Gaber R, Hultgren B, Eilers S, Blatt H, Stapleton J, et al. Skin self-examination education for early detection of melanoma: a randomized controlled trial of Internet, workbook, and in-person interventions. *J Med Internet Res* 2014 Jan 13;16(1):e7 [FREE Full text] [doi: [10.2196/jmir.2883](https://doi.org/10.2196/jmir.2883)] [Medline: [24418949](https://pubmed.ncbi.nlm.nih.gov/24418949/)]
37. Wang H, Wang Y, Liang C, Li Y. Assessment of Deep Learning Using Nonimaging Information and Sequential Medical Records to Develop a Prediction Model for Nonmelanoma Skin Cancer. *JAMA Dermatol* 2019 Sep 04;1 [FREE Full text] [doi: [10.1001/jamadermatol.2019.2335](https://doi.org/10.1001/jamadermatol.2019.2335)] [Medline: [31483437](https://pubmed.ncbi.nlm.nih.gov/31483437/)]
38. Han SS, Lim W, Kim MS, Park I, Park GH, Chang SE. Interpretation of the Outputs of a Deep Learning Model Trained with a Skin Cancer Dataset. *J Invest Dermatol* 2018 Oct;138(10):2275-2277 [FREE Full text] [doi: [10.1016/j.jid.2018.05.014](https://doi.org/10.1016/j.jid.2018.05.014)] [Medline: [29864434](https://pubmed.ncbi.nlm.nih.gov/29864434/)]
39. Zipprich J, Winter K, Hacker J, Xia D, Watt J, Harriman K, Centers for Disease Control and Prevention (CDC). Measles outbreak--California, December 2014-February 2015. *MMWR Morb Mortal Wkly Rep* 2015 Feb 20;64(6):153-154 [FREE Full text] [Medline: [25695321](https://pubmed.ncbi.nlm.nih.gov/25695321/)]

*Edited by G Eysenbach; submitted 16.04.20; peer-reviewed by H Santos, R Kaczmarczyk; comments to author 06.07.20; revised version received 17.07.20; accepted 20.07.20; published 26.10.20.*

*Please cite as:*

Park TH, Kim WI, Park S, Ahn J, Cho MK, Kim S

Public Interest in Acne on the Internet: Comparison of Search Information From Google Trends and Naver

*J Med Internet Res* 2020;22(10):e19427

URL: <http://www.jmir.org/2020/10/e19427/>

doi: [10.2196/19427](https://doi.org/10.2196/19427)

PMID: [33104003](https://pubmed.ncbi.nlm.nih.gov/33104003/)

©Tae Heum Park, Woo Il Kim, Suyeon Park, Jaeouk Ahn, Moon Kyun Cho, Sooyoung Kim. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Recommendations From the Twitter Hashtag #DoctorsAreDickheads: Qualitative Analysis

Anjana Estelle Sharma<sup>1,2</sup>, MAS, MD; Ziva Mann<sup>3</sup>, MA; Roy Cherian<sup>2,4</sup>, MHS; Jan Bing Del Rosario<sup>1,5</sup>, BS; Janine Yang<sup>1,6</sup>, BS; Urmimala Sarkar<sup>2</sup>, MD, MPH

<sup>1</sup>Department of Family & Community Medicine, University of California San Francisco, San Francisco, CA, United States

<sup>2</sup>Center for Vulnerable Populations, University of California San Francisco, San Francisco, CA, United States

<sup>3</sup>Ziva Mann Consulting, Newton, MA, United States

<sup>4</sup>Department of Culture and Theory, School of Humanities, University of California, Irvine, Irvine, CA, United States

<sup>5</sup>Berkeley School of Public Health, University of California Berkeley, Berkeley, CA, United States

<sup>6</sup>Drexel University College of Medicine, Philadelphia, PA, United States

**Corresponding Author:**

Anjana Estelle Sharma, MAS, MD  
Department of Family & Community Medicine  
University of California San Francisco  
995 Potrero Ave  
Ward 83  
San Francisco, CA, 94110  
United States  
Phone: 1 6179454776  
Email: [anjana.sharma@ucsf.edu](mailto:anjana.sharma@ucsf.edu)

**Related Article:**

This is a corrected version. See correction statement: <https://www.jmir.org/2020/11/e25511/>

## Abstract

**Background:** The social media site Twitter has 145 million daily active users worldwide and has become a popular forum for users to communicate their health care concerns and experiences as patients. In the fall of 2018, a hashtag titled #DoctorsAreDickheads emerged, with almost 40,000 posts calling attention to health care experiences.

**Objective:** This study aims to identify common health care conditions and conceptual themes represented within the phenomenon of this viral Twitter hashtag.

**Methods:** We analyzed a random sample of 5.67% (500/8818) available tweets for qualitative analysis between October 15 and December 31, 2018, when the hashtag was the most active. Team coders reviewed the same 20.0% (100/500) tweets and the remainder individually. We abstracted the user's health care role and clinical conditions from the tweet and user profile, and used phenomenological content analysis to identify prevalent conceptual themes through sequential open coding, memoing, and discussion of concepts until an agreement was reached.

**Results:** Our final sample comprised 491 tweets and unique Twitter users. Of this sample, 50.5% (248/491) were from patients or patient advocates, 9.6% (47/491) from health care professionals, 4.3% (21/491) from caregivers, 3.7% (18/491) from academics or researchers, 1.0% (5/491) from journalists or media, and 31.6% (155/491) from non-health care individuals or other. The most commonly mentioned clinical conditions were chronic pain, mental health, and musculoskeletal conditions (mainly Ehlers-Danlos syndrome). We identified 3 major themes: disbelief in patients' experience and knowledge that contributes to medical errors and harm, the power inequity between patients and providers, and metacommentary on the meaning and impact of the #DoctorsAreDickheads hashtag.

**Conclusions:** People publicly disclose personal and often troubling health care experiences on Twitter. This adds new accountability for the patient-provider interaction, highlights how harmful communication affects diagnostic safety, and shapes the public's viewpoint of how clinicians behave. Hashtags such as this offer valuable opportunities to learn from patient experiences.

Recommendations include developing best practices for providers to improve communication, supporting patients through challenging diagnoses, and promoting patient engagement.

(*J Med Internet Res* 2020;22(10):e17595) doi:[10.2196/17595](https://doi.org/10.2196/17595)

## KEYWORDS

social media; patient engagement; Twitter messaging; missed diagnosis; internet; physician patient relationship

## Introduction

Twitter is a social media platform for users to share 280-character posts. Globally, Twitter included an average of 145 million daily active users in 2019; [1] 22% of Americans use the platform [2]. Twitter serves as an appealing resource to 61% of American adults who look for web-based health information [3], of which 12% use Twitter or social networking to share health updates [4-6]. Patients and families can develop communities for specific medical conditions and health care-related education [7,8]. Health care professionals utilize Twitter for networking, disseminating medical information, policy and research, disease and health communication monitoring, and advocacy for specific issues [9-11]. By providing this open forum, Twitter enables novel forms of dialogue among and across patients, patient advocacy groups, and health care professionals.

Twitter *hashtags* are words or phrases that Twitter users include in their posts to demarcate posts by a common theme or content. The hashtag can be entered as a search term and demarcate a particular dialogue on the website; therefore, it is a method of indexing conversations. In October 2018, a hashtag titled #DoctorsAreDickheads emerged on Twitter. The term originated from a professional YouTube video maker, who posted a video on Twitter explaining that she had been diagnosed with Ehlers-Danlos syndrome and postural orthostatic tachycardia syndrome (POTS) [12]. After describing her 8-year-long process with various health care professionals to receive the diagnosis, she closed the video with the phrase, “doctors are dickheads.” As users began to respond, #DoctorsAreDickheads emerged as a viral hashtag; patients, caregivers, and health professionals responded to the phrase by sharing their own experiences and criticizing the use of the hashtag. As of July 2020, the hashtag had been used in a total of 37,624 tweets by 12,731 Twitter user accounts and is still in active use (personal communication with Symplur, August 7, 2020).

We sought to describe the concepts represented in this hashtag as part of a broader depiction of patient-driven communications about health. To this end, we conducted a qualitative, phenomenological analysis of this hashtag. Our goal was to describe the *who*, the *what*, and the *how* of this phenomenon: who is posting the hashtag (as per their health care-related role), what is being stated with this hashtag or what are the common medical conditions associated with the hashtag, and how dialogue and prevalent concepts related to it arose. These concepts highlight specific patient and clinician challenges that are voiced publicly on social media and will inform future efforts to improve the patient’s experience of care.

## Methods

### Data Collection

We reviewed all tweets with the hashtag using a report generated from Symplur, a health care social media analytics company [13]. The greatest peak of this hashtag was in late October 2018 (see [Multimedia Appendix 1](#) for the frequency of hashtags over time).

We then obtained tweets containing the hashtag dated between October 24 and December 31, 2018, comprising the predominant wave of use of the hashtag. We were able to obtain all tweets in this period, except for October 25, 2018, because of Symplur’s export limit of 2000 tweets at a time. In total, there were 9670 original tweets during the selected period, of which we were able to extract 8818. We randomly sampled 5.67% (500/8818) tweets for analysis using a random number generator. We selected 500 as the sample size based on a precedent social media analysis; Chan et al [14] demonstrated that a qualitative analysis of a sample of 540 tweets was sufficient to identify themes in how individuals understand and engage in health behaviors. We downloaded all tweets into Google Sheets spreadsheets (Google) for data abstraction and coding and used Microsoft Excel (Microsoft) for tabulation and thematic analysis.

### Data Processing of User Characteristics

We excluded tweets from abstraction if they were not in English, were nonsensical in content, or clearly posted by a bot (eg, had *bot* in the Twitter user profile). As Symplur exported the link to the original tweet, coders in the team reviewed the original Twitter post on the web and read the first tweet above or below the tweet of interest for context when necessary, as tweets can be either standalone comments or part of a conversation or *thread*. Team members abstracted the demographic role of Twitter users by reviewing the user profile and content of the tweet. For example, if a user described an experience while receiving medical care in their tweet, we abstracted their role as *patient*. If a user described themselves as an advocate for a clinical issue in their Twitter profile, we classified them as *advocate*. As there was a substantial overlap in role between *patient* and *patient advocate*, as evidenced by tweet content or profile, we developed a shared category called *patient or patient advocate*. The team also abstracted the clinical conditions mentioned in the tweets, for example, *depression* or *fibromyalgia*.

### Analysis of Themes

This was an interpretative phenomenological analysis, in which a *phenomenon* or lived experience is described by exploring the perspectives and shared meaning of those who have experienced it [15]. We sought to understand for whom, for what, and how

the *viral* hashtag #DoctorsAreDickheads became a way to frame and represent patients' lived experiences on Twitter. We used the iterative process known as a hermeneutic cycle, moving continuously between data collection, interpretation, and theorization and incorporating awareness of our subjective perspectives as researchers in order to develop a nuanced analysis of the phenomenon [16]. This helped us understand how patients interpolate their individual experiences into a wider discourse of patient experience through the hashtag #DoctorsAreDickheads.

Team members (AS, ZM, RC, JY, and JD) coded 20.0% (100/500) sampled tweets together and the remainder individually. Each team member first independently reviewed the content of 50 tweets, selecting a short *code* of a word or short phrase describing the conceptual topics represented in tweets with the hashtag #DoctorsAreDickheads. These codes ascribed what phenomenologically could be called the *essence* or the core meaning of a tweet. Afterward, the team compared codes into a preliminary codebook. The team then reapplied the preliminary codebook to the 50 previously coded tweets and then used it to code another 50 tweets independently. The team then met once more to review code application, discuss new codes, and begin the process of interpretation. Two research team members reconciled differences in code application for the initial 100 tweets that were coded by the entire research team. Once consensus and agreement had been established, we divided the remaining 400 posts among 5 members of the team and coded them independently using the codebook and developing new codes when necessary.

Throughout the process, team members wrote memos or notes to describe shared meaning from the individual codes. After all

remaining tweets were coded, the team sorted and categorized memos based on thematic content, developing an interpretive framework of the hashtag #DoctorsAreDickheads to result in final themes. When discussing possible themes, we also conducted frequency checks to prioritize the themes that were most frequently identified in the data. We aimed to identify all *meanings* of the phenomenon included in the sample; thematic saturation is not a priority in a phenomenological approach [17].

This study was reviewed by the University of San Francisco, California, institutional review board and categorized as *exempt*, with the approval number 19-27965.

## Results

### User Characteristics

Of the 500 tweets analyzed in our sample, 9 were excluded. Reasons for exclusion included being in a language other than English (n=2), written by a bot owing to having *bot* in the profile or nonsensical content (n=5), or containing no content (n=2). The sample included 344 independent Twitter users, with a median tweet frequency of 1 per user; 1 account had 7 posts, 1 had 20 posts, and 1 had 22 posts. In total, 50.5% (248/491) of tweets were posted by patients or patient advocates. Almost one-third or 31.6% (155/491) of tweets were posted by people in the other or unknown category. Health care professionals contributed 9.6% (47/491) of tweets; 4.3% (21/491) tweets were posted by caregivers, 3.7% (18/491) were posted by academics or researchers, and 1.0% (5/491) were posted by media or non-health care organizations. A list of roles identified in the coded sample is available in [Table 1](#).

**Table 1.** Characteristics of Twitter users posting the hashtag #DoctorsAreDickheads (N=491).

Demographics represented in the sample <sup>a</sup>	Values, n (%)
Patient and/or patient advocate	248 (50.5)
Health care provider	47 (9.6)
Caregiver and/or family member	21 (4.3)
Researcher or academic	18 (3.7)
Media, non-health care organization	5 (1.0)
Non-health care individuals or unknown or other	155 (31.6)

<sup>a</sup>Some Twitter posts pertained to multiple demographics.

In our sample, 60.2% (296/491) tweets mentioned a clinical condition. The most common condition mentioned was chronic pain (44 tweets). Mental health, musculoskeletal, and obstetrical or gynecologic conditions and procedures were also common. Ehlers-Danlos syndrome was the most common specific

condition, followed by fibromyalgia, chronic fatigue syndrome (also known as myalgic encephalitis), POTS, and mast cell activation syndrome. A full list of conditions is available in [Table 2](#).



**Table 2.** Clinical conditions mentioned in sample tweets (n=296).

Condition	Number of tweets
Chronic pain	44 (general pain: 38; fibromyalgia: 5)
Mental health	31
Musculoskeletal conditions	26 (Ehlers-Danlos syndrome: 19; other: 7)
Obstetrical or gynecological conditions or procedures	21
Neurological conditions	18 (chronic fatigue syndrome or myalgic encephalitis: 5; POTS <sup>a</sup> : 4; other conditions: 9)
Disability	17
Chronic illness (unspecified condition)	14
Gastrointestinal conditions	8
Autoimmune conditions	7 (mast cell activation syndrome: 4; other autoimmune conditions: 3)

<sup>a</sup>POTS: postural orthostatic tachycardia syndrome.

### Major Themes

We identified 3 core thematic results that were manifested in the experiences represented within our sample. Full definitions of each theme and additional exemplar quotes are found in [Table 3](#). Of note, we are publishing verbatim tweets with usernames

to give credit to Twitter users and their contributions to this discourse when possible. We obtained permission from cited users to publish these tweets. For tweets about which we received no response, we anonymized the content in accordance with recommendations regarding social media research [18].

**Table 3.** Major themes and definitions identified in the content analysis of the Twitter hashtag.

Theme	Definition	Example tweets
Belief and diagnosis	Describing experiences with medical providers being skeptical, dismissive, or “gaslighting”; this disbelief then causing delayed or incorrect diagnosis and/or medical harm	<ul style="list-style-type: none"> <li>“It took 10 years for my MS to be diagnosed. Doctors thought I was embellishing my symptoms and doing too much internet research. If they had spent that time listening, running the correct tests, and treating me, I might not be disabled to the point I am now. #DoctorsAreDickheads” (@VenusDoom14)</li> <li>“Two cardiologists dismissed my POTS as ‘nothing wrong’ or ‘it’s all in your head’ before the third one figured out my POTS. He’s a life-saver, but the other #DoctorsAreDickheads” (@Snarcoleptic_13)</li> <li>“this stings so hard when #DoctorsAreDickheads do this to you while gaslighting you about the psychosomatic nature of your symptoms” (@moniquedhooghe)</li> <li>“I went to an urgent care for what turned out to be pneumonia but had to spend half of the appointment being grilled over why I ‘think’ I have epilepsy. ‘Because the neurologist I’ve been seeing for a decade told me,’ was not good enough. #DoctorsAreDickheads” (@Jenny_Trout)</li> <li>“One doctor I went to, without even knowing me or my history, interrupted me while I was explaining my symptoms &amp; just said ‘You have a psychological condition.’ I said no I don’t.. &amp; he cut me off again &amp; said ‘Yes you do.’ #DoctorsAreDickheads (@d_vaz)</li> </ul>
Power inequity in the patient-provider interaction	Differential in power (due to medical hierarchy as well as misogyny, White supremacy, and ableism) affecting communication and behaviors between clinicians and patients	<ul style="list-style-type: none"> <li>“All I want is to be believed. To have people understand that when sick/stressed, I can’t pretend or act and so my intonation is flat. But they won’t. And, if I wasn’t hairy, if I didn’t have external ‘plumbing,’ this would be worse. #DoctorsAreDickheads” (@theAutistech)</li> <li>“Well, the time to care about my well-being is when I’m in the clinic, but physicians often will not. More often if they are men, and particularly more often if they’re white. This isn’t a stereotype, it’s established in the research. #DoctorsAreDickheads”</li> <li>“#DoctorsAreDickheads is being driven by people living with disabilities and activists that I know. I feel this so deeply – I’ve experienced this bullsh*t even if I’m closer to neurotypical – they confuse us and ignore our own knowledge about our bodies.”</li> <li>“Physicians have all the power. They could help us get better, but for all of us with chronic illness, they’ve traumatized us. We’re too scared to come in to be seen. You can’t get it unless you’ve lived it. #DoctorsAreDickheads”</li> </ul>
Metacommentary	Discussion about the rationale for and impact of this hashtag in public discourse	<ul style="list-style-type: none"> <li>“To all the people that are using #DoctorsAreDickheads first off all Get stuffed (<i>insert: crying laughing emoji</i>) our grouping all doctors into a group that in reality is only made up of like 1% of them. Now am I saying that all doctors are amazing? No but a lot of them work f**king hard and spend time helping others when they could be at home with there family” (@PineappleYT123)</li> <li>“People complaining about the #DoctorsAreDickheads hashtag because it contains a vulgarity... Do you know if patients use curse words (what I call “cuss words” from home) in a medical practice, they can be labeled ‘difficult’?” (@DrZackaryBerger)</li> <li>“I’m sorry, but #DoctorsAreDickheads is simply honest. Some doctors are rude, some are abusive, some are incompetent. Some are brilliant, but that doesn’t mean we can’t discuss the generally poor response to patients who raise issues.” (@WTBDavidG)</li> <li>“I’ve seen as many if not more medical professionals responding positively to #DoctorsAreDickheads in 24hr than I have to more polite debate in the last 8 months.” (@stendec6)</li> <li>“Decent doctors knows they are decent. They understand why the hashtag exists and why patients are suffering. They can deal with a few hurt feelings because they see the greater change that is possible when we stand up for ourselves #DoctorsAreDickheads” (@In-tactCervix)</li> <li>“So much trauma is due to us doctors. We learn best from our patients, but these lessons come too late. The stigma about weight isn’t something we talk about in our training. Let’s do better. #DoctorsAreDickheads”</li> </ul>

## Belief and Diagnosis

Patients and caregivers described a common experience of clinicians not listening, not believing, minimizing, or not valuing their accounts of illness. The experience of being disbelieved was often linked to experiencing an incorrect, delayed, or missed diagnosis. These diagnostic adverse events were associated with physical harm:

*Had terrible blood clots for several years—they said my legs were hurting from fibro and they couldn't do anything. Then I started having trouble breathing and we dashed to the ER. The clots had ended up spreading from my legs to my lungs.*

In addition to physical harm, others described emotional harm and guilt:

*I felt horrible. I was ruining all the holidays, and I could not do a single thing about it. I felt like a worthless piece of crap. And all because... the doctors didn't look at me, did not see my pain as valid. Even now, with my diagnosis.. it is hard... #DoctorsAreDickheads. [@WheelieNick]*

The narratives with this theme described prevalent gaslighting, meaning a manipulative tactic in which someone questions a person's perceptions, memories, and sense of reality. They also described egotistical behavior from clinicians, lording medical training or expertise over patients or being dismissive of patient input when their diagnosis or assessment was challenged:

*A doc told me that I had a cancer syndrome. I said no way – I had been in an accident right before my symptoms started. He told me, "I have an Ivy League degree, so don't ask questions." Turns out he was incorrect. No cancer. #DoctorsAreDickheads.*

Within this theme, we observed how patients countered the narrative of being dismissed by using #DoctorsAreDickheads to create a community where people are believed. In response to a thread in which a patient shared how they "sobbed . . . heaved with the realization that yet again (I'm) being gaslighted about (my) own damn body," another Twitter user responded:

*...please, consider the #DoctorsAreDickheads conversation. This hashtag shows that you're not on your own in this. It isn't just you imagining things. Look at all the people here who believe your words.*

## Power Inequity in the Patient-Provider Interaction

Twitter users described their experiences using this hashtag as a result of the power inequity and hierarchy in medical care. Clinicians hold power in decision making and medical orders, serving as *gatekeeper* for desired services. This included experiences where clinicians denied patient bodily autonomy:

*And then after I was finished having kids and wanted my tubes tied, the first two doctors I asked, refused. For non evidence based reasons. It took me 2.5 years to find a doctor who would. #DoctorsAreDickheads. [@MxPeachyKi]*

The power imbalance impacts communication, and a number of patients described being aware of what they felt they could

or could not say, or *self-edit* what they would express in the visit, to protect themselves from consequences that would affect their care:

*Honestly, if we suddenly go very silent and compliant, we're actually fighting back rage and tears bc we know damn well if we let it show you'll just label us hysterical and FIRE us as patients. Like we somehow serve YOU. #DoctorsAreDickheads. [@rhysfelis]*

Many of the accounts recounted experiences where patients felt that the clinicians were abusing their power. These experiences ranged from subtler, verbal diminishment of the patient experience to physical, verbal, and even sexual abuse and/or severely unethical care:

*cw: sexual assault I had many doctors actively try to cover up or push under the rug the fact that their coworker sexually assaulted me when I was 14. All these people are still practicing at a major hospital, including assaulter. Suffice to say those #DoctorsAreDickheads. [@atoradegay]*

Within this theme was a call for attention to how the patient or clinician power differential is compounded by structural inequities in society. Tweets addressed how White, cisgender, and neurotypical patients have more privileges in medical visits because of structural power imbalances. People of color, lesbian, gay, bisexual, transgender, queer individuals, and people living with disabilities described an intersectional experience, shaped from their identities as members of marginalized groups, in which there was a higher risk of a negative encounter or inappropriate care:

*My #DoctorsAreDickheads story: before I'm a physician, I'm a queer woman. Physicians, nurses, and everyone in healthcare have a long systemic history of abuse of power and broken trust with the LGBTQ+ community. My family & my community fear medicine because of it. Don't @ me. [@ShannonOMac]*

*THANK YOU. It was a little frustrating bc most of the participants were white and I didn't quite know how to articulate that #DoctorsAreDickheads is different when you have other marginalized IDs outside of being disabled or sick. [@Twitchyspoonie]*

## Metacommentary on Hashtag

Although much of the conversation using this hashtag focused on narratives of experiences with clinicians, there was also a *meta* conversation about the meaning of the hashtag's use on Twitter. This included the risks and benefits of using such an inflammatory term:

*Yes this #DoctorsAreDickheads represents poor experiences of care. We try to and should improve this if needed. But doctors are humans, patients, parents, and professionals. Attacking us is counterintuitive when we campaign for improvements. #twitter is this really helpful? [@dr\_nigel\_lane]*

Patient advocates noted that the hashtag got attention, precisely because it was sensational:

*The provocative hashtag #DoctorsAreDickheads drew people's attention to widespread, systemic medical maltreatment. A more polite hashtag couldn't have done this. [ @jeff\_says\_that ]*

Patients, advocates, and health care providers shared their frustration and hoped that this attention and dialogue would allow better understanding and possibly change:

*The medical professionals getting butthurt by #doctorsaredickheads need to read it for what it is - our cries to be treated as PEOPLE first, CONDITIONS second. A desire for inclusivity and a genuine desire to help IMPROVE patient care and the doctor-patient relationship. [ @Chrisa\_Hickey ]*

Within the metacommentary, a parallel hashtag emerged, *#DoctorIRspect*, used to share accounts of laudable or appropriate medical care. These tweets contained recommendations for improved communication during medical visits. Some tweets contained suggestions for engaging patients as partners in their care and system redesign to ensure more patient-centered care:

*So how do you become a #DoctorIRspect? It's really easy. If you have no clue what is going on, just say so. Tell us you don't know. That's all.*

*If you're not involving patients, I urge you to begin doing so. Heck, there are many patients that have experience in orchestrating such change within large medical spaces (insert: waving hand emoji) We're here with experience and even degrees. Hire us to help you. #DoctorsAreDickheads [ @GraysonGoal ]*

## Discussion

Our study paints a picture of patients living with chronic conditions, lacking power within the medical encounter, and turning to social media to share testimonies of being disbelieved and disrespected. Our study is the first empirical approach to analyze the phenomenon of this hashtag on Twitter, utilizing a random sample. Another response was an editorial piece, which concluded that the degree of rancor in the conversation would not aid in improving medical care [19]. Rare or challenging clinical syndromes were commonly mentioned, such as Ehlers-Danlos syndrome, chronic pain, and chronic fatigue syndrome, which have been seen in other analyses of medical topics discussed on Twitter [20,21]. Twitter may be an underutilized resource for understanding the patient's perspective and provider dynamics within the diagnostic pathway, particularly for challenging-to-diagnose conditions; social media data can be mined to monitor care quality and patient experience [22].

The connection between communication within the clinical encounter and the ability to make a correct and timely diagnosis was an unexpected finding. The hashtag *#DoctorsAreDickheads* highlighted how clinician engagement with patients is not just a matter of patient experience, but a priority for diagnostic safety. Patients shared how disrespectful treatment was connected to a missed, delayed, or incorrect diagnosis. Several previous studies on Twitter data found descriptions of

procedural, medication, and diagnostic errors [17,23]; our larger sample validates that patients can self-identify and report diagnostic adverse events. How information is conveyed or received is highly dependent on belief and respect throughout the encounter. Improvements are urgently needed in how clinicians communicate and how patients are involved in the diagnostic pathway [24,25].

The central takeaway from our analysis was how patients felt disempowered, disrespected, and disbelieved. Patients described a range of ways in which they are vulnerable during a health care encounter [26]. Beginning with the existing power differential between provider and patient, tweets described how patients enter the health care system struggling with active symptoms of illness or health needs and can be further burdened by bias, including racism, misogyny, or ableism in the medical system, or by previous traumatic experiences with health care. These vulnerabilities could lead to avoidance or *self-editing*, which hashtag users described as trying to behave like a *good patient* rather than continuing to share their knowledge or expertise in their condition. When experiencing mistreatment, patients are not engaged in sharing knowledge and expertise in their conditions, further reducing the quality of the clinical encounter. Our findings resonate with other works detailing the impacts of power and hierarchy in medicine, in which the lack of agency or respect further perpetuates harm [27,28].

In addition to describing harm from misdiagnosis, patients described assaultive behavior from health care professionals, ranging from hurtful comments (eg, ableist or sexist remarks, carelessness about prognosis) to egregious, unethical, or illegal acts (including sexual abuse). Traditionally, the patient-clinician interaction is private, with few witnesses. Should there be abuse, patients do not have clear-cut routes of action. In the climate of *#MeToo* and the Black Lives Matter movement, people in positions of power are being held to greater accountability for their behavior. As this hashtag demonstrates, Twitter provides greater transparency to clinician behaviors in a public forum.

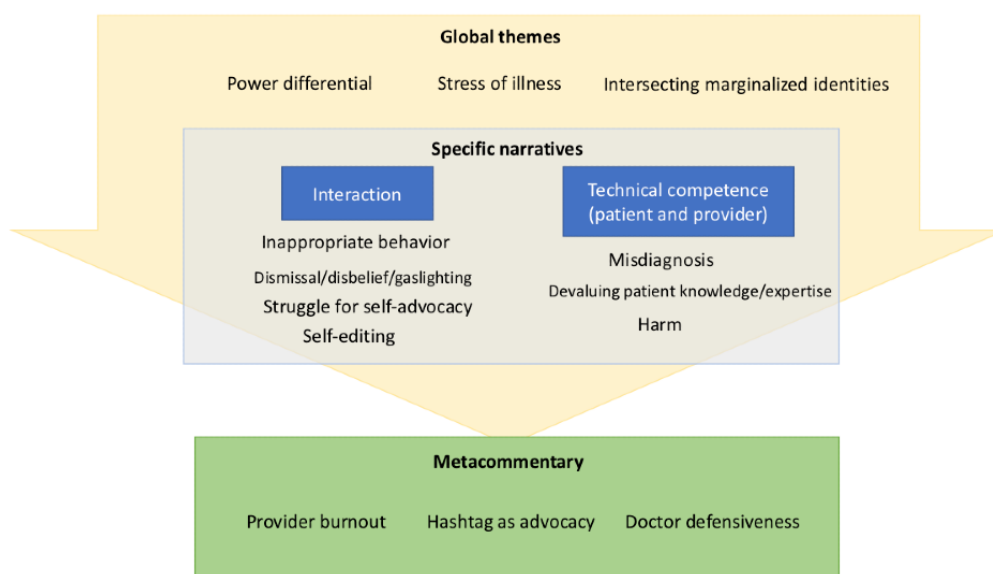
Using this hashtag, tweets demanded empathy on behalf of both clinicians and patients. Within the metacommentary theme, some hashtag users accused patients of exacerbating clinician burnout and mental illness. Patients and advocates responded by highlighting clinicians' privilege, encouraging them to be less defensive and listening to critiques. This opposition between patient needs and clinician burnout is a false dichotomy and speaks to the deeper shortcomings of a medical system that erodes empathy both for patients and for health care personnel. As health care systems continue to explore changes to improve quality and patient experience, work to achieve the quadruple aim is aligned with improving both clinician workplace satisfaction and patient experience [29]. Both patient and clinician wellness are priorities; however, the *#DoctorsAreDickheads* conversation shows how power differentials and limitations in the current health care system disproportionately impact patients, putting their needs in conflict with providers.

Figure 1 shows how the themes relate to each other. The direct experiences of gaslighting, minimization, and ignoring patient expertise (Theme 2) are couched in the overarching themes of

power differential, the burden of illness, and historically marginalized identities (Theme 1). In metacommentary, clinicians have responded both defensively and in support, and

patients and advocates noted the hashtag's utility to gain attention for advocacy for systemic changes (Theme 3).

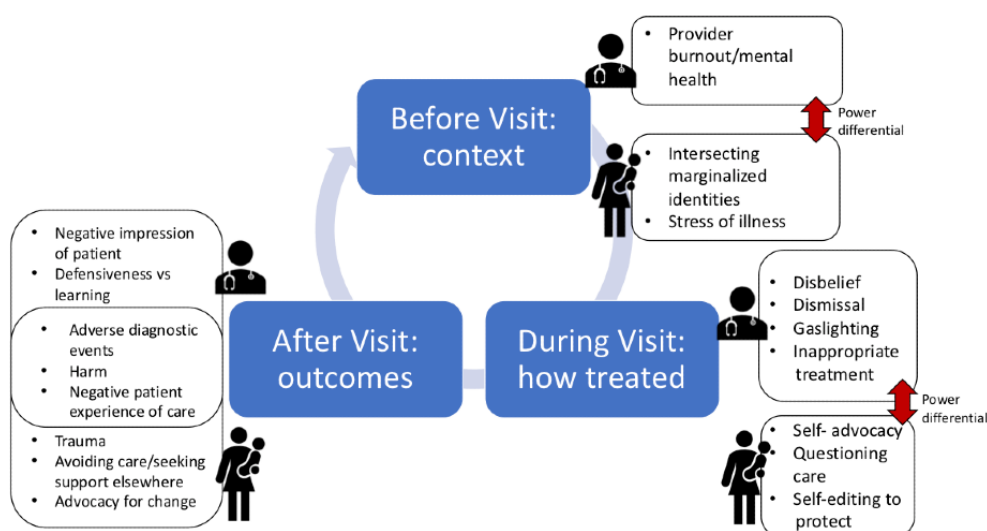
**Figure 1.** Map of major themes identified in Twitter hashtag.



We organized the experiences described in a process map of the patient and clinician encounter, showing how major themes function in a cyclical fashion (Figure 2). First, contextual factors, including the power differential, the patient's intersectional identity, and provider well-being, affect the clinical interaction before a visit. During the encounter, the power differential affects the interaction in which the clinician may disbelieve, devalue, or *gaslight* the patient, who may engage in various

coping behaviors such as self-advocacy, self-editing, or questioning care. In this interaction, downstream negative outcomes include misdiagnosis, medical harm, and negative patient experience or trauma. In this conceptual map, harmful outcomes are shared, highlighting the message of the #DoctorsAreDickheads hashtag as a call to action to improve outcomes for both patients and clinicians.

**Figure 2.** Cyclical process map of themes identified in Twitter hashtag.



Actionable recommendations for clinicians were provided with this phenomenon. First, clinicians can strive to diminish their defensiveness when patients share a negative experience. Second, there is the potential to improve how clinicians communicate in a clinical encounter. Trainings may improve clinician capacity to listen about a health concern, voice diagnostic uncertainty, even by stating "I don't know," and

accompany patients through challenging diagnoses or chronic conditions. Third, patients encouraged increased patient engagement within health care systems, such as through community advisory boards or hiring patients as consultants, to develop more patient-centered care systems [30].



## Limitations

Limitations of this study include a relatively small sample of events; however, our initial sample of 500 is similar to other qualitative analyses of social media posts [8,11]. Those who post on Twitter are not representative of all patients; however, given that 55.2 million people in the United States use Twitter, it is clear that Twitter's users are a sizable proportion of enrolled patients [31]. We do not know if patients who are higher utilizers of health care are more or less likely to post content on Twitter using this hashtag. Finally, we do not know detailed demographic information about Twitter users or their geographic location.

## Conclusions

Twitter and social media are growing platforms where patients discuss health care; this public forum holds clinicians to a higher

level of accountability and transparency. #DoctorsAreDickheads is an intentionally sensational hashtag, born out of frustration with health care interactions. The hashtag is meant to raise awareness of common negative patient experiences, particularly for those living with challenging, rare, and chronic conditions. Patients experience disbelief, mistrust, and lack of listening from their clinicians, which they link to delayed or missed diagnoses. Patients asked for a deeper recognition of the capacity and expertise they bring to the clinical visit and awareness of how power and bias affect the encounter. Although clinicians may feel resistant to concerns expressed through social media, patient advocates on Twitter advocate for system-level improvements to improve medical treatment and patient experience. By systematically exploring views expressed on these platforms, clinicians and health care leaders may identify important areas for improvement, such as improved communication during a challenging diagnosis.

## Acknowledgments

AS was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH) under award number KL2TR001870. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

The authors wish to thank those who generously shared their original posts for this manuscript. The authors thank Jane Lee for assistance with formatting.

## Authors' Contributions

AS and US conceived of the original research idea. AS assembled the research team with input and support from US. RC compiled the data sample. AS, RC, ZM, JD, and JY developed an analytic protocol and conducted data coding, analysis, and interpretation with leadership from AS. AS composed the bulk of the manuscript, with RC contributing to the Methods section, ZM contributing to the Discussion section, and JD and JY contributing to the Introduction and Discussion sections. RC, ZM, JD, JY, and US provided edits to the manuscript. AS, ZM, and US developed the conceptual figures.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Prevalence of hashtag #DoctorsAreDickheads over time.

[PNG File, 45 KB - [jmir\\_v22i10e17595\\_app1.png](#)]

## References

1. Q3 2019 Letter to Shareholders. Twitter Investor Relations. 2019. URL: [https://s22.q4cdn.com/826641620/files/doc\\_financials/2019/q3/Q3-2019-Shareholder-Letter.pdf](https://s22.q4cdn.com/826641620/files/doc_financials/2019/q3/Q3-2019-Shareholder-Letter.pdf) [accessed 2020-08-04]
2. Omnicore Agency. 2019. URL: <https://www.omnicoreagency.com/twitter-statistics/> [accessed 2020-08-04]
3. Perrin A, Anderson M. Share of US Adults Using Social Media, Including Facebook, is Mostly Unchanged Since 2018. Pew Research Center. 2018 Apr 10. URL: <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/> [accessed 2020-08-04]
4. The Social Life of Health Information. Pew Research Center. URL: <https://www.pewresearch.org/internet/2009/06/11/the-social-life-of-health-information/> [accessed 2020-08-04]
5. Pershad Y, Hange P, Albadawi H, Oklu R. Social medicine: Twitter in healthcare. J Clin Med 2018 May 28;7(6):- [FREE Full text] [doi: [10.3390/jcm7060121](https://doi.org/10.3390/jcm7060121)] [Medline: [29843360](#)]
6. Rivas R, Sadah SA, Guo Y, Hristidis V. Classification of health-related social media posts: evaluation of post content-classifier models and analysis of user demographics. JMIR Public Health Surveill 2020 Apr 1;6(2):e14952 [FREE Full text] [doi: [10.2196/14952](https://doi.org/10.2196/14952)] [Medline: [32234706](#)]
7. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. J Med Internet Res 2015 Jul 30;17(7):e188 [FREE Full text] [doi: [10.2196/jmir.4721](https://doi.org/10.2196/jmir.4721)] [Medline: [26228234](#)]

8. Himelboim I, Han JY. Cancer talk on twitter: community structure and information sources in breast and prostate cancer social networks. *J Health Commun* 2014;19(2):210-225. [doi: [10.1080/10810730.2013.811321](https://doi.org/10.1080/10810730.2013.811321)] [Medline: [24111482](https://pubmed.ncbi.nlm.nih.gov/24111482/)]
9. McClellan C, Ali MM, Mutter R, Kroutil L, Landwehr J. Using social media to monitor mental health discussions - evidence from Twitter. *J Am Med Inform Assoc* 2017 May 1;24(3):496-502. [doi: [10.1093/jamia/ocw133](https://doi.org/10.1093/jamia/ocw133)] [Medline: [27707822](https://pubmed.ncbi.nlm.nih.gov/27707822/)]
10. Lee JL, DeCamp M, Dredze M, Chisolm MS, Berger ZD. What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on twitter. *J Med Internet Res* 2014 Oct 15;16(10):e237 [FREE Full text] [doi: [10.2196/jmir.3765](https://doi.org/10.2196/jmir.3765)] [Medline: [25591063](https://pubmed.ncbi.nlm.nih.gov/25591063/)]
11. Alnemer KA, Alhuzaim WM, Alnemer AA, Alharbi BB, Bawazir AS, Barayyan OR, et al. Are health-related tweets evidence based? Review and analysis of health-related tweets on twitter. *J Med Internet Res* 2015 Oct 29;17(10):e246 [FREE Full text] [doi: [10.2196/jmir.4898](https://doi.org/10.2196/jmir.4898)] [Medline: [26515535](https://pubmed.ncbi.nlm.nih.gov/26515535/)]
12. Boebi S. Stevie's On Twitter: New Video! My Biggest Secret: I'm Disabled. Twitter. 2018 Oct 22. URL: <https://twitter.com/stevieboebi/status/1054467827447386112> [accessed 2019-12-09]
13. Symplur - Connecting the Dots in Healthcare Social Media. URL: <https://www.symplur.com/> [accessed 2019-12-16]
14. Chan B, Lopez A, Sarkar U. The canary in the coal mine tweets: social media reveals public perceptions of non-medical use of opioids. *PLoS One* 2015;10(8):e0135072 [FREE Full text] [doi: [10.1371/journal.pone.0135072](https://doi.org/10.1371/journal.pone.0135072)] [Medline: [26252774](https://pubmed.ncbi.nlm.nih.gov/26252774/)]
15. Biggerstaff D, Thompson AR. Interpretative Phenomenological Analysis (IPA): A Qualitative Methodology of Choice in Healthcare Research. *Qualitative Research in Psychology* 2008 Sep 20;5(3):214-224. [doi: [10.1080/14780880802314304](https://doi.org/10.1080/14780880802314304)]
16. Neubauer BE, Witkop CT, Varpio L. How phenomenology can help us learn from the experiences of others. *Perspect Med Educ* 2019 Apr;8(2):90-97 [FREE Full text] [doi: [10.1007/s40037-019-0509-2](https://doi.org/10.1007/s40037-019-0509-2)] [Medline: [30953335](https://pubmed.ncbi.nlm.nih.gov/30953335/)]
17. Brocki JM, Wearden AJ. A critical evaluation of the use of interpretative phenomenological analysis (IPA) in health psychology. *Psychol Health* 2007 Feb 1;21(1):87-108. [doi: [10.1080/14768320500230185](https://doi.org/10.1080/14768320500230185)]
18. Ayers JW, Caputi TL, Nebeker C, Dredze M. Don't quote me: reverse identification of research participants in social media studies. *NPJ Digit Med* 2018;1:30 [FREE Full text] [doi: [10.1038/s41746-018-0036-2](https://doi.org/10.1038/s41746-018-0036-2)] [Medline: [31304312](https://pubmed.ncbi.nlm.nih.gov/31304312/)]
19. Oliver D. David Oliver: incendiary healthcare hashtags. *Br Med J* 2018 Nov 2;363:k4593. [doi: [10.1136/bmj.k4593](https://doi.org/10.1136/bmj.k4593)] [Medline: [30389697](https://pubmed.ncbi.nlm.nih.gov/30389697/)]
20. Nicholl H, Tracey C, Begley T, King C, Lynch AM. Internet use by parents of children with rare conditions: findings from a study on parents' web information needs. *J Med Internet Res* 2017 Feb 28;19(2):e51 [FREE Full text] [doi: [10.2196/jmir.5834](https://doi.org/10.2196/jmir.5834)] [Medline: [28246072](https://pubmed.ncbi.nlm.nih.gov/28246072/)]
21. Kloth YM, Deutsch KM, Danielson KA, Strack J, Law C. What Twitter teaches us about patient-provider communication on pain. *PLoS One* 2019;14(12):e0226321 [FREE Full text] [doi: [10.1371/journal.pone.0226321](https://doi.org/10.1371/journal.pone.0226321)] [Medline: [31877158](https://pubmed.ncbi.nlm.nih.gov/31877158/)]
22. Hawkins JB, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. *BMJ Qual Saf* 2016 Jun;25(6):404-413. [doi: [10.1136/bmjqs-2015-004309](https://doi.org/10.1136/bmjqs-2015-004309)] [Medline: [26464518](https://pubmed.ncbi.nlm.nih.gov/26464518/)]
23. Nakhasi A, Bell SG, Passarella RJ, Paul MJ, Dredze M, Pronovost PJ. The potential of twitter as a data source for patient safety. *J Patient Saf* 2019 Dec;15(4):e32-e35. [doi: [10.1097/PTS.0000000000000253](https://doi.org/10.1097/PTS.0000000000000253)] [Medline: [26756726](https://pubmed.ncbi.nlm.nih.gov/26756726/)]
24. Singh H, Naik AD, Rao R, Petersen LA. Reducing diagnostic errors through effective communication: harnessing the power of information technology. *J Gen Intern Med* 2008 Apr;23(4):489-494 [FREE Full text] [doi: [10.1007/s11606-007-0393-z](https://doi.org/10.1007/s11606-007-0393-z)] [Medline: [18373151](https://pubmed.ncbi.nlm.nih.gov/18373151/)]
25. McDonald KM, Bryce CL, Graber ML. The patient is in: patient involvement strategies for diagnostic error mitigation. *BMJ Qual Saf* 2013 Oct;22(Suppl 2):ii33-ii39. [doi: [10.1136/bmjqs-2012-001623](https://doi.org/10.1136/bmjqs-2012-001623)] [Medline: [23893394](https://pubmed.ncbi.nlm.nih.gov/23893394/)]
26. Berry LL, Danaher TS, Beckham D, Awdish RL, Mate KS. When patients and their families feel like hostages to health care. *Mayo Clin Proc* 2017 Sep;92(9):1373-1381. [doi: [10.1016/j.mayocp.2017.05.015](https://doi.org/10.1016/j.mayocp.2017.05.015)] [Medline: [28843434](https://pubmed.ncbi.nlm.nih.gov/28843434/)]
27. Sharma M. 'Can the patient speak?': postcolonialism and patient involvement in undergraduate and postgraduate medical education. *Med Educ* 2018 May;52(5):471-479. [doi: [10.1111/medu.13501](https://doi.org/10.1111/medu.13501)] [Medline: [29349892](https://pubmed.ncbi.nlm.nih.gov/29349892/)]
28. Abraar K. Medicine's Power Problem. *Scientific American Blogs*. 2019 Nov 09. URL: <https://blogs.scientificamerican.com/voices/medicines-power-problem/> [accessed 2020-08-04]
29. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med* 2014;12(6):573-576. [doi: [10.1370/afm.1713](https://doi.org/10.1370/afm.1713)] [Medline: [25384822](https://pubmed.ncbi.nlm.nih.gov/25384822/)]
30. Bombard Y, Baker GR, Orlando E, Fancott C, Bhatia P, Casalino S, et al. Engaging patients to improve quality of care: a systematic review. *Implement Sci* 2018 Jul 26;13(1):98 [FREE Full text] [doi: [10.1186/s13012-018-0784-z](https://doi.org/10.1186/s13012-018-0784-z)] [Medline: [30045735](https://pubmed.ncbi.nlm.nih.gov/30045735/)]
31. Rimma K. How Many People Are Using Twitter? eMarketer: Data and Research on Digital for Business. 2018 Oct 24. URL: <https://www.emarketer.com/content/the-social-series-who-s-using-twitter-infographic> [accessed 2020-08-02]

## Abbreviations

**NIH:** National Institutes of Health

**POTS:** postural orthostatic tachycardia syndrome

*Edited by G Eysenbach; submitted 24.12.19; peer-reviewed by A Cyr, C Quinlan, A Holden; comments to author 12.06.20; revised version received 06.08.20; accepted 15.09.20; published 28.10.20.*

*Please cite as:*

*Sharma AE, Mann Z, Cherian R, Del Rosario JB, Yang J, Sarkar U*

*Recommendations From the Twitter Hashtag #DoctorsAreDickheads: Qualitative Analysis*

*J Med Internet Res 2020;22(10):e17595*

*URL: <http://www.jmir.org/2020/10/e17595/>*

*doi: [10.2196/17595](https://doi.org/10.2196/17595)*

*PMID: [33112246](https://pubmed.ncbi.nlm.nih.gov/33112246/)*

©Anjana Estelle Sharma, Ziva Mann, Roy Cherian, Jan Bing Del Rosario, Janine Yang, Urmimala Sarkar. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Development and Evaluation of a Digital Intervention for Fulfilling the Needs of Older Migrant Patients With Cancer: User-Centered Design Approach

Hande Sungur<sup>1</sup>, PhD; Nida Gizem Yilmaz<sup>2</sup>, MSc; Brittany Ming Chu Chan<sup>1</sup>, BSc; Maria E T C van den Muijsenbergh<sup>3</sup>, Prof Dr; Julia C M van Weert<sup>1</sup>, Prof Dr; Barbara C Schouten<sup>1</sup>, PhD

<sup>1</sup>Department of Communication Science, Amsterdam School of Communication Research/ASCoR, University of Amsterdam, Amsterdam, Netherlands

<sup>2</sup>Department of Public and Occupational Health, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

<sup>3</sup>Department of Primary and Community Care, Radboud University Medical Centre, Nijmegen, Netherlands

**Corresponding Author:**

Hande Sungur, PhD

Department of Communication Science

Amsterdam School of Communication Research/ASCoR

University of Amsterdam

Postbus 15791

Amsterdam, 1001 NG

Netherlands

Phone: 31 20 525 3680

Email: [h.sungur@uva.nl](mailto:h.sungur@uva.nl)

## Abstract

**Background:** Older migrant patients with cancer face many language- and culture-related barriers to patient participation during medical consultations. To bridge these barriers, an eHealth tool called *Health Communicator* was developed in the Netherlands. Essentially used as a digital translator that can collect medical history information from patients, the Health Communicator did not include an oncological module so far, despite the fact that the prevalence of Dutch migrant patients with cancer is rising.

**Objective:** This study aims to systematically develop, implement, and conduct a pilot evaluation of an oncological module that can be integrated into the Health Communicator to stimulate patient participation among older Turkish-Dutch and Moroccan-Dutch patients with cancer.

**Methods:** The Spiral Technology Action Research model, which incorporates 5 cycles that engage key stakeholders in intervention development, was used as a framework. The *listen* phase consisted of a needs assessment. The *plan* phase consisted of developing the content of the oncological module, namely the question prompt lists (QPLs) and scripts for patient education videos. On the basis of pretests in the *do* phase, 6 audiovisual QPLs on *patient rights, treatment, psychosocial support, lifestyle and access to health care services, patient preferences*, and *clinical trials* were created. Additionally, 5 patient education videos were created about *patient rights, psychosocial support, clinical trials*, and *patient-professional communication*. In the *study* phase, the oncological module was pilot-tested among 27 older Turkish-Dutch and Moroccan-Dutch patients with cancer during their consultations. In the *act* phase, the oncological model was disseminated to practice.

**Results:** The *patient rights* QPL was chosen most often during the pilot testing in the *study* phase. Patients and health care professionals perceived the QPLs as easy to understand and useful. There was a negative correlation between the tool's ease of use and patient age. Patients reported that using the module impacted the consultations positively and thought they were more active compared with previous consultations. Health care professionals also found patients to be more active than usual. Health care professionals asked significantly more questions than patients during consultations. Patients requested to see the *patients' rights* video most often. Patients rated the videos as easy to understand, useful, and informative. Most of the patients wanted to use the tool in the future.

**Conclusions:** Older migrant patients with cancer, survivors, and health care professionals found the oncological module to be a useful tool and have shown intentions to incorporate it into future consultation sessions. Both QPLs and videos were evaluated positively, the latter indicating that the use of narratives to inform older, low-literate migrant patients with cancer about health-related topics in their mother tongue is a viable approach to increase the effectiveness of health care communication with this target group.

**KEYWORDS**

cancer; patient participation; health services needs and demand; eHealth; migrants; physician-patient relations; culture; mobile phone

## Introduction

### Background

Migrant patients across the world are facing numerous adversities within the health care system. For instance, health care professionals communicate less adequately with migrant patients compared with nonmigrant patients [1], and their affective (ie, relating to psychosocial issues such as psychological support) and instrumental needs (ie, relating to more medically related issues such as being informed about treatments) are often not fully met [2]. Health care professionals particularly tend to overlook the importance of ensuring that migrant patients in language-discordant consultations comprehend all the information presented. This can induce dissatisfaction among migrant patients and their family members, ineffective consultation sessions, and ultimately worse patient health outcomes compared with language-concordant consultations [3]. In addition to language barriers, unresolved cultural barriers, such as culturally shaped beliefs about health and illness and communication styles, might also provoke migrant patients' perceptions of lack of health care professionals' respect, inability to participate in the decision-making process, and even perceived inequalities and discrimination in care quality [4-6].

The number of Dutch migrant patients with cancer is expected to triple within the next 20 years, mainly because of the aging of the first generation Turkish-Dutch and Moroccan-Dutch population [7]. However, cancer rates are still low among Dutch migrant patients with cancer; nevertheless, it becomes increasingly important to address language and cultural barriers to adequate health care professional-patient communication and enhance older migrant patients' participation during medical consultations in oncological care. Patient participation refers to "the extent to which patients produce verbal responses that have the potential to significantly influence the content and structure of the interaction as well as the healthcare professional's beliefs and behaviours" [8]. In particular, when it comes to the first generation Turkish-Dutch and Moroccan-Dutch patients with cancer, who often prefer to take along relatives to their consultation to bridge the language barrier instead of professional interpreters, there is a risk of relatives inhibiting patient participation during medical consultations because of culturally shaped beliefs about communication and treatment [9], resulting in providers not getting the patient information they need to provide good quality health care. Hence, interventions that address these barriers are urgently needed.

In general, patient-targeted intervention strategies that have been found to enhance patient participation significantly include coaching, providing educational materials, and having patients offer their opinions to their health care professionals [10,11]. An example of a successful intervention that employed the abovementioned strategies comes from Epstein et al [12]. The

results of this intervention, which was neither targeted at migrant patients nor meant to address language barriers, showed that more patient-centered communication took place during consultations with patients who received question prompt lists (QPLs) and individualized communication coaching, which assisted patients in identifying issues that needed to be addressed during consultation sessions, than during consultations with patients who did not receive these communication strategies. In addition, previous research indicated that a written QPL is beneficial to the communication and psychological and cognitive outcomes of patients with cancer [13].

However, Fukui et al [14] concluded that interventions need to be remodeled to align with the cultural characteristics of their participants to yield satisfactory results. In an intervention tailored to Americans, patients with cancer were provided with statistical information such as recurrence and survival rates. However, information related to the truth of illnesses in Japan was frowned upon among the Japanese participants. In the modified version, patients with cancer were given the liberty to ask for medical information in their own terms. Upon modification, more than 80% of Japanese patients with cancer expressed high satisfaction with the intervention, and no participants dropped out of the intervention compared with a 30% dropout rate in the American version of the intervention. Hence, the results of this study show that interventions must be culturally tailored to be effective.

### This Study

Given the lack of interventions tailored specifically to older Turkish-Dutch and Moroccan-Dutch migrant patients with cancer to combat the barriers they experience in communicating with their health care professionals [15,16], this study set out to systematically develop and implement an intervention to improve their participation and satisfaction with care. As the Turkish and Moroccan community has been growing as a result of migration, and recent figures have shown that they not only make up approximately 5% of the Dutch population but also account for being two of the fastest-growing populations in the last 5 years [17,18], we focused on older Turkish-Dutch and Moroccan-Dutch patients with cancer (>55 years). We developed an intervention, more specifically an oncological module, to be implemented in an existing Dutch eHealth tool called *Health Communicator*. The Health Communicator is a web-based digital tool that aims to resolve language barriers between non-Western patients with low Dutch language proficiency and their health care professionals. The Health Communicator is used to collect patient medical anamnestic data via digital questionnaires and to provide educational videos for patients in multiple languages about their illness. Although the Health Communicator includes medical history questionnaires and educational videos for various medical topics (eg, diabetes, pregnancy), it lacked a module specifically aimed at oncology patients, especially one systematically developed



for older Turkish-Dutch and Moroccan-Dutch migrant patients with cancer ([Multimedia Appendix 1](#)).

Our goal of implementing this new oncological module is further supported by evidence pointing toward the fact that older adults and Turkish-Dutch patients are more inclined to use media than interpersonal sources to gather health information, ranging from the television to the internet [19]. In a recent study, health care professionals have expressed positive attitudes and intentions to use the existing Health Communicator tool to increase patient participation (Yilmaz et al, unpublished data, 2020). Furthermore, older adults display more competence in recalling health information in an audiovisual format than a text-only format [20]. Hence, the overall aim of this study was to systematically develop, implement, and conduct a pilot evaluation of a web-based oncological module that can be integrated with the Health Communicator to stimulate patient

participation among older Turkish-Dutch and Moroccan-Dutch patients with cancer.

## Methods

The Spiral Technology Action Research (STAR) [21] model was used as the guiding framework for developing the oncological module. This model encourages several evaluation cycles among stakeholders, enabling continuous improvement of the intervention until it is finalized. To establish a high degree of relevance for target users, the model incorporates 5 guiding cycles that engage relevant community members in developing the intervention. These 5 cycles, *listen*, *plan*, *do*, *study*, and *act*, continuously weave technological and community development together [21]. In the following sections, we describe the steps taken in each phase with a special emphasis on the study phase and its results. [Table 1](#) provides an overview of the aims and actions during each phase.

**Table 1.** Overview of the phases, their aims, actions, and publications.

Phase	Aim	Actions	Publication
Listen	Target group analysis	<ul style="list-style-type: none"> <li>Conduct a literature review to identify (older) ethnic minority cancer patients' information and participation preferences and needs</li> <li>Conduct a qualitative study to identify unfulfilled instrumental and affective needs of older Turkish-Dutch and Moroccan-Dutch patients with cancer, and the barriers perceived by health care professionals (ie, general practitioners and oncology nurses) to fulfil patients' needs</li> </ul>	<ul style="list-style-type: none"> <li>Review [22]</li> <li>Submitted; paper under review (Yilmaz et al, unpublished data, 2020)</li> </ul>
Plan	Methodology development	<ul style="list-style-type: none"> <li>Develop question prompt lists and pilot testing them among patients</li> <li>Developing narrative patient education videos and pilot testing them among both patients and health care professionals (ie, general practitioners and oncology nurses)</li> </ul>	<ul style="list-style-type: none"> <li>Results shortly described in the Methods section of this paper</li> </ul>
Do	Development of the oncological module prototype	<ul style="list-style-type: none"> <li>Finalize the 6 question prompt lists and 5 narrative patient education videos by revising them based on feedback from patients and health care professionals (ie, general practitioners and oncology nurses) from the <i>plan</i> phase</li> </ul>	<ul style="list-style-type: none"> <li>No publication</li> </ul>
Study	Pilot evaluation of the oncological module	<ul style="list-style-type: none"> <li>Conduct a qualitative study to pilot-evaluate the oncological module prototype among patients and health care professionals (ie, general practitioners and oncology nurses)</li> </ul>	<ul style="list-style-type: none"> <li>Findings described in the Results section of this paper</li> </ul>
Act	Creating a dissemination plan	<ul style="list-style-type: none"> <li>Create and disseminate a stand-alone module: <i>the Conversation Starter</i></li> </ul>	<ul style="list-style-type: none"> <li>No publication</li> </ul>

### Listen: Target Group Analysis

To identify existing findings on the needs of older migrant patients with cancer, a systematic literature review was conducted. The results of the literature review revealed that most non-Western ethnic minority patients with cancer and survivors have a high preference and need for information and shared or active participation. However, no information was available regarding the preferences and needs of our target population, Moroccan-Dutch and Turkish-Dutch older patients with cancer [22].

Given the limited findings in the literature regarding the needs of older Turkish-Dutch and Moroccan-Dutch patients with cancer, further in-depth interviews were conducted to determine the topics that need to be addressed within the oncological module. A total of 19 interviews were carried out with

Turkish-Dutch and Moroccan-Dutch patients with cancer and cancer survivors. Interviews were conducted by bilingual interviewers in patients' first language (ie, Turkish, Moroccan Arabic, or Berber dialects). Participants were recruited by reaching out to prominent figures within Turkish and Moroccan communities that work in the health sector and via snowball sampling. The results revealed unmet instrumental needs concerning the treatment of cancer and the health care system and unmet affective needs concerning psychosocial support and affective doctor-patient communication. Acceptance of the Health Communicator, which was studied based on concepts of the Technology Acceptance Model (TAM), revealed that patients thought that the Health Communicator would be useful to fulfill their unfulfilled needs, but ultimately did not intend to use the tool (Yilmaz et al, unpublished data, 2020).

Following patient interviews, 2 focus groups with general practitioners (GPs) and oncology nurses were conducted. Health care professionals were asked to reflect on the findings of the patient interviews, share their own experiences, and indicate the type of information they would like to receive from their patients to better fulfill their unmet needs. Although health care professionals acknowledged the most unmet instrumental and affective needs of patients that emerged from the interviews, they appeared not to be aware of patients' need for psychosocial support or misunderstandings surrounding clinical trial requests. Health care professionals also needed more information about patients' instrumental needs and the role of family members in the treatment process (Yilmaz et al, unpublished data, 2020). Overall, the patient interviews and focus groups with GPs and oncology nurses provided valuable insights into the content creation phases.

### Plan: Methodology Development

On the basis of the results of the target group analysis, 2 types of content, namely QPLs and narrative patient education videos, were created for the oncological module. QPLs are essentially structured question lists that aim to make it easier for patients to ask questions to their providers during consultations [13]. Typically, patients can look at these premade lists before the consultations and select the questions they would like to ask their health care providers during consultations. Selecting questions from a premade list is easier than formulating the questions on their own for the patients. In addition, selecting them before consultations decreases the risk of forgetting to ask them during consultations because of time pressure or other distractions. Narrative patient education videos that depict patient stories are increasingly used in health communication as they are shown to enhance learning, recall of information, and intentions to stimulate healthy behavior and attitudes [23]. They are especially suitable for relaying cancer-related information as they can transport the patient into the story and make it easier for them to identify with the positive role models in the videos [24]. In the methodological development phase, several pilot tests with patients and professionals were conducted for the QPLs and videos before moving on to the third phase where the prototype was built.

### Developing QPLs

On the basis of the emergent instrumental and affective needs of patients, 5 QPLs were developed. The content of the QPLs was not disease-specific (ie, not about breast cancer or other cancer types) but rather included topics relevant to general oncological care that addresses the issues that older migrant patients with cancer face during their illness, namely (1) patient rights, (2) lifestyle and access to health care services, (3) treatment, (4) psychosocial support, and (5) clinical trials. Each QPL consisted of 4 or 5 simple questions aimed at stimulating patients to ask more questions to their health care professionals and have a more active role during consultations (eg, "Can I discuss my problems in my mother tongue with someone sharing my culture?" and "Can I ask for a second opinion?"). In addition, to address health care professionals' instrumental needs, a sixth QPL was developed, which enabled them to learn more about patients' instrumental and decision-making preferences and

health behaviors (eg, "Are you using any medication bought from another country?" and "Who do you want, next to your doctor, to help you make healthcare decisions for you?"). Additional illustrations representing each question were developed to assist patients with low literacy.

All QPLs and accompanying illustrations were pilot-tested during in-depth interviews conducted with 11 older migrant patients with cancer and survivors (8 Moroccan and 3 Turkish; mean age 61.50, SD 9.36 years). Participants were recruited by reaching out to prominent figures within Turkish and Moroccan communities that work in the health sector and via snowball sampling.

Interviews were conducted by bilingual interviewers in the patients' first language (ie, Turkish, Moroccan Arabic, or Berber dialects). Patients evaluated each question and illustration for ease of understanding of the content and language (eg, "Do you understand this question?" and "Do you think this picture is a clear illustration of the question?"). Results showed that patients found the QPL questions easy to understand, whereas some illustrations were found to be too abstract. Patients had specific recommendations for word choices and requested increasing concreteness and familiarity of illustrations, such as adding headscarves to some of the female figures and removing abstract symbols. On the basis of their recommendations, revisions were made and incorporated in the *do* phase of our intervention development.

### Developing Videos

We created 5 scripts for each video. Similar to QPLs, the content of the videos addressed general oncological issues that older migrant patients with cancer typically face during their illness, namely (1) patient rights and access to health care services, (2) doctor-patient communication, (3a and 3b) psychosocial support, and (4) clinical trials. We created 2 separate scripts featuring a male and female patient as the main characters for the psychosocial support video. This was done to enhance identification with characters for both male and female patients. Each script was related to the experiences of an older migrant patient who survived cancer from the patients' point of view. The scripts incorporated actual experiences and specific language used by patients during interviews in the target group analysis phase as much as possible. All scripts were prepared in Dutch for health care providers and in Turkish, Moroccan Arabic, and Berber dialects for patients.

The scripts were pilot-tested with 8 older migrant patients with cancer (Turkish,  $n=3$ ; Moroccan,  $n=5$ ; mean age 63.75, SD 6.39 years) during individual in-depth interviews. Participants were recruited by reaching out to prominent figures within Turkish and Moroccan communities that work in the health sector and via snowball sampling. Interviews were conducted by bilingual interviewers in patients' first language (ie, Turkish, Moroccan Arabic, or Berber dialects). Patients evaluated each script on ease of understanding ("Can you understand everything said in the video/happening in the video easily?"), familiarity ("Does this story sound familiar to you?"), usefulness ("Do you find the information provided in the story useful?"), believability ("Do you find the information believable?"), emotions induced ("Does this story induce any emotion for you and if yes what

type of emotions?”), and level of identification with the main characters (“Can you put yourself in the shoes of the character in the story?”). Patients found the scripts easy to understand and reported very high familiarity, believability, and identification with the characters. They stated that they could see themselves in these stories and understand the emotions shared by the characters, and they also found them to be very useful for other patients.

We also tested the scripts in 2 separate focus group meetings with GPs ( $n=6$  [2 women and 4 men]; mean age 45.17, SD 11.89 years) and oncology nurses ( $n=5$ , all women; mean age 49.60, SD 12.16 years). The health care professionals evaluated the scripts on accuracy (“Is all the information provided in the script correct?”), the usefulness of the provided information (“How useful do you find the information in the script?”), and their willingness to share videos once they are available (“Would you share these videos with your patients in the future?”). The results showed that health care professionals found the scripts to be accurate and useful, and they expressed their intentions to use them in the future.

### **Do: Developing the Oncological Module Prototype**

On the basis of the results of the pilot tests with patients and health care professionals, we created a prototype for the oncological module. The prototype included 6 QPLs tested in the previous phase. We used voice actors and added audio support to the QPLs in Turkish, Moroccan Arabic, and Berber (Taraftit dialect), enabling (illiterate) patients to listen to the QPLs in their mother tongue.

Similarly, feedback on the scripts was incorporated, and 5 short videos (1.5–3.5 min long) featuring Turkish-Dutch and Moroccan-Dutch actors were filmed. Once again, voice actors narrated these stories in Turkish, Moroccan Arabic, and Berber (Taraftit). Scripts were largely based on patients’ answers during the interviews in the target group analysis. The first video addressed patients’ instrumental needs about patient rights (eg, right to informed consent) and access to health care services

(eg, dietitian, home care, psychological support). The second video included suggestions to improve GP-patient communication by encouraging patients to prepare before consultations, ask more questions, make use of interpreters, and inform the doctor about their affective needs. The third and fourth videos aimed to provide psychosocial support to patients. The videos acknowledged the negative emotions experienced by patients with cancer and tried to counter them by giving a positive but also realistic message of hope. Topics about self-care, such as the importance of good diet, social contacts, psychological and spiritual support, were incorporated into these videos. The final video aimed to clarify the misunderstandings surrounding clinical trial requests; the video emphasized that a request to join a clinical trial does not mean that the patient has no hope of treatment and that patients can take time to decide on joining these trials, can refuse to join without risking their relationship with their doctors, and can quit if they have accepted to join previously. Furthermore, information regarding the general aim of these trials was provided.

### **Study: Pilot Evaluation of the Oncological Module**

#### **Sample**

The oncological module was pilot-evaluated in practice among 27 Turkish-Dutch and Moroccan-Dutch older patients with cancer aged 50 years and older and cancer survivors (Turkish:  $n=15$ , 9 women, mean age 63.47, SD 2.59 years; Moroccan:  $n=12$ , mean age 63.33 years, SD 2.70 years) and their health care professionals (GPs and oncology nurses:  $n=12$ , mean age 53.50 years, SD 13.34 years; see [Tables 2](#) and [3](#) for sample characteristics). Dutch language proficiency was self-reported by patients on a 4-point scale (1=poor, 2=mediocre, 3=reasonable, and 4=good). Patients were recruited by first targeting their health care professionals; on the basis of snowball sampling, with the help of health care professionals who participated in the earlier phases of the study, we reached other health care professionals, who then invited their patients to participate in the study.

**Table 2.** Background characteristics of patients.

Demographics	Turkish (n=15)	Moroccan (n=12)	Total (n=27)
Age (years), mean (SD)	63.47 (2.59)	63.33 (2.70)	63.41 (9.55)
Years residing in the Netherlands, mean (SD)	39.40 (3.21)	34.09 (3.54)	37.15 (12.19)
Dutch language proficiency, mean (SD)	1.89 (0.30)	2.21 (0.29)	2.02 (1.08)
<b>Sex, n (%)</b>			
Female	9 (60)	9 (75)	18 (67)
Male	6 (40)	3 (25)	9 (33)
<b>Education level, n (%)</b>			
No schooling	4 (26)	6 (50)	10 (36)
Primary school in Turkey or Morocco	7 (47)	5 (40)	12 (44)
Secondary school in Turkey or Morocco	1 (7)	— <sup>a</sup>	1 (4)
Primary school in the Netherlands	—	1 (10)	1 (4)
Secondary school in the Netherlands	2 (13)	—	2 (8)
Higher education in the Netherlands	1 (7)	—	1 (4)

<sup>a</sup>No participants belonging to the category.

**Table 3.** Background characteristics of health care professionals.

Demographic	General practitioners (n=10)	Oncology nurses (n=2)	Total (n=12)
Age (years), mean (SD)	57.50 (10.50)	33.50 (2.12)	53.50 (13.34)
Work experience in years, mean (SD)	23.44 (10.37)	6.00 (2.82)	20.27 (11.67)
<b>Sex, n (%)</b>			
Female	5 (50)	2 (100)	7 (58)
Male	5 (50)	— <sup>a</sup>	5 (42)
<b>Number of older patients with cancer aged ≥50 years with a Turkish or Moroccan background treated in the previous 2 years</b>			
1-2	1 (10)	—	1 (8)
2-4	4 (40)	—	4 (30)
5-10	5 (50)	1 (50)	6 (50)
≥10	—	1 (50)	1 (8)

<sup>a</sup>No participants belonging to the category.

## Procedure

Before the beginning of the pilot testing with patients, all participating health care professionals received an hour-long individual training on how to use the oncological module. Patients evaluated the QPL content and use before their consultations in GP practices or hospitals by means of surveys that were verbally administered to patients in their mother tongue by bilingual trained interviewers. Interviewers presented patients with QPL themes in the oncological module and asked them to choose one or more QPLs that they would like to fill out. After choosing a theme (eg, patient rights), patients saw all available questions within that QPL (eg, “Can I record consultations?”) and selected the ones they wanted to discuss with their professionals during the consultation (Figure 1). Before the consultation started, patients evaluated their experience with the QPLs that they chose. Health care

professionals rated the usefulness and ease of use of QPLs at the end of the pilot study.

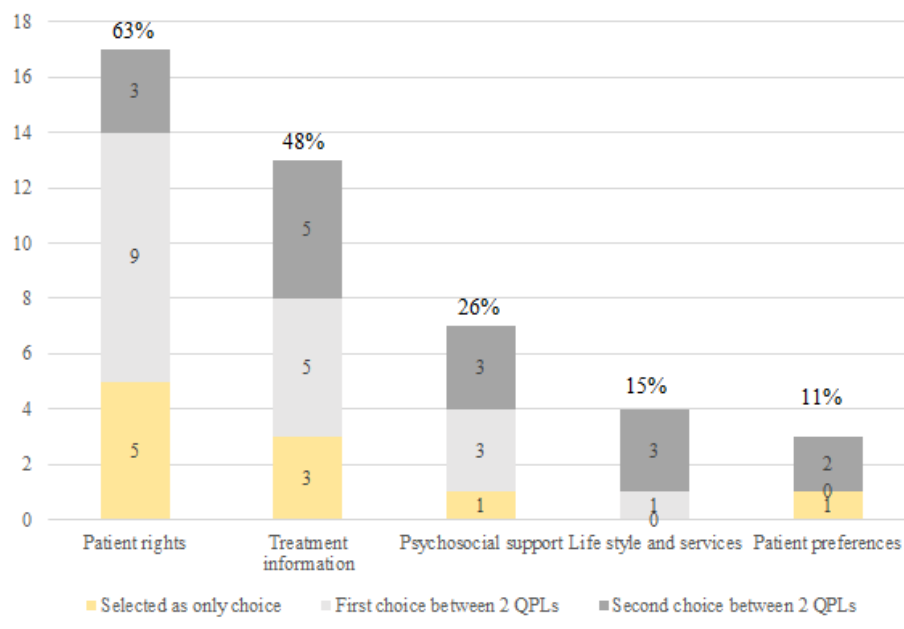
Following the QPL selection, patients and professionals consulted the same GP practice or hospital; 18 patients gave consent to have their consultations audio-taped. Patients who did not consent felt that the topics were too private and did not want anyone else to hear them, although their anonymity and confidentiality were guaranteed. After the consultations, patients completed a survey again, measuring their evaluation of the consultation.

After the consultations, the topics of all available narrative videos were briefly described to patients. Patients selected the videos they wanted to watch (Figure 2). Most of the patients preferred to receive videos on their smartphones. When this was not possible, patients requested that the videos were sent to patients’ family members’ phones. After patients watched the videos, phone interviews were conducted within a week, on

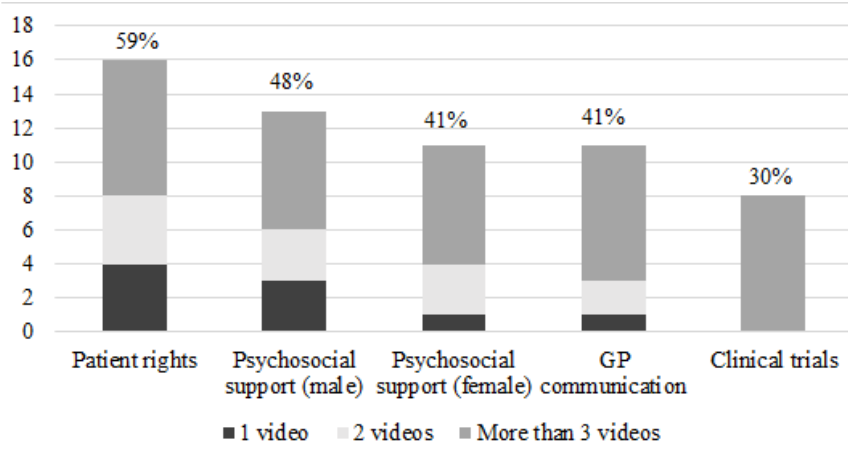
average, to assess how they perceived the videos. During the same phone interviews, patients were asked to rate the overall usefulness of the oncological module for improving their communication with their health care providers (ie, “Overall, this tool is useful to improve my communication with my

healthcare provider”). Health care professionals also responded to a similar question during the survey, which they filled at the end of their participation (ie, “The oncological module is useful in bridging communication barriers between migrant patients and their providers”).

**Figure 1.** Breakdown of question prompt lists selected by the patients in the evaluation study. Percentages reflect the percentage of participants that selected that question prompt list. Among the 27 participants, 10 selected 1 and 17 selected 2 question prompt lists. QPLs: question prompt lists.



**Figure 2.** Videos selected by the patients. Percentages reflect the percentage of participants that selected that video. 33% of patients selected 1 video, 22% selected 2 videos, 4% selected 3 videos, 19% selected 4 videos, and 11% selected 5 videos, amounting to an average of 2.27 (SD 1.59) videos requested per patient. GP: general practitioner.



## Measures

All survey items were measured using 5-point Likert scales (eg, 1=completely disagree, 5=completely agree; 1=very difficult, 5=very easy; and 1=very unsatisfied, 5=very satisfied).

### Patients' Evaluation of QPLs

After making their selection, patients assessed the ease of using the tool (both with and without assistance), ease of understanding QPL content and accompanying graphic images, and finally how convenient it was to choose QPLs before a consultation.

### Patients' Evaluations After Consultation

Patients' general satisfaction with the consultation was measured with 1 item (ie, “Overall, how satisfied are you with your consultation?”). Patients' perceptions of the informativeness and supportiveness of the health care professional were measured by 3 and 4 items, respectively, based on Street et al [25] (informativeness Cronbach  $\alpha=.84$ ; supportiveness Cronbach  $\alpha=.95$ ). Patients' attitudes toward the oncological module were measured with 2 items (using this tool “...made communicating with my doctor or nurse easier than usual” and “...had a positive influence on my consultation with my doctor or nurse”; Cronbach  $\alpha=.87$ ) and intention to use it again was measured



with 1 item (“I would like to use this tool again in the future”) [25].

Patients’ perceived participation during the consultation was measured using 5 items from Street et al [25] adaptation of the Lerman et al [26] perceived involvement in care scale (PICS; eg, “I asked my healthcare professional to explain the discussed topics in detail”; Cronbach  $\alpha=.82$ ). An additional question was used to measure the level of participation in comparison with previous consultations (ie, “I felt I could participate more than usual in this consultation”).

### Patient Participation

Patient participation was assessed with a self-developed codebook containing 2 measures: (1) relative talk for each person (ie, patient, health care professional, interpreter or companion) and (2) number and type of questions asked by patients, professionals, and interpreters or companion. In 11 of 18 consultations, an informal interpreter and/or companion was present; 8 consultations were held by nurses, 10 by GPs. Questions were divided according to the QPLs, except for clinical trials because this QPL could not be chosen because it was deemed irrelevant for GPs (ie, patient rights, treatment-related information, psychosocial support, lifestyle and access to health care services, patient information). In addition, a *miscellaneous* category was added to questions unrelated to these categories (eg, social questions, checking understanding, etc). The number of questions was counted per speech turn. If more than one question per turn was asked to address the same topic, this was counted as 1 question. Rhetorical questions were not addressed. Interpreters or companions’ questions that were a translation of a patient’s or professional’s questions were not counted.

### Patients’ Evaluation of the Narrative Videos

During the phone interviews, patients rated how easy it was for them to access (“How easy was it for you to access the videos”) and understand the videos (“The video was easy to understand”) and the usefulness (“The video was useful to improve my understanding of the topic about...”) and informativeness (“The video was very informative concerning the topic about...”) of each video that they had seen. The level of identification with the main characters in the narrative videos was measured with 4 items (Cronbach  $\alpha=.81$ ); 2 items were based on Murphy et al [24] and measured similarity and liking of the main characters. The other 2 items were based on the Cohen [27] identification scale and measured empathy for emotions experienced by the main characters. Finally, the patient’s intention to watch similar videos in the future was measured (ie, “In the future, would you like to watch videos that are similar to the ones that you watched?”) and whether they mentioned the videos to others (eg, friends, family members).

### Health Care Professionals’ Evaluation

After each consultation, health care professionals evaluated the participation of the patients during the consultation. To measure this, the same questions that patients responded to when reporting their participation were used. Specifically, these were the 5 items from the PICS (eg, “The patient asked me to explain the discussed topics in detail”; Cronbach  $\alpha=.74$ ) and the

additional question about comparison of the participation level (ie, “The patient participated more than usual in this consultation”). During the final survey study, health care professionals also evaluated the QPLs and narrative videos on perceived usefulness, ease of sharing with patients, and their intentions to use in the future.

### Analysis

Correlations between variables were calculated by running bivariate correlations using Pearson  $r$  as the correlation coefficient. The relative talk was measured by counting all words per person and calculating the word ratio. Question-asking was analyzed by means of descriptive analyses (mean and SD). Differences in the amount of question-asking were assessed using paired-samples  $t$  tests. The relationship between the first QPL topic choice by the patients and the number of questions asked by patients and health care professionals about that topic during consultations were assessed by simple linear regression analyses. Interrater reliability was calculated for 5 of 18 transcripts (27.7%; two-way mixed-effects model) and was 0.99 for professionals’ questions, 1.0 for patients’ questions, and 0.99 for interpreter or companions’ questions.

### Act

After the *study* phase was completed, a dissemination plan was devised in collaboration with the Dutch expertise center on migrant health care, Pharos, to distribute the oncological module to target audiences. First, to clearly communicate the goal of the oncological module, it was named the *Conversation Starter*. Next, the key organizations and actors that could help us reach health care professionals and older migrant patients with cancer were identified. Different newsletters highlighting the relevant parts of the Conversation Starter were prepared for these different audiences. Finally, a web version of the Conversation Starter that can be freely accessed without the Health Communicator application was created [28].

## Results

### Evaluation of QPL Content and Use

#### Patients

Overall, patients perceived the questions in the QPLs as easy to understand (mean 4.30, SD 0.65) and useful (mean 3.96, SD 0.72). Graphic images accompanying the QPLs were perceived as easy to understand (mean 3.75, SD 1.14). Patients found the usefulness of the images in aiding the understanding of the questions as neutral (mean 3.23, SD 1.27). Overall, participants found it relatively easy to use the QPL function in the oncological module (mean 3.54, SD 1.07). They had low confidence, although still neutral, about their ability to use the tool easily without any assistance (mean 3.19, SD 1.44). There was a strong negative correlation between the ease of using the tool and patient age. It was harder for older patients to use the QPLs ( $r=-0.48$ ;  $P=.01$ ), and they were also less confident in their ability to use the tool without any assistance ( $r=-0.67$ ;  $P<.001$ ). Approximately 70% of the patients reported that it was convenient for them to select the QPLs before their consultations (mean 3.85, SD 1.35). Similarly, younger age was

related to reporting more convenience in using the QPL function before consultations ( $r=-0.56$ ;  $P=.003$ ).

### Professionals

Health care professionals rated the QPLs as rather useful (mean 3.67, SD 1.07) and overall somewhat easy to use (mean 3.44, SD 0.96).

### Patient Participation

The mean consultation length was 14.09 min (SD 7.60; range 3.37–35.06 min). Professionals spoke, on average, most words (48.15%; range 166–2481), followed by patients (33.5%; range 130–2544), and interpreters or companions (18.35%; range 38–1672). The mean number of questions asked per consultation was 24.55 (SD 17.21; range 11–89). Professionals asked significantly more questions (mean 15.72, SD 9.36; range 3–43) than both patients (mean 4.56, SD 5.45; range 0–23;  $t_{17}=6.61$ ;  $P<.001$ ) and interpreters or companions (mean 4.28, SD 6.44; range 0–23;  $t_{17}=4.81$ ;  $P<.001$ ). The most asked questions were about treatment-related information (mean 9.89, SD 15.87), followed by questions on miscellaneous topics (mean 9.28, SD 4.34), patient rights (mean 2.17, SD 2.04), psychosocial support (mean 1.67, SD 2.97), and lifestyle and access to health care services (mean 1.50, SD 2.75). The least asked questions were about patient information (mean 0.06, SD 0.24).

### Relationship Between QPL Choice and Participation During Consultations

#### Patients

Overall, no significant relationship was found between the first QPL selection and the number of questions asked by patients during the consultation about that topic (patient rights:  $R^2=0.17$ ,  $b^*=0.12$ ;  $t_{16}=0.68$ ,  $P=.51$ ; 95% CI  $-0.267$  to  $0.517$ ; treatment:  $R^2=0.05$ ,  $b^*=-0.01$ ;  $t_{16}=-0.20$ ,  $P=.84$ ; 95% CI  $-0.058$  to  $0.048$ ; and psychosocial support:  $R^2=0.15$ ,  $b^*=-0.11$ ;  $t_{16}=-0.60$ ,  $P=.56$ ; 95% CI  $-0.502$  to  $0.279$ ). As there were not enough patients who selected the QPLs about lifestyle and access to health care services and patient preferences as their first choice, it was not possible to run the analyses to test these relationships.

#### Professionals

Similarly, we tested the relationship between patients' first QPL choice and the number of questions posed by health care professionals about that topic. The number of questions asked by the health care professionals in a given topic was marginally significantly predicted by the first QPL topic selected by the patients for patient rights ( $R^2=0.42$ ,  $b^*=0.14$ ;  $t_{16}=1.85$ ,  $P=.08$ ; 95% CI  $-0.021$  to  $0.309$ ) but not for treatment ( $R^2=0.09$ ,  $b^*=-0.01$ ;  $t_{16}=-0.37$ ,  $P=.72$ ; 95% CI  $-0.040$  to  $0.028$ ) and psychosocial support ( $R^2=0.15$ ,  $b^*=0.02$ ;  $t_{16}=0.61$ ,  $P=.55$ ; 95% CI  $-0.060$  to  $0.108$ ).

### Evaluation of Consultations

#### Patients

Overall, patients reported being highly satisfied with their consultations (mean 4.31, SD 0.55). They perceived their

providers as informative (mean 4.17, SD 0.57) and supportive (mean 4.34, SD 0.59). They also reported that using the module positively impacted consultations (mean 3.90, SD 0.55). Patients perceived themselves to be rather active during the consultations (mean 3.75, SD 0.84) and thought they were somewhat more active in comparison with previous consultations (mean 3.75, SD 0.79). Most (70%) of the participants reported that they would like to use the tool in the future (mean 3.85, SD 1.25). The age of the participants had a strong negative correlation with their wish to use the tool in the future ( $r=0.45$ ;  $P=.02$ ).

#### Professionals

Professionals' evaluation of the patients' participation revealed similar, although slightly lower scores. Professionals found patients to be slightly more active in asking questions and expressing themselves (mean 3.42, SD 0.70) and somewhat more active than their usual levels of participation (mean 3.61, SD 1.06). Both patients' and providers' evaluations were positively correlated for each of these measures ( $r=0.58$ ,  $P=.002$  and  $r=0.64$ ,  $P<.001$ , respectively).

### Evaluation of Narrative Videos

#### Patients

Patients rated the videos as easy to understand (mean 4.30, SD 0.47), useful (mean 4.00, SD 0.62), and informative (mean 4.20, SD 0.57). Identification with the characters in the videos was high (mean 3.93, SD 0.54). Patients were somewhat motivated to watch similar videos in the future (mean 3.55, SD 0.85). Around 30% of the patients talked about videos with others (ie, told their friends and family members that they watched them).

#### Professionals

Professionals rated the videos as useful (mean 3.92, SD 1.08). They were neutral about the ease of sharing the videos with their patients (mean 3.25, SD 1.06) and showed some intentions to share them (mean 3.42, SD 1.08).

### Overall Evaluation of the Oncological Module

Overall, patients found the tool to be useful in improving their communication with their health care professionals (mean 4.27, SD 0.88). Health care professionals rated the usefulness of the oncological module less favorable than the patients (mean 3.42, SD 1.16).

## Discussion

### Principal Findings

This paper elaborates on the development of an oncological module aimed at increasing patient participation between older migrant patients with cancer and their health care professionals. Using the STAR model as our framework, the module was developed with continuous input from relevant community members, namely older migrant patients with cancer, oncology nurses, GPs, and researchers. This multifaceted contribution allowed us to develop a multilingual intervention that received positive evaluations from both migrant patients with cancer and professionals. Patients most often chose the QPL for patient rights. Both patients and health care professionals perceived the QPLs as somewhat easy to understand and useful. The

correlation between the ease of using the tool and patient age was negative. Health care professionals asked significantly more questions than patients, whereas patients reported that using the module impacted the consultations positively and thought they were more active compared with previous consultations. Health care professionals echoed that patients were more active than usual. However, there were no significant relations between patients' first choice of QPLs and the number of questions asked about that topic during consultations. Patients requested to see the patients' rights video most often, and overall rated the videos as easy to understand, useful, and informative. Most of the patients reported that they like to use the tool in the future.

Aligned with the study by Walczak et al [29], migrant patients with cancer in this study evaluated the QPLs as easy to understand and quite useful. This is likely to be the result of the fact that the QPLs served the function of breaking down all cancer-related information that existed in small segments. By doing so, large chunks of information were labeled, which helped migrant patients to identify their concerns and needs more easily, thus preventing the possibility of experiencing cognitive overload [29]. Nonetheless, the evaluation became less positive with an increase in patients' age. The older the patient, the less easy they found using the tool, which might partly be explained by a decrease in one's cognitive competence over time and the digital divide (ie, a disparity in using digital technologies between young and old migrants and nonmigrants) [30-32]. To be readily able to adopt eHealth tools, older patients first have to be in a physically good condition and, second, be cognitively competent enough to develop internet literacy [33]. On the basis of the results of a systematic review, a suggested solution to help (older) migrant patients to better understand how QPLs work would be through incorporating a training component for patients [34]. In our intervention, only health professionals were trained. However, this—and even more in-depth training—is not sufficiently useful for patients because language barriers professionals are limited in their ability to pass on their training to patients. Therefore, a training component specifically targeted for patients is needed not only to ensure that the consultation sessions would be effectively guided by the QPLs but also to offer guidance for older migrant patients who might need help using the tool owing to old age [35]. This component can perhaps be added after each consultation session and shared with family members of the patient so that older migrant patients and their relatives are exposed to the tool more frequently, which in turn helps to increase their familiarity with the tool and enable relatives to help their older family members in using it.

Results collected to examine patient participation during the consultation sessions revealed that professionals were more active than patients in terms of leading the conversation and asking questions. A recent study revealed that patients' passive participation could be owing to the knowledge and language barriers they face [36], also indicated by the lack of relation between the choice of QPLs and the type of questions asked. Although an interpreter or a companion was present during most of the consultation sessions, it is still possible that the patients had a lack of understanding about the subject matters raised by the professionals and/or did not have the ability to ask questions

about the topics they indicated to want to discuss by their QPL choice because of an unresolved language barrier during the consultations. As a result of these barriers, there is a possibility that patients possessed insufficient competence to formulate the exact question or response they would like to make, especially when it comes to disagreeing with a statement proposed by the professional, as this requires more effortful processing, compared with simply agreeing with a statement [36,37]. As such, they may experience a lack of confidence and choose to refrain from making their point, thus generating fewer questions and talk, leading to a relatively passive outcome in terms of patient participation. As indicated by Cegala and Post [38], a lack of response from the patients' end hinders the active adoption of a more patient-centered communication from the professionals' end. This shows that there is a need for patients to be more empowered to secure a consultation session with active patient participation [38].

Given that the results indicated that most questions raised were about treatment-related information, a possible solution to tackle patients' passiveness would be through distributing information related to their cancer diagnosis in their native languages along with the QPLs before each consultation session, if they had a preference for this information. This approach can enhance patients' knowledge of the subject, which, in turn, allows them to become more confident in expressing their views and opinions about the professionals' suggestions and thus facilitate patient-centered communication, resulting in active patient participation. In addition, working with professional interpreters instead of nonprofessional interpreters during consultations is recommended, as most patient rights topics discussed during the consultations were related to patients' need for a professional interpreter. This will lead to better translations and enhance patients' understanding of the conversation [39,40], enabling more active participation during consultations.

The assessment of migrant patients with cancer of their QPL-incorporated consultation sessions was fairly positive, with patients expressing high satisfaction and concluding that the QPLs helped to impact their consultation sessions positively. Professionals' overall evaluation of patients' participation was less satisfactory, possibly because patients did ask fewer questions and contributed less to the conversation than they might have expected as a result of the intervention. On the other hand, for patients, the QPLs were able to help them identify topics of interest in a more direct manner, and, in turn, this helped professionals to formulate a clearer picture of their unmet instrumental and/or affective needs and prepare the consultation sessions based on this feedback. Although only marginally significant, the positive relationship between patients' choice of the patient rights QPL and health care professionals' questions about patient rights seems to lend some substantiation for this positive effect of QPLs, helping both parties to bring up the topic of patient rights more immediately during their consultation, thus increasing patient satisfaction with the consultation sessions [34]. Corresponding with previous research, this indicates that using QPLs allowed for more effective communication between patients and professionals, as indicated by the perceived higher than usual patient



participation, helping to improve the consultation sessions and encourage active patient participation [34,41].

Finally, both patients and professionals evaluated educational videos positively. Patients reported fairly high levels of identification with the characters in the narrative videos and expressed moderate intentions to watch similar videos in the future. High levels of identification with the character in the videos might likely have helped migrant patients with the ease of processing information presented because of sharing a common background with the characters [42]. Ultimately, this leads to more informed and empowered patients and, in turn, results in positive attitudes toward videos and better quality in health communication and care. What was especially interesting here was that the results reflected the patients showing more interest in narrative videos that covered the theme of patient rights in comparison with the other themes. This preference was also reflected in the selection of themes in the QPLs. Corresponding to previous research, this shows that there is a possibility that migrant patients often feel that they are not taken seriously, and to a certain extent, even discriminated against by their doctors [5,43]. Nonetheless, the precise concerns migrant patients might have concerning their rights are still scarcely researched and, thus, pointing to the fact that more empirical evidence is needed to understand inadequacies in the health care system for older migrant patients. Overall, this warrants that future research is needed to determine the underlying motivations and reasons behind choosing this theme.

### Study Limitations and Future Research

Although older migrant patients with cancer showed some intentions to use the oncological module in the future, the results indicated that older patients had concerns and expressed little ease and confidence in their ability to use the tools on their own, especially regarding using the QPLs. Again, this shows that training patients remains a crucial component of such intervention to be effective, as this study did not sufficiently target the issue of adopting and adhering to the eHealth tool, which is especially important among older adults because of their low internet self-efficacy [44]. Previous research has indicated that just short-term training can lead to successful information and communication technology adoption and outcomes for ethnically diverse older people [45]. In addition, future studies could incorporate the extended TAM and senior technology acceptance model (STAM) into the developmental process [44,46] instead of using the original TAM that was used in this study. The extended TAM sheds light on several spearheads, such as the amount of text used, page organization, and the incorporation of offline support to increase the perceived ease of use of an eHealth tool among older adults [44], whereas the STAM also incorporates age-related health and ability characteristics, such as gerontechnology self-efficacy [46]. Taking these additional factors suggested by the extended TAM

and STAM into consideration, this might further increase the tool's overall user-friendliness and lead to higher levels of perceived ease of use, usefulness, and intention to use.

Furthermore, as there is a dearth of research assessing digital literacy skills among older migrant patients, future studies are needed to gain more insight into their (lack of) competence to use new technologies in everyday life [47] and what is needed to enhance their skills. The results of such studies will help future intervention developers to gain a deeper insight into understanding what features are indeed appropriate and easy to use. The combination of incorporating the extended TAM and measuring older patients' digital literacy might help to produce higher levels of adoption and adherence to QPLs, which ultimately increases patient participation during the consultation sessions.

Another limitation of this study is related to the developmental process of the prototype. The STAR model was designed fundamentally to promote healthy behaviors among young adolescents by encouraging them to actively participate in the e-tool developmental process [48]. This involves all stakeholders being present in the discussions and decisions taken during all developmental phases. Although we did include patients in as many phases as possible, this approach was deemed unfeasible in the *do* phase when the prototype was developed, given the fact that our target group is rather vulnerable because of their sickness and old age. As the inclusion of people who represent different areas of expertise could increase the effectiveness of brainstorming sessions largely [49], future studies should explore the possibility and feasibility of inviting ethnic minority patients who are not ill yet or migrant patients' close relatives to take part directly in this phase of the intervention development. This could help eHealth developers to obtain a fuller picture of the attributes that can be added to the module and help maximize the customization of the tool for the target user.

### Conclusions

We conclude that QPLs make the oncological module a beneficial tool to assist health care professionals in question-asking during consultations and to better fulfill the instrumental and affective needs of older migrant patients with cancer. In addition, the use of narratives to inform older and low-literate migrant patients with cancer about health-related topics in their mother tongue is a viable approach to increase the effectiveness of health care communication with this target group. However, given that older migrant patients are less able to use the QPLs on their own, health care professionals should also look into the feasibility of adding a training component to offer offline guidance in navigating the QPLs. Finally, the oncology module developed in this study is a promising tool for both patients and health care professionals.

### Acknowledgments

The authors would like to thank all the participating patients, health care providers, and research assistants who made this project possible. The authors would also like to thank the reviewers for their valuable contributions, which helped us to improve this paper.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

An infographic depicting the difference between the existing Health Communicator tool and the new oncological module developed in this study.

[PNG File , 3733 KB - [jmir\\_v22i10e21238\\_app1.png](#) ]

## References

1. Butow P, Bell M, Goldstein D, Sze M, Aldridge L, Abdo S, et al. Grappling with cultural differences; communication between oncologists and immigrant cancer patients with and without interpreters. *Patient Educ Couns* 2011 Sep;84(3):398-405. [doi: [10.1016/j.pec.2011.01.035](#)] [Medline: [21388772](#)]
2. Ahmed S, Lee S, Shommu N, Rumana N, Turin T. Experiences of communication barriers between physicians and immigrant patients: a systematic review and thematic synthesis. *Patient Experience Journal* 2017 Apr 24;4(1):122-140. [doi: [10.35680/2372-0247.1181](#)]
3. Jacobs EA, Diamond L. *Providing Health Care in the Context of Language Barriers: International Perspectives*. New York, USA: Multilingual Matters; 2017.
4. Akhavan S, Karlsen S. Practitioner and client explanations for disparities in health care use between migrant and non-migrant groups in Sweden: a qualitative study. *J Immigr Minor Health* 2013 Feb;15(1):188-197. [doi: [10.1007/s10903-012-9581-y](#)] [Medline: [22323124](#)]
5. Jangland E, Gunningberg L, Carlsson M. Patients' and relatives' complaints about encounters and communication in health care: evidence for quality improvement. *Patient Educ Couns* 2009 May;75(2):199-204. [doi: [10.1016/j.pec.2008.10.007](#)] [Medline: [19038522](#)]
6. Suurmond J, Seeleman C. Shared decision-making in an intercultural context. Barriers in the interaction between physicians and immigrant patients. *Patient Educ Couns* 2006 Feb;60(2):253-259. [doi: [10.1016/j.pec.2005.01.012](#)] [Medline: [16442467](#)]
7. KWF. *Allochtonen en kanker: sociaal-culturele en epidemiologische aspecten ethnic minorities and cancer: Socio-cultural and epidemiological aspects*. Amsterdam 2006:-.
8. Street RL, Millay B. Analyzing patient participation in medical encounters. *Health Commun* 2001 Jan;13(1):61-73. [doi: [10.1207/s15327027hc1301\\_06](#)] [Medline: [11370924](#)]
9. de Graaff FM, Francke AL, van den Muijsenbergh ME, van der Geest S. Talking in triads: communication with Turkish and Moroccan immigrants in the palliative phase of cancer. *J Clin Nurs* 2012 Nov;21(21-22):3143-3152. [doi: [10.1111/j.1365-2702.2012.04289.x](#)] [Medline: [23083389](#)]
10. Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: a concept analysis based on a literature review. *Patient Educ Couns* 2016 Dec;99(12):1923-1939. [doi: [10.1016/j.pec.2016.07.026](#)] [Medline: [27450481](#)]
11. Haywood K, Marshall S, Fitzpatrick R. Patient participation in the consultation process: a structured review of intervention strategies. *Patient Educ Couns* 2006 Oct;63(1-2):12-23. [doi: [10.1016/j.pec.2005.10.005](#)] [Medline: [16406464](#)]
12. Epstein RM, Duberstein PR, Fenton JJ, Fiscella K, Hoerger M, Tancredi DJ, et al. Effect of a patient-centered communication intervention on oncologist-patient communication, quality of life, and health care utilization in advanced cancer: the voice randomized clinical trial. *JAMA Oncol* 2017 Jan 1;3(1):92-100 [FREE Full text] [doi: [10.1001/jamaoncol.2016.4373](#)] [Medline: [27612178](#)]
13. Brandes K, Linn AJ, Butow PN, van Weert JC. The characteristics and effectiveness of question prompt list interventions in oncology: a systematic review of the literature. *Psychooncology* 2015 Mar;24(3):245-252. [doi: [10.1002/pon.3637](#)] [Medline: [25082386](#)]
14. Fukui S, Kamiya M, Koike M, Kugaya A, Okamura H, Nakanishi T, et al. Applicability of a Western-developed psychosocial group intervention for Japanese patients with primary breast cancer. *Psychooncology* 2000;9(2):169-177. [doi: [10.1002/\(sici\)1099-1611\(200003/04\)9:2<169::aid-pon441>3.0.co;2-v](#)] [Medline: [10767755](#)]
15. Ferguson WJ, Candib L. Culture, language, and the doctor-patient relationship. *Fam Med* 2002 May;34(5):353-361. [Medline: [12038717](#)]
16. Paternotte E, van Dulmen S, van der Lee N, Scherpbier AJ, Scheele F. Factors influencing intercultural doctor-patient communication: a realist review. *Patient Educ Couns* 2015 Apr;98(4):420-445. [doi: [10.1016/j.pec.2014.11.018](#)] [Medline: [25535014](#)]
17. The World Factbook Europe: The Netherlands. Central Intelligence Agency. URL: <https://www.cia.gov/library/publications/the-world-factbook/geos/nl.html> [accessed 2020-02-16]
18. Relatively High Population Growth Due to Migration. CBS. 2017. URL: <https://www.cbs.nl/en-gb/news/2017/05/relatively-high-population-growth-due-to-migration> [accessed 2020-02-16]
19. Schinkel S, Van Weert JC, Kester JA, Smit EG, Schouten BC. Does media use result in more active communicators? Differences between native dutch and turkish-dutch patients in information-seeking behavior and participation during



- consultations with general practitioners. *J Health Commun* 2015 Aug;20(8):910-919. [doi: [10.1080/10810730.2015.1018600](https://doi.org/10.1080/10810730.2015.1018600)] [Medline: [26073918](https://pubmed.ncbi.nlm.nih.gov/26073918/)]
20. Bol N, van Weert JC, de Haes HC, Loos EF, Smets EM. The effect of modality and narration style on recall of online health information: results from a Web-based experiment. *J Med Internet Res* 2015 Apr 24;17(4):e104 [FREE Full text] [doi: [10.2196/jmir.4164](https://doi.org/10.2196/jmir.4164)] [Medline: [25910416](https://pubmed.ncbi.nlm.nih.gov/25910416/)]
  21. Skinner HA, Maley O, Norman CD. Developing internet-based ehealth promotion programs: the spiral technology action research (STAR) model. *Health Promot Pract* 2006 Oct;7(4):406-417. [doi: [10.1177/1524839905278889](https://doi.org/10.1177/1524839905278889)] [Medline: [16840770](https://pubmed.ncbi.nlm.nih.gov/16840770/)]
  22. Yilmaz NG, Schouten BC, Schinkel S, van Weert JC. Information and participation preferences and needs of non-Western ethnic minority cancer patients and survivors: a systematic review of the literature. *Patient Educ Couns* 2019 Apr;102(4):631-650. [doi: [10.1016/j.pec.2018.11.018](https://doi.org/10.1016/j.pec.2018.11.018)] [Medline: [30594322](https://pubmed.ncbi.nlm.nih.gov/30594322/)]
  23. de Graaf A, Sanders J. Characteristics of narrative interventions and health effects: a review of the content, form, and context of narratives in health-related narrative persuasion research. *Rev Commun Res* 2016;4:88-131.
  24. Murphy ST, Frank LB, Chatterjee JS, Baezconde-Garbanati L. Narrative versus non-narrative: the role of identification, transportation and emotion in reducing health disparities. *J Commun* 2013 Feb;63(1):- [FREE Full text] [doi: [10.1111/jcom.12007](https://doi.org/10.1111/jcom.12007)] [Medline: [24347679](https://pubmed.ncbi.nlm.nih.gov/24347679/)]
  25. Street RL, Gordon H, Haidet P. Physicians' communication and perceptions of patients: is it how they look, how they talk, or is it just the doctor? *Soc Sci Med* 2007 Aug;65(3):586-598 [FREE Full text] [doi: [10.1016/j.socscimed.2007.03.036](https://doi.org/10.1016/j.socscimed.2007.03.036)] [Medline: [17462801](https://pubmed.ncbi.nlm.nih.gov/17462801/)]
  26. Lerman CE, Brody DS, Caputo GC, Smith DG, Lazaro CG, Wolfson HG. Patients' perceived involvement in care scale: relationship to attitudes about illness and medical care. *J Gen Intern Med* 1990;5(1):29-33. [doi: [10.1007/BF02602306](https://doi.org/10.1007/BF02602306)] [Medline: [2299426](https://pubmed.ncbi.nlm.nih.gov/2299426/)]
  27. Cohen J. Defining identification: a theoretical look at the identification of audiences with media characters. *Mass Commun Soc* 2001 Aug;4(3):245-264. [doi: [10.1207/s15327825mcs0403\\_01](https://doi.org/10.1207/s15327825mcs0403_01)]
  28. Amsterdam Center for Health Communication. URL: <https://www.healthcommunication.nl/blog/category/gesprekstarter/> [accessed 2020-09-30]
  29. Walczak A, Mazer B, Butow PN, Tattersall MH, Clayton JM, Davidson PM, et al. A question prompt list for patients with advanced cancer in the final year of life: development and cross-cultural evaluation. *Palliat Med* 2013 Sep;27(8):779-788. [doi: [10.1177/0269216313483659](https://doi.org/10.1177/0269216313483659)] [Medline: [23630055](https://pubmed.ncbi.nlm.nih.gov/23630055/)]
  30. Bolle S, Romijn G, Smets EM, Loos EF, Kunnean M, van Weert JC. Older cancer patients' user experiences with web-based health information tools: a think-aloud study. *J Med Internet Res* 2016 Jul 25;18(7):e208 [FREE Full text] [doi: [10.2196/jmir.5618](https://doi.org/10.2196/jmir.5618)] [Medline: [27457709](https://pubmed.ncbi.nlm.nih.gov/27457709/)]
  31. Nikou S, Agahari W, Keijzer-Broers W, de Reuver M. Digital healthcare technology adoption by elderly people: a capability approach model. *Telematics Info* 2020 Oct;53:101315. [doi: [10.1016/j.tele.2019.101315](https://doi.org/10.1016/j.tele.2019.101315)]
  32. Walker DM, Hefner JL, Fareed N, Huerta TR, McAlearney AS. Exploring the digital divide: age and race disparities in use of an inpatient portal. *Telemed J E Health* 2020 May;26(5):603-613. [doi: [10.1089/tmj.2019.0065](https://doi.org/10.1089/tmj.2019.0065)] [Medline: [31313977](https://pubmed.ncbi.nlm.nih.gov/31313977/)]
  33. Nguyen MH, Smets EM, Bol N, Loos EF, Van Weert JC. How tailoring the mode of information presentation influences younger and older adults' satisfaction with health websites. *J Health Commun* 2018;23(2):170-180. [doi: [10.1080/10810730.2017.1421729](https://doi.org/10.1080/10810730.2017.1421729)] [Medline: [29345531](https://pubmed.ncbi.nlm.nih.gov/29345531/)]
  34. Brandes K, Butow PN, Tattersall MH, Clayton JM, Davidson PM, Young J, et al. Advanced cancer patients' and caregivers' use of a question prompt list. *Patient Educ Couns* 2014 Oct;97(1):30-37. [doi: [10.1016/j.pec.2014.06.010](https://doi.org/10.1016/j.pec.2014.06.010)] [Medline: [25023487](https://pubmed.ncbi.nlm.nih.gov/25023487/)]
  35. Rodenbach RA, Brandes K, Fiscella K, Kravitz RL, Butow PN, Walczak A, et al. Promoting end-of-life discussions in advanced cancer: effects of patient coaching and question prompt lists. *J Clin Oncol* 2017 Mar 10;35(8):842-851. [doi: [10.1200/jco.2016.68.5651](https://doi.org/10.1200/jco.2016.68.5651)] [Medline: [28135140](https://pubmed.ncbi.nlm.nih.gov/28135140/)]
  36. Landmark AM, Svennevig J, Gerwing J, Gulbrandsen P. Patient involvement and language barriers: problems of agreement or understanding? *Patient Educ Couns* 2017 Jun;100(6):1092-1102. [doi: [10.1016/j.pec.2016.12.006](https://doi.org/10.1016/j.pec.2016.12.006)] [Medline: [28065435](https://pubmed.ncbi.nlm.nih.gov/28065435/)]
  37. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. *Health Educ Res* 2008 Jun;23(3):454-466 [FREE Full text] [doi: [10.1093/her/cyn004](https://doi.org/10.1093/her/cyn004)] [Medline: [18349033](https://pubmed.ncbi.nlm.nih.gov/18349033/)]
  38. Cegala DJ, Post DM. The impact of patients' participation on physicians' patient-centered communication. *Patient Educ Couns* 2009 Nov;77(2):202-208. [doi: [10.1016/j.pec.2009.03.025](https://doi.org/10.1016/j.pec.2009.03.025)] [Medline: [19395225](https://pubmed.ncbi.nlm.nih.gov/19395225/)]
  39. Butow PN, Goldstein D, Bell ML, Sze M, Aldridge LJ, Abdo S, et al. Interpretation in consultations with immigrant patients with cancer: how accurate is it? *J Clin Oncol* 2011 Jul 10;29(20):2801-2807. [doi: [10.1200/JCO.2010.34.3335](https://doi.org/10.1200/JCO.2010.34.3335)] [Medline: [21670462](https://pubmed.ncbi.nlm.nih.gov/21670462/)]
  40. Karliner LS, Jacobs EA, Chen AH, Mutha S. Do professional interpreters improve clinical care for patients with limited English proficiency? A systematic review of the literature. *Health Serv Res* 2007 Apr;42(2):727-754 [FREE Full text] [doi: [10.1111/j.1475-6773.2006.00629.x](https://doi.org/10.1111/j.1475-6773.2006.00629.x)] [Medline: [17362215](https://pubmed.ncbi.nlm.nih.gov/17362215/)]

41. Noordman J, Drienaar JA, Henselmans I, Verboom J, Heijmans M, van Dulmen S. Patient participation during oncological encounters: barriers and need for supportive interventions experienced by elderly cancer patients. *Patient Educ Couns* 2017 Dec;100(12):2262-2268. [doi: [10.1016/j.pec.2017.06.009](https://doi.org/10.1016/j.pec.2017.06.009)] [Medline: [28619272](https://pubmed.ncbi.nlm.nih.gov/28619272/)]
42. Moyer-Gusé E. Toward a theory of entertainment persuasion explaining the persuasive effects of entertainment-education messages. *Commun Theory* 2008;18(3):- . [doi: [10.1111/j.1468-2885.2008.00328.x](https://doi.org/10.1111/j.1468-2885.2008.00328.x)]
43. Schinkel S, Schouten BC, Kerplick F, van den Putte B, van Weert JC. Perceptions of barriers to patient participation: are they due to language, culture, or discrimination? *Health Commun* 2019 Nov;34(12):1469-1481. [doi: [10.1080/10410236.2018.1500431](https://doi.org/10.1080/10410236.2018.1500431)] [Medline: [30040497](https://pubmed.ncbi.nlm.nih.gov/30040497/)]
44. Chung JE, Park N, Wang H, Fulk J, McLaughlin M. Age differences in perceptions of online community participation among non-users: an extension of the technology acceptance model. *Comput Hum Behav* 2010 Nov;26(6):1674-1684. [doi: [10.1016/j.chb.2010.06.016](https://doi.org/10.1016/j.chb.2010.06.016)]
45. Czaja SJ, Boot W, Charness N, Rogers W, Sharit J. Improving social support for older adults through technology: findings from the prism randomized controlled trial. *Gerontologist* 2018 May 8;58(3):467-477 [FREE Full text] [doi: [10.1093/geront/gnw249](https://doi.org/10.1093/geront/gnw249)] [Medline: [28201730](https://pubmed.ncbi.nlm.nih.gov/28201730/)]
46. Chen K, Chan AH. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). *Ergonomics* 2014;57(5):635-652. [doi: [10.1080/00140139.2014.895855](https://doi.org/10.1080/00140139.2014.895855)] [Medline: [24655221](https://pubmed.ncbi.nlm.nih.gov/24655221/)]
47. Schäffer B. The digital literacy of seniors. *Res Comp Int Educ* 2016 Jul 31;2(1):29-42. [doi: [10.2304/rcie.2007.2.1.29](https://doi.org/10.2304/rcie.2007.2.1.29)]
48. Buccieri K, Molleson G. Empowering homeless youth: building capacity through the development of mobile technology. *J Community Pract* 2015 May 18;23(2):238-254. [doi: [10.1080/10705422.2015.1027802](https://doi.org/10.1080/10705422.2015.1027802)]
49. Markman AB. *Open Innovation: Academic and Practical Perspectives on the Journey From Idea to Market*. Oxford, UK: Oxford University Press; 2016.

## Abbreviations

**GP:** general practitioner

**PICS:** perceived involvement in care scale

**QPL:** question prompt list

**STAM:** senior technology acceptance model

**STAR:** Spiral Technology Action Research

**TAM:** technology acceptance model

*Edited by G Eysenbach; submitted 08.06.20; peer-reviewed by J Suurmond, Z Su; comments to author 11.07.20; revised version received 05.09.20; accepted 07.09.20; published 26.10.20.*

*Please cite as:*

Sungur H, Yılmaz NG, Chan BMC, van den Muijsenbergh METC, van Weert JCM, Schouten BC

*Development and Evaluation of a Digital Intervention for Fulfilling the Needs of Older Migrant Patients With Cancer: User-Centered Design Approach*

*J Med Internet Res* 2020;22(10):e21238

URL: <http://www.jmir.org/2020/10/e21238/>

doi: [10.2196/21238](https://doi.org/10.2196/21238)

PMID: [33104008](https://pubmed.ncbi.nlm.nih.gov/33104008/)

©Hande Sungur, Nida Gizem Yılmaz, Brittany Ming Chu Chan, Maria E T C van den Muijsenbergh, Julia C M van Weert, Barbara C Schouten. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org/>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Harnessing Telemedicine for the Provision of Health Care: Bibliometric and Scientometric Analysis

Ahmed Waqas<sup>1</sup>, MBBS, FRSPH; Soo Huat Teoh<sup>2</sup>, MD; Luís Velez Lapão<sup>3</sup>, PhD; Luiz Ary Messina<sup>4</sup>, PhD; Jorge César Correia<sup>5</sup>, MD

<sup>1</sup>Institute of Population Health, University of Liverpool, Liverpool, United Kingdom

<sup>2</sup>Lifestyle Science Cluster, Advanced Medical and Dental Institute, Universiti Sains Malaysia, Penang, Malaysia

<sup>3</sup>Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal

<sup>4</sup>Telemedicine University Network, Rede Nacional de Ensino e Pesquisa, Brasília, Brazil

<sup>5</sup>Unit of Patient Education, Division of Endocrinology, Diabetology, Nutrition and Patient Education, Department of Medicine, Geneva University Hospitals and University of Geneva, Geneva, Switzerland

**Corresponding Author:**

Jorge César Correia, MD

Unit of Patient Education

Division of Endocrinology, Diabetology, Nutrition and Patient Education

Department of Medicine, Geneva University Hospitals and University of Geneva

Chemin Venel 7

Geneva, 1206

Switzerland

Phone: 41 22 372 97 22

Email: [jorgecesar.correia@hcuge.ch](mailto:jorgecesar.correia@hcuge.ch)

## Abstract

**Background:** In recent decades, advances in information technology have given new momentum to telemedicine research. These advances in telemedicine range from individual to population levels, allowing the exchange of patient information for diagnosis and management of health problems, primary care prevention, and education of physicians via distance learning.

**Objective:** This scientometric investigation aims to examine collaborative research networks, dominant research themes and disciplines, and seminal research studies that have contributed most to the field of telemedicine. This information is vital for scientists, institutions, and policy stakeholders to evaluate research areas where more infrastructural or scholarly contributions are required.

**Methods:** For analyses, we used CiteSpace (version 4.0 R5; Drexel University), which is a Java-based software that allows scientometric analysis, especially visualization of collaborative networks and research themes in a specific field.

**Results:** We found that scholarly activity has experienced a significant increase in the last decade. Most important works were conducted by institutions located in high-income countries. A discipline-specific shift from radiology to telestroke, teledermatology, telepsychiatry, and primary care was observed. The most important innovations that yielded a collaborative influence were reported in the following medical disciplines, in descending order: public environmental and occupational health, psychiatry, pediatrics, health policy and services, nursing, rehabilitation, radiology, pharmacology, surgery, respiratory medicine, neurosciences, obstetrics, and geriatrics.

**Conclusions:** Despite a continuous rise in scholarly activity in telemedicine, we noticed several gaps in the literature. For instance, all the primary and secondary research central to telemedicine was conducted in the context of high-income countries, including the evidence synthesis approaches that pertained to implementation aspects of telemedicine. Furthermore, the research landscape and implementation of telemedicine infrastructure are expected to see exponential progress during and after the COVID-19 era.

(*J Med Internet Res* 2020;22(10):e18835) doi:[10.2196/18835](https://doi.org/10.2196/18835)

**KEYWORDS**

telemedicine; scientometric analysis; evidence synthesis; health information technology; research; theme

## Introduction

Advances in information and communication technologies (ICTs) have virtually reduced the world to a global village. The recent progress in ICTs has also shown incredible promise in addressing significant challenges in health care in disparate regions worldwide. Specifically, telemedicine ensures the provision of accessible, cost-effective, and specialized health care services in disparate areas. According to the World Health Organization (WHO), telemedicine pertains to the delivery of health care using different modalities embedded in the realms of information and communication technologies. It aims to advance health care, ranging from individual to population levels, by allowing exchange of patient information for diagnosis and management of health problems, primary care prevention, and education of physicians via distance learning [1]. Telemedicine is a new channel for health care services, which also enables opportunities to strengthen collaborative research.

The earliest evidence for telemedicine can be traced back to a clinical report published in *The Lancet* in 1879, which described the successful diagnosis of a child over the phone [2]. In addition, use of telegraphs was also evident in the American Civil War for transfer of mortality data and remote delivery of medical care [3]. A fine example of telemedicine was seen when, 20 years ago, the National Aeronautics and Space Administration monitored the astronauts' well-being during the Apollo mission to the moon. The modern form of telemedicine, however, appeared with the advent and maturation of the internet, which made possible the use of videoconferencing, high-quality data transfer, and distance learning platforms at a lower cost. The potential of telemedicine in strengthening health systems was also recently recognized by the WHO, leading to the establishment of the Global Observatory for eHealth in 2005 [1]. In 2009, the telemedicine module of the Global Observatory for eHealth mapped the development of telemedicine in 4 specialties of medicine—pathology, radiology, psychology, and dermatology—in 114 member states. This report found that the greatest development had been made in the provision of teleradiology services (33%) among the WHO member states, while 20% of the countries reported conducting a national review or evaluation of telemedicine. In addition, 50% of the member states reported that they had institutions dedicated to the development of telemedicine solutions [1].

The most recent report by the WHO, published in 2016, emphasized the role of eHealth in achieving universal health coverage [4]. The development of telemedicine was found crucial in the attainment of sustainable development goal 8, “achieve universal health coverage,” and goal 3, “ensure healthy lives and promote well-being for all at all ages,” thus ensuring health for all, regardless of creed, ethnicity, color, or finances. It also identified a rapid progress among its member states from 2010 to 2016. At least 83% of the countries had reported a mobile health (mHealth) initiative, widespread use of teleradiology (from 33% to 77%), telepathology or teledermatology (about 50%), and telepsychiatry (33%). In addition, e-learning initiatives were reported in 84% of countries and the use of national electronic health records in 47% of the member states, and 78% of the countries reported legislations

ensuring privacy of the electronic data. However, similar to the previous survey, very few countries had conducted evaluations of mHealth programs, which limits our understanding of the use of telemedicine, its barriers and facilitators, and which elements actually work [4]. These reports are a milestone in the field of telemedicine. However, these were dependent on data provided by government organizations and were heavily focused on government-run telehealth initiatives [4].

While the WHO-commissioned reports and evidence synthesis publications were indispensable in gauging worldwide infrastructure and legislation in telemedicine, scholarly research is a true marker for progress and evolution in every scientific field, and it is crucial to map research output in the field of telemedicine to determine prevalent research themes in order to guide policy makers and funders to improve or restrict flow of funding when required. Recognizing the importance of mapping progress in a field, scientists have devised several reproducible statistical methods that form the disciplines of bibliometrics and scientometrics. Scientometrics is defined as the “quantitative study of science, communication in science, and science policy” [5], and it helps evaluate the impact of journals, scientists, and institutions on the development and innovation of a scientific field.

Several studies published recently have used bibliometric methods to study progress in telemedicine, albeit in a very narrow context. For instance, Fatehi and Wootton [6] focused on delineating the use of different terminology to describe telemedicine, Groneberg et al [7] detailed the country-specific publication output and annual trends of publication and citation outputs, Gu et al [8] described the intellectual structure of telemedicine research by focusing on collaborative networks between different countries and authors, and Askari et al [9] provided an overview of the top 60 most frequently cited studies in telemedicine-specific journals. There is, however, a paucity of studies providing a holistic snapshot of advances in telemedicine from 2010 to 2019. The present analysis leverages the use of scientometric techniques to analyze publication output in the field of telemedicine worldwide, regardless of the government, industrial, or academic affiliations.

## Methods

### Bibliographic Search

An academic search of the Web of Science (WOS) core database was performed covering January 2010 to December 2019 to retrieve English language papers, using the following keyword: TS=(telemedicine). We restricted our search results to papers published in English only. The bibliographic records of these studies, including titles, abstracts, author details, affiliations, keywords, and citing references, were downloaded. We restricted our search results to the last 10 years to restrict our analyses in order to achieve a snapshot of recent research activity in telemedicine. The WOS core database was used for this academic search for two important reasons: it provides large coverage of over 20,000 peer-reviewed journals pertaining to 250 disciplines in health and social sciences and engineering and it is the only database that allows curation of citing



references of each indexed publication to allow cocitation analyses.

## Operational Definitions

Telemedicine literally means “healing at a distance.” However, there are no definitive definitions of telemedicine, with a recent review reporting over 104 peer-reviewed definitions found in the literature [10]. However, for the purpose of this scientometric investigation, we adapted a definition of telemedicine embodying 4 crucial elements: (1) provision of clinical support, (2) connection of users from different physical locations, (3) improved health outcomes, and (4) use of ICT [4,6].

## Knowledge Visualization Analyses

For analyses, we used CiteSpace (version 4.0 R5; Drexel University), which is a Java-based software that allows scientometric analysis, especially visualization of collaborative networks and research themes in a specific field [11,12]. The visualization of these collaborative networks in a discipline is based on the theory of cocitation, which posits that 2 documents share a cocitation relationship when they are cited together by another document [11,12]. For mapping of these networks, we ran network analyses using the cosine link reduction method and pathfinder networking scaling. Term sources were set as titles, abstracts, and author keywords, while nodes were set as cited references to delineate collaborative networks and cluster analyses. The time-splicing method was used to explore publications in 2 periods, 2010 to 2014 and 2015 to 2019, where each slice comprised the top 50 cited papers every year.

To obtain clusters or themes of research, we ran cluster analyses in which each cluster was termed according to publication keywords using 2 text-mining methods: term frequency-inverse

document frequency (TF-IDF) and log likelihood ratio (LLR). The first method, TF-IDF, uses terms that are weighted by term frequencies multiplied by inverted document frequencies [11,12]. LLR tests choose the most appropriate clustering label by assessing the strength of the bond between a term and a cluster [11,12]. A cluster is said to be parsimonious when it comprises a larger number of items and an acceptable silhouette and modularity value (Q). The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared with other clusters (separation) [11,12]. The value of Q and silhouette ranges between  $-1$  to  $1$ , where a value closer to  $1$  is considered acceptable.

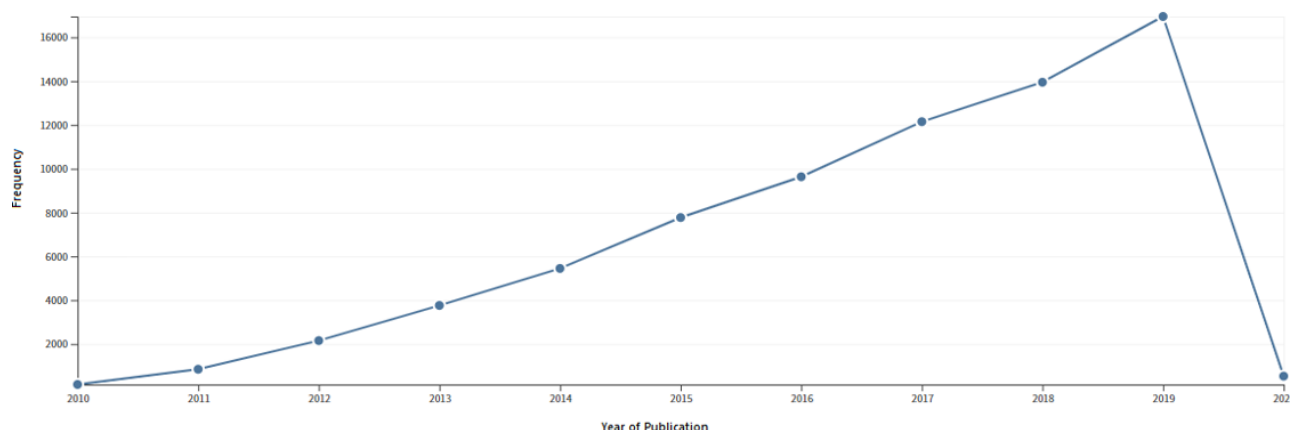
Each paper is presented as a node and links between two nodes as edges. The collaborative network mapped from this analysis yields several important results. Any research study with centrality values  $\geq 0.1$  are considered influential in their collaborative networks. Citation rings show annual citation patterns, while purple nodes represent landmark theories or works that give rise to a new body of work. Citation bursts revealing short periods of high scholarly activity are presented as red rings. Based on these cues, researchers can identify important works in a field and important themes of research.

## Results

### Bibliometrics

Web of Science (core database) yielded 6896 publications from 2010 to 2019, with a total h-index of 87 and an average 10.64 citations per study. These were cited a total of 73,354 times by a total of 42,381 citing papers. There was an increasing trend in both the publication and citation output from 2010 to 2019 (Figure 1).

**Figure 1.** Trends of citations received by papers in telemedicine.



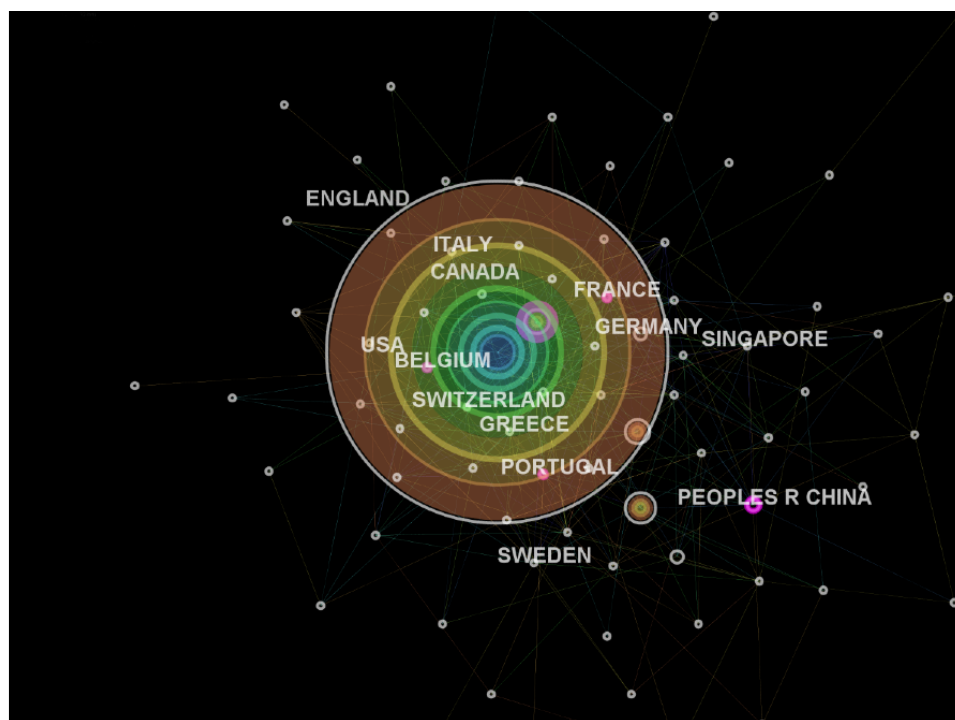
### Regional Trends in Telemedicine Research

Regionally, the highest publication output (in English) was contributed by high-income countries: the United States, Australia, England, Canada, and Germany. There were 2 middle-income countries, India and the People’s Republic of China, that also ranked in the top 10 for publication output. However, in terms of having a centrality score of 0.1 or greater, 6 countries—England, France, Belgium, Portugal, the People’s

Republic of China, and Greece—appeared to hold significant influence in worldwide collaborations in telemedicine. These countries are also presented as purple nodes in Figure 2 (ie, contributing groundbreaking research). These central nations, although mainly high-income countries, also formed collaborations with a number of low- and middle-income countries (LMICs), such as Ethiopia, Mali, Botswana, Nepal, Zimbabwe, Pakistan, and Uganda.



**Figure 2.** Regional collaborative networks in the area of telemedicine research. Peoples R China: People's Republic of China; USA: United States of America.

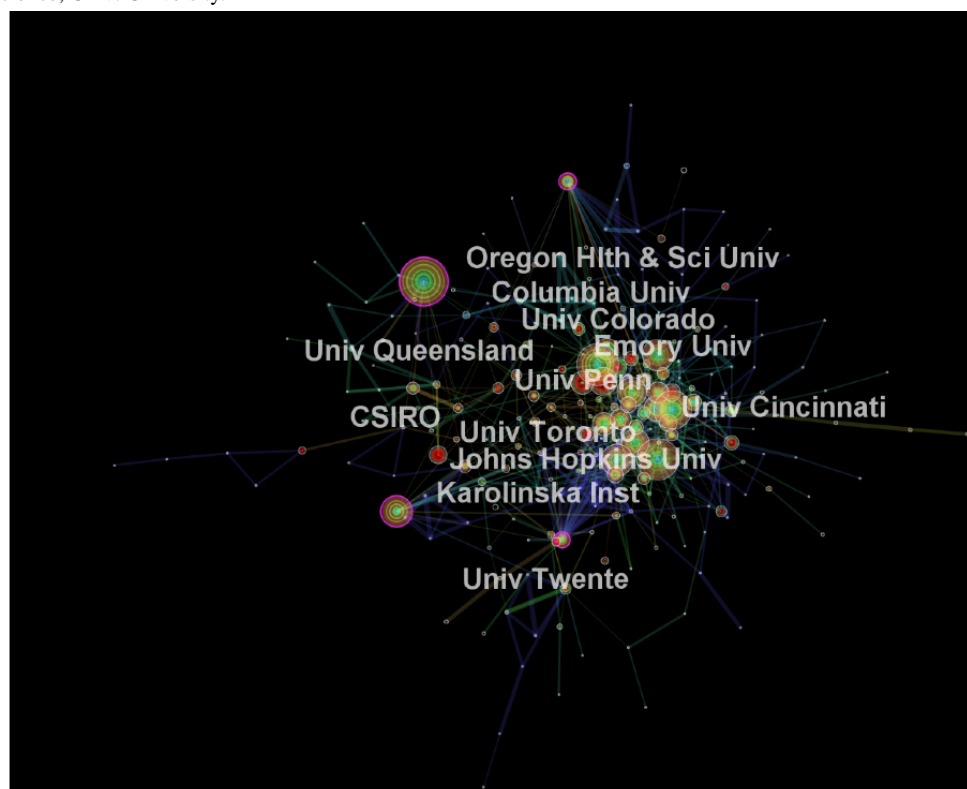


### Institutional Trends in Telemedicine Research

Among institutions, the top 9 contributing institutions in terms of publication output were based in the United States, including the University of California system (n=304), Harvard University (n=227), and the Pennsylvania Commonwealth System of Higher Education (n=152). Outside of the United States, the

University of Queensland in Australia was the fifth-highest contributing region. Institutions with centrality  $\geq 0.1$  included Columbia University, University of Queensland, University of Toronto, and Karolinska Institute. It is noteworthy that none of the top 9 US institutions in terms of publication output were deemed important in their collaborative networks (Figure 3).

**Figure 3.** Institutional collaborations in telemedicine research. CSIRO: Commonwealth Scientific and Industrial Research Organisation; Hlth: Health; Inst: Institute; Sci: Science; Univ: University.

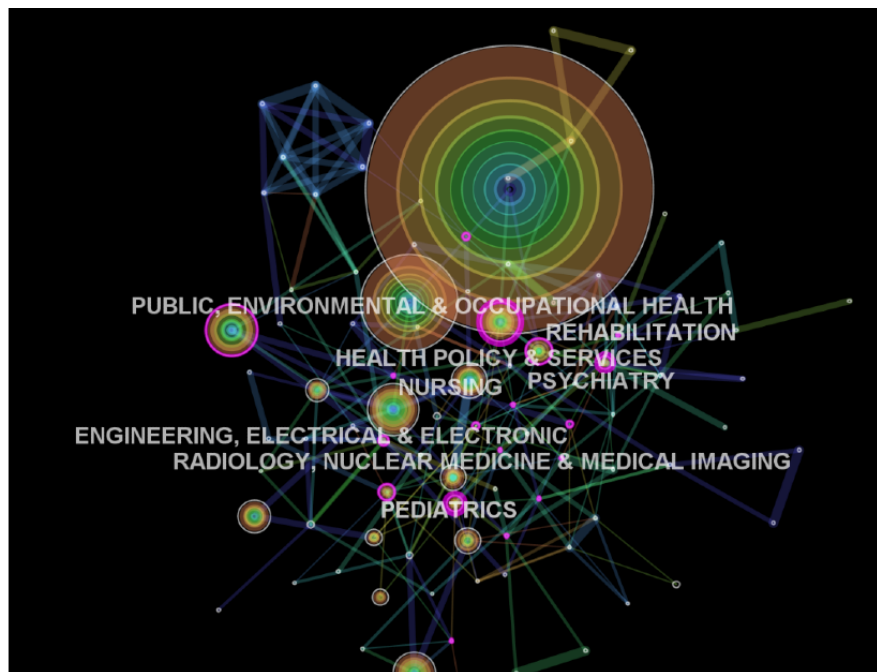


### Discipline-Specific Trends in Telemedicine Research

An analysis of the Web of Science revealed that the disciplines of health care sciences and services ( $n=2338$ ) and medical informatics ( $n=835$ ) reported the highest number of publication items. However, a total of 19 disciplines had the greatest centrality values ( $\geq 0.1$ ). Most important innovations yielding a

collaborative influence were reported in the following medical disciplines, in descending order: public environmental and occupational health, psychiatry, pediatrics, health policy and services, nursing, rehabilitation, radiology, pharmacology, surgery, respiratory medicine, neurosciences, obstetrics, and geriatrics (Figure 4).

**Figure 4.** Discipline-specific trends in telemedicine research.

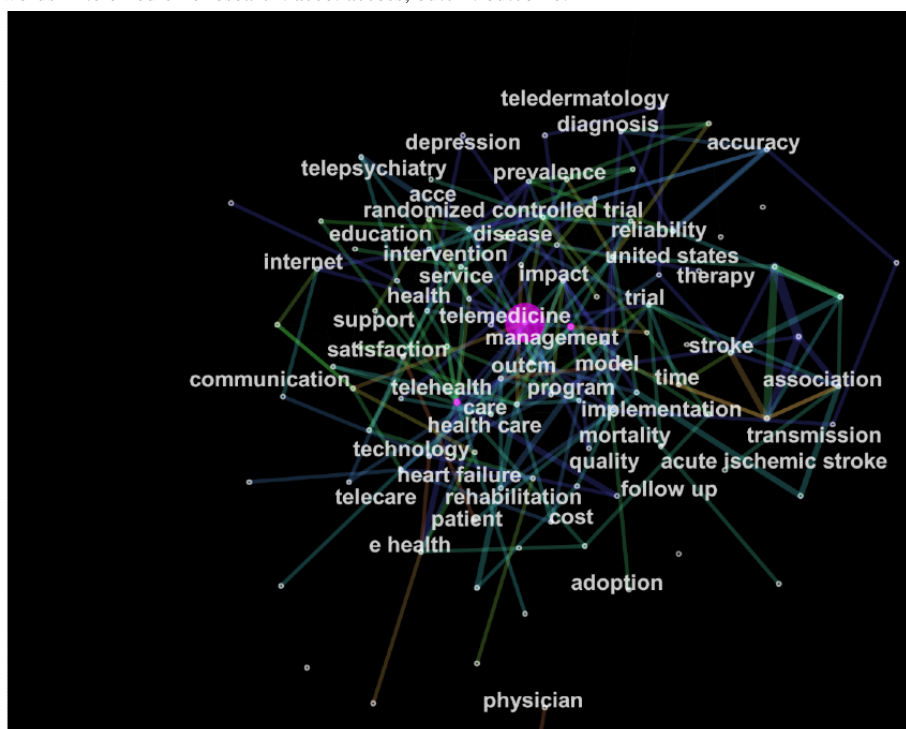


### Keyword Analysis in Telemedicine

A total of 106 research keywords were identified in the field of telemedicine, revealing the most-researched topics (Figure 5). From 2010 to 2019, 32 keywords appeared to have citation

outbursts showing the greatest research activity in telemedicine (Figure 6). The top 50 cited keywords in tandem with citation outbursts were divided into themes to identify the most frequently researched diseases, outcomes, study designs, and populations, shown in Table 1.

**Figure 5.** Research keywords in telemedicine research. acce: access; outcm: outcome.



**Figure 6.** Research hotspots in telemedicine from 2010 to 2019.

Keywords	Year	Strength	Beginning	End	2010-2019
Information technology	2010	17.007	2010	2012	
Accuracy	2010	12.802	2010	2011	
Telepsychiatry	2010	18.634	2010	2015	
Telecare	2010	22.799	2010	2014	
Therapy	2010	12.430	2010	2013	
Home	2010	10.414	2010	2012	
Consultation	2010	6.987	2010	2013	
Internet	2010	9.719	2010	2011	
Videoconferencing	2010	10.150	2011	2012	
Communication	2010	14.967	2011	2013	
Satisfaction	2010	5.538	2011	2013	
Telemonitoring	2010	16.053	2011	2013	
Information	2010	10.394	2011	2012	
Association	2010	3.462	2012	2013	
Follow-up	2010	7.651	2012	2016	
Telestroke	2010	7.125	2012	2014	
Performance	2010	9.818	2012	2014	
Telecommunication	2010	8.851	2013	2014	
Cost	2010	4.394	2013	2014	
Thrombolysis	2010	15.310	2013	2014	
Self-management	2010	16.288	2013	2015	
Experience	2010	6.828	2014	2015	
Diabetic retinopathy	2010	8.019	2014	2015	
Veteran	2010	11.995	2014	2015	
Support	2010	9.430	2014	2016	
Smartphone	2010	3.520	2015	2019	
Teledermatology	2010	7.738	2015	2016	
Barrier	2010	11.633	2016	2019	
Meta-analysis	2010	22.009	2016	2019	
Feasibility	2010	4.247	2016	2017	
United States	2010	15.402	2017	2019	
Adherence	2010	5.924	2017	2019	

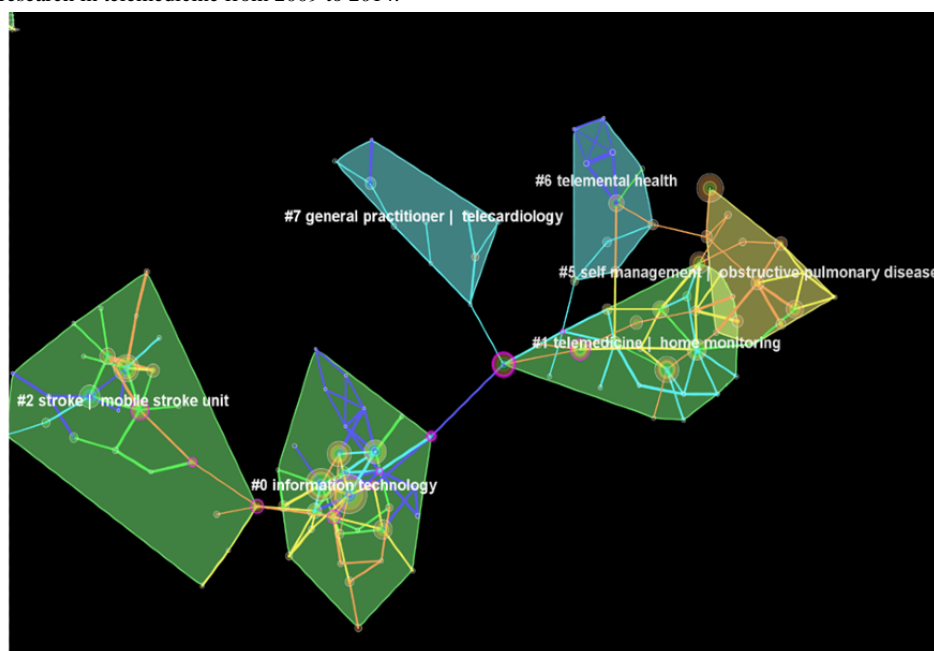
**Table 1.** Research keywords in telemedicine by theme.

Theme	Most frequent keywords
Diseases	Diabetic retinopathy, stroke, depression, heart failure, rehabilitation, thrombolysis
Performance indicators	Feasibility, accuracy, reliability, barrier, adherence, satisfaction, performance, cost-effectiveness
Outcomes	Self-management, support, impact, diagnosis, education, mortality, quality of life, telemonitoring
Study design	Association, follow-up, meta-analysis, randomized controlled trials, implementation, prevalence
Disciplines	Telestroke, teledermatology, telepsychiatry, primary care
Setting	Home, internet, videoconferencing, communication, telecommunication, smartphone
Population	Veterans, United States, children

## Clusters of Research in Telemedicine From 2009 to 2014

From 2009 to 2014, a total of 2527 papers were published, which were cited by 141,702 references. These were analyzed to study landmark publications and clusters of research during

this period. There was a total of 228 nodes and 273 edges. Cluster analysis yielded a parsimonious network of clusters (Figure 7) with a modularity of 0.85 and silhouette value of 0.42. It yielded a total of 56 clusters, out of which 8 comprised at least 10 studies each and an acceptable silhouette value (range 0.85 to 1.0).

**Figure 7.** Clusters of research in telemedicine from 2009 to 2014.

## Acute Medicine

### Stroke

The zeroth cluster comprised 32 items with a silhouette value of 0.98, described by terms such as information technology (TF-IDF) and acute ischemic stroke and thrombolysis (LLR). The third cluster pertained to mobile stroke units (TF-IDF) and associated mortality and economic outcomes (LLR).

### Telecardiology

The second cluster pertained to telemonitoring at home (TF-IDF) for chronic diseases such as heart failure (LLR). The fifth cluster was focused on clinical management and home monitoring (TF-IDF) of pacemaker activity and implantable cardioverter defibrillator (LLR), especially in patients with myocardial infarction.

## Chronic Diseases

### Diabetes

Cluster 7 pertained to general practitioner-mediated telecardiology (TF-IDF) and user acceptance of this program (LLR). The sixth cluster pertained to nutrition in diabetes, especially in the context of the Columbia University Informatics for Diabetes Education and Telemedicine project.

### Telepsychiatry

The third cluster was defined as neuropsychological assessment (by TF-IDF method), focusing on posttraumatic stress disorder (PTSD) and cognitive behavioral therapy (LLR method).

### Respiratory Medicine

This cluster (the fifth cluster) pertained to self-management of chronic obstructive pulmonary disease (COPD) and its exacerbation (TF-IDF and LLR).

## Landmark Publications From 2009 to 2014

The period from 2009 to 2014 revealed 10 landmark publications pertaining to different themes, where most of the publications pertained to intensive care, especially stroke (Table 2) [13-16]. Audebert et al published 3 important studies [17-19] demonstrating the success of the telemanagement of stroke in rural hospitals in Bavaria, Germany. In a similar context, Schwamm et al [13] provided evidence for telestroke consultations via videoconferencing and Lilly et al [14] showed that the implementation of a tele-intensive care unit (ICU) intervention was associated with reduced adjusted odds of mortality and reduced length of hospital stay, as well as with

changes in best-practice adherence and lower rates of preventable complications. In their cross-sectional survey, Silva et al [16] described the status of telestroke programs in the United States. In addition, 2 important literature reviews were published during this period. Sood et al [10] improved our understanding of the theoretical underpinnings of modern telemedicine after a careful evaluation of 104 peer-reviewed publications, while Kahn et al [15] summarized the recommendations of a working group for the adoption of a standardized framework for the standardized conduct of tele-ICU studies. Lastly, May et al [20] described the complexity that exists in the scale-up of telemedicine programs, which is often underestimated, leading to their failures.

**Table 2.** Lessons learned from landmark publications from 2009 to 2014.

Author (year)	Study design	Theme	Disease	Lesson learned
Broens (2007) [21] <sup>a</sup>	Qualitative literature review	Implementation	N/A <sup>b</sup>	Determinants of successful implementation and scale-up of telemedicine programs. Important determinants of telemedicine programs include (1) technology, (2) acceptance, (3) financing, (4) organization, and (5) policy and legislation.
Schwamm et al (2004) [13]	Retrospective	Feasibility	Stroke	Telestroke consultation via videoconferencing improved care in 95% of the cases.
Silva et al (2012) [16] <sup>a</sup>	Cross-sectional survey	Barriers and facilitators	Stroke	Status of telestroke in the United States. The top 3 clinical needs met by telestroke were emergency department consultation (100%), patient triage (83.8%), and inpatient teleconsultation (46.0%).
Kahn et al (2011) [15] <sup>a</sup>	Working group statement	Guidelines	Intensive care	This working group meeting was convened to address methodological and knowledge gaps in the field. It proposed adoption of a common framework to facilitate standardized conduct of telemedicine studies in the ICU <sup>c</sup> .
Audebert et al (2005) [18]	Retrospective	Feasibility	Stroke	Telemedicine provided a cost-effective method to recommend use of thrombolysis among patients presenting with stroke in rural regions.
Lilly et al (2011) [14] <sup>a</sup>	Prospective stepped-wedge clinical trial	Effectiveness	Intensive care	Implementation of a tele-ICU intervention was associated with reduced adjusted odds of mortality and reduced length of hospital stay, as well as with changes in best-practice adherence and lower rates of preventable complications.
May et al (2003) [20]	Qualitative study	Implementation	N/A	Complexity exists at 4 discrete levels in any given telehealth context: implementation, adoption, translation, and stabilization. This complexity is often underestimated, leading to failed scale-ups.
Sood et al (2007) [10] <sup>a</sup>	Literature review	Theoretical underpinnings	N/A	Defined modern telemedicine after a careful review of 104 publications.
Audebert et al 2006 [17]	Prospective	Feasibility	Stroke	The telestroke concept promises better coverage of systemic thrombolysis in nonurban areas.
Audebert et al 2006 [19] <sup>a</sup>	Nonrandomized clinical trial	Trial	Stroke	Treatment in rural hospitals independently reduced the probability of a poor outcome compared with controls.

<sup>a</sup>Purple nodes in Figure 7 representing seminal work in the area of telemedicine.

<sup>b</sup>N/A: not applicable.

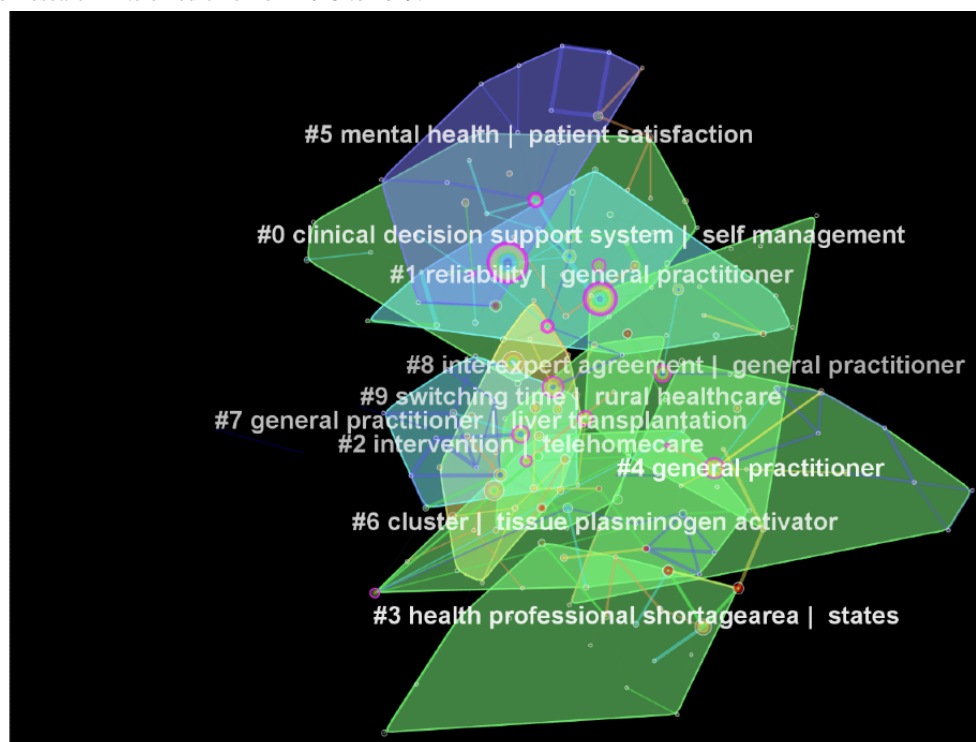
<sup>c</sup>ICU: intensive care unit.

## Clusters of Research in Telemedicine From 2015 to 2019

From 2015 to 2019, a total of 4493 records were published, which were cited a total of 141,702 times. Cluster analyses yielded a parsimonious cluster network with a modularity of 0.69 and a silhouette value of 0.39. To get a snapshot of research themes in this period, we analyzed a total of 205 nodes with

345 edges. A total of 27 clusters of research in telemedicine were identified, out of which 12 had an acceptable silhouette value. In size, these clusters ranged from 8 to 23 studies, and modularity values ranged from 0.97 to 0.71. These clusters fell into 4 major themes: (1) clinical decision support systems, (2) reliability, (3) access to health care, and (4) medical conditions (Figure 8).



**Figure 8.** Clusters of research in telemedicine from 2015 to 2019.

### *Clinical Decision Support Systems*

The zeroth cluster pertained to clinical decision support systems to aid in self-management (TF-IDF), explored in the context of ulcerative colitis (LLR) and lung cancer (mutual information).

### *Reliability*

A total of 2 clusters (1 and 8) focused on reliability and interexpert agreement (TF-IDF) pertaining to telemedicine-aided diagnoses mediated by general practitioners (TF-IDF), especially in the field of teler dermatology (LLR).

### *Access to Health Care*

A total of 3 clusters (2, 3, and 9) pertained to this thematic area of telemedicine, defined by access to telemedicine in shortage areas (TF-IDF) to aid in the diagnosis of diabetic retinopathy. Rural health care was an important component of this research theme, where the issue of switching time between pediatric consultations was thoroughly researched.

### *Medical Conditions*

A total of 4 clusters (4, 5, 6, and 7) focused on both acute and chronic conditions, for instance, general practitioner-mediated mental health care, especially in PTSD. Cluster 5 pertained to telecardiology, in which heart failure, remote monitoring of pacemaker activity, and patient satisfaction were important areas of research. In accordance with research prior to 2015, acute ischemic stroke and thrombolysis were important research areas in telestroke. In the specialty of tele-ICU, the care of critically ill patients, especially those undergoing liver transplantation, and economic outcome were the most researched areas.

### *End Consumer Research*

Patient compliance, safety, and satisfaction were explored in 2 clusters (10 and 11).

### **Landmark Publications From 2015 to 2019**

This period of scholarly activity in telemedicine continued to be influenced by 4 studies published prior to 2015 [13,14,22-24], pertaining to tele-ICU, telestroke, tele-mental health, and facilitators and barriers to telemedicine. The majority of publications unique to this time period were literature reviews, systematic or otherwise (n=10), followed by retrospective studies (n=2) and a clinical trial (n=1). Major themes in this era were effectiveness and cost-effectiveness research (both primary and secondary). In addition, the connected health model of health care, which governs telemedicine, and the standardized framework for assessment of telemedicine commissioned by the European Commission were deemed central in these collaborative networks.

The most influential review in this period pertained to chronic diseases and was an evidence synthesis report on 141 randomized controlled trials relating to asthma, COPD, diabetes, heart failure, and hypertension [25]. It reported strong evidence of publication bias, with 108 randomized controlled trials reporting positive results and almost none reporting harm [25]. Wade et al [26] presented a systematic review regarding economic analysis of telemedicine and concluded that the delivery of health services by real-time video communication was cost-effective for home care and access to on-call hospital specialists.

Effectiveness research was conducted using both primary interventional and evidence synthesis approaches. For instance, Chaudhry et al [27] conducted a clinical trial and showed that telemonitoring did not improve outcomes among patients hospitalized for heart failure. In their reviews, Ekland et al [28] and Flodgren et al [29] examined effectiveness of telemedicine in multiple conditions [28,29]; Elbert et al [30] focused on effectiveness and cost-effectiveness for somatic

diseases and Hubley et al [31], on psychiatric diseases. Bashshur et al [32] examined 3 tracer diseases (heart failure, stroke, and COPD), which, when treated using telemedicine approaches, showed several markers of improvement, such as reduced hospital admissions and readmissions, length of hospital stay, and emergency department visits. Fierson et al [33] reviewed the currently available literature on telemedicine-based remote digital fundus imaging evaluations for retinopathy of prematurity and outlined pertinent practical and risk management considerations.

Kvedar et al [34] presented a model of care to make telemedicine an important part of the US health care system. He reported that

care processes in the United States are insufficient to address the mismatch in supply and demand of health care providers [34]. This review presented connected health as a new care model to improve patient care with telemedicine and telehealth. Kidholm et al [35], after synthesizing evidence from a stakeholders meeting, presented a framework for the assessment of telemedicine with 7 important domains: (1) health problem and description of the application, (2) safety, (3) clinical effectiveness, (4) patient perspectives, (5) economic aspects, (6) organizational aspects, and (7) sociocultural, ethical, and legal aspects. A summary of these publications is provided in Table 3.

**Table 3.** Lessons learned from landmark publications from 2015 to 2019.

Author (year)	Study design	Theme	Disease	Lesson learned
Wootton (2012) [25] <sup>a</sup>	Review	Evidence synthesis	Chronic diseases	This study presents an evidence synthesis report on 141 RCTs <sup>b</sup> pertaining to asthma, COPD <sup>c</sup> , diabetes, heart failure, and hypertension. There was strong evidence of publication bias, with 108 RCTs reporting positive results and almost none reporting harm.
Wade et al (2010) [26]	Systematic review	Economic analysis	— <sup>d</sup>	Delivery of health services by real-time video communication was cost-effective for home care and access to on-call hospital specialists, showed mixed results for rural service delivery, and was not cost-effective for local delivery of services between hospitals and primary care.
Chaudhry et al (2010) [27] <sup>a</sup>	Clinical trial	Effectiveness	Heart failure	Telemonitoring did not improve outcomes among patients hospitalized for heart failure.
Ekeland et al (2010) [28] <sup>a</sup>	Systematic review of systematic reviews	Effectiveness	—	Out of 80 included systematic reviews, 21 showed that telemedicine was effective, and 18 reported that evidence regarding telemedicine was limited and inconsistent.
Kvedar et al (2014) [34] <sup>a</sup>	Literature review	Model of health care	—	Care processes in the United States are insufficient to address the mismatch in supply and demand of health care providers. This review presented connected health as a new care model to improve patient care with telemedicine and telehealth.
Elbert et al (2014) [30]	Systematic review of systematic reviews	Effectiveness and cost-effectiveness	Somatic diseases	Out of 31 eligible reviews, 7 found eHealth to be clinically effective and cost-effective and 13 found it to be promising, while the rest found the evidence to be limited or inconsistent.
Bashshur et al (2014) [32] <sup>a</sup>	Systematic review	General review	Chronic diseases: heart failure, stroke, and COPD	The 3 diseases, when treated using telemedicine approaches, showed several improvements, such as reduced hospital admissions and readmissions, length of hospital stay, and emergency department visits.
Flodgren et al (2015) [29] <sup>a</sup>	Systematic review and meta-analysis	Effectiveness	Cardiovascular disease, diabetes, respiratory conditions, mental health or substance abuse conditions, conditions requiring a specialist consultation, comorbidities, urogenital conditions, neurological injuries and conditions, gastrointestinal conditions, neonatal conditions requiring specialist care, solid-organ transplantation, and cancer	There was high- to moderate-certainty evidence that there was no significant difference between telemedicine and usual health care in improving all-cause mortality and admissions to the hospital. There was some evidence of improved quality of life, lower HbA <sub>1c</sub> <sup>e</sup> among patients with diabetes, and decreased LDL <sup>f</sup> and blood pressure. Participants with different mental health and substance abuse problems reported no differences in the effect of therapy delivered over videoconferencing compared with face-to-face delivery.
Kidholm et al (2012) [35] <sup>a</sup>	Recommendations based on workshops with users and stakeholders of telemedicine, initiated by European Commission	Framework for assessment of telemedicine	—	There are 7 domains in MAST <sup>g</sup> : (1) health problem and description of the application, (2) safety, (3) clinical effectiveness, (4) patient perspectives, (5) economic aspects, (6) organizational aspects, and (7) sociocultural, ethical, and legal aspects.
Schwamm et al (2009) [13] <sup>a</sup>	See Table 2	See Table 2	See Table 2	See Table 2
Lilly et al (2011) [14] <sup>a</sup>	See Table 2	See Table 2	See Table 2	See Table 2

Author (year)	Study design	Theme	Disease	Lesson learned
Hilty et al (2013) [23] <sup>a</sup>	Review	Effectiveness	Mental health	This review reported that tele-mental health interventions are effective and improve access to care. More research is required on service models and ethical and cross-cultural aspects of tele-mental health.
Dharmar et al (2013) [22] <sup>a</sup>	Retrospective	Quality improvement	Pediatric critical care	Telemedicine consultations were associated with higher physician-rated quality of care and parent satisfaction.
Sanders et al (2012) [24]	Qualitative	Evaluation and barriers to adoption	Telehealth in general	This qualitative investigation examined barriers to participation and adoption of telehealth among people who withdrew from a UK-based clinical trial on telemedicine.
Fierston et al (2015) [33]	Review	Evaluation for retinopathy of prematurity	Retinopathy of prematurity	This report reviewed the currently available literature on RDTF-TM <sup>h</sup> evaluations for retinopathy of prematurity and outlined pertinent practical and risk management considerations.
Ashwood et al (2017) [36]	Retrospective	Effectiveness and cost-effectiveness	—	Direct-to-consumer telehealth may increase access to care but does not decrease spending; 12% of direct-to-consumer telehealth visits replaced visits to other providers, and 88% represented new use. Net annual spending on acute respiratory illness increased \$45 per telehealth user.
Hubble et al (2016) [31]	Systematic review	Effectiveness	Psychiatric diseases	A large evidence base supported telepsychiatry as a delivery method for mental health services. Future studies will inform optimal approaches to implementing and sustaining telepsychiatry services.

<sup>a</sup>Purple nodes in Figure 7 representing seminal work in the area of telemedicine.

<sup>b</sup>RCTs: randomized controlled trials.

<sup>c</sup>COPD: chronic obstructive pulmonary disease.

<sup>d</sup>Not available.

<sup>e</sup>HbA<sub>1c</sub>: glycated hemoglobin.

<sup>f</sup>LDL: low-density lipoprotein.

<sup>g</sup>MAST: model for assessment of telemedicine.

<sup>h</sup>RDTF-TM: telemedicine-based remote digital fundus imaging.

## Discussion

### Summary

This scientometric analysis presents an overview of scholarly work in the field of telemedicine in the last 10 years. It shows the transition of scholarly work in this field from teleradiology in the previous decade to mental health, stroke, and critical care medicine. Barriers and facilitators to successful implementation of telemedicine were also seen as an important area of research in telemedicine. Collaborative networks between regions and institutions revealed collaborative links between central global institutions and LMICs, showing a transfer of technology and expertise to disparate regions. Among the LMICs, China and India are emerging as big players in telemedicine.

### General Trends and Transcontinental Collaborations

Our analysis revealed a steadily increasing publication output and citation activity in the field of telemedicine, which is in consonance with previous literature [7,9,25,37,38]. In terms of regional output, a bibliometric assessment of literature in telemedicine from 1980 to 2013 showed that the top 5 countries in terms of publication output were the United States, the United Kingdom, Germany, Canada, and Australia, while China ranked tenth [37]. However, we opine that the centrality or influence of a particular entity in their collaborative networks and the volume of innovative work may be better indicators of progress

in a field. In this vein, England, France, Belgium, Portugal, the People's Republic of China, and Greece appeared to hold significant influence worldwide. Similarly, almost all of the top institutions with regard to publication output were from the United States, which reflects previous literature [7,9,25,37,38]. However, only 1 of the United States-based universities was found to be central in its domain. Top institutions were Columbia University, University of Queensland, University of Toronto, and Karolinska Institute.

Several of the top institutions were involved in collaborations with institutions from LMICs, indicating transfer of technology and expertise. This is an important endeavor, as studying the effectiveness and uptake of telemedicine may decrease disparities in these regions. Portugal, for instance, provides a good case study to examine collaborative networks between high-income and low-income countries. A transcultural pediatric telecardiology service has been established in several Portuguese-speaking African countries in collaboration with Portugal-based universities [39]. This program has been highly successful. For instance, in Angola alone, it has performed 32,685 outpatient teleconsultations (1998 to 2016), saving health system costs [39]. Another important endeavor includes echocardiography services through a telecardiology initiative being provided in Tanzania, Malawi, Mali, and Mozambique with a telereporting center in Italy [40]. On October 26, 2017, another impetus for telemedicine research and implementation

was provided by a resolution that created the Comunidade dos Países de Língua Portuguesa's Permanent Working Group on Telemedicine and Telehealth during the fourth Health Ministers Meeting of the Portuguese-Speaking Countries in Brasília [41].

### Transition of Research Themes

We noticed a transition in research themes in telemedicine during these periods. For instance, the WHO reports cited teleradiology services as being most prevalent worldwide. In line with this, Armfield et al [42], using text-mining approaches, reported that during the early period of telemedicine research from 1970 to 1995, teleradiology and telepathology were the most dominant fields, as well as the first fields to adopt telemedicine. In contrast, research trends in a more recent period (2009 to 2013) focused on cost-effectiveness, and the clinical and discipline-specific terms “diabetic” and “stroke” emerged in this period. Our analyses revealed that these themes progressed into the established fields of telecardiology, telestroke, and tele-ICU. Moreover, we also saw a rise in cost-effectiveness as well as implementation and feasibility research, which are very important aspects in the uptake of telemedicine. All the influential studies in our analyses pertained to these themes.

### Research Gaps and Recommendations for Future Work

Despite a continuous rise in scholarly activity in telemedicine, we noticed several gaps in the literature. For instance, all the primary and secondary research central to telemedicine was conducted in the context of high-income countries, including the evidence synthesis approaches pertaining to implementation aspects of telemedicine. In addition, patient confidentiality and ethical perspectives on the use of telemedicine were nonexistent in our analysis. Most of the telemedicine research in LMICs was driven in collaboration with high-income countries. There is a huge gap in needs-based analysis, eHealth literacy, and

inclusion of Indigenous end consumers and stakeholders in the design of telemedicine platforms in LMICs.

There were also no research clusters on improving eHealth literacy, especially in the context of use of telemedicine in LMICs. The lack of strengths, weaknesses, opportunities, and threats analyses, particularly in the evaluation of eHealth literacy among physicians, is a big factor in the failure of telemedicine. This was a major contributory factor in the failure of the Réseau en Afrique Francophone pour la Télémédecine (RAFT) telemedicine software platform in Angola, which enjoyed commitment from the Ministry of Health and local stakeholders but was not taken up by the participating physicians [43].

Telemedicine financing is a critical aspect for sustainability and most often not covered in studies. The development of telemedicine on a global scale will require more sophisticated business models. Additionally, telemedicine skills development is very seldom provided by medical schools.

Telemedicine is still in its infancy in LMICs, and there is a lack of clarity in several important aspects, such as the development and adoption of ethical standards, treatment protocols, and guidelines. Medical informaticians should liaise with health care centers, physicians, and medical ethicists to develop software promoting an ethos of confidentiality, privacy, and security during the sharing of sensitive data.

### Conclusion

The findings in this investigation suggest a rapid development in the field of telemedicine, albeit prior to the COVID-19 pandemic. We expect that the research landscape and implementation of telemedicine infrastructure may see exponential progress during and after the COVID-19 period. This is also echoed in the recent report by the American Medical Association, which predicts that “after COVID-19, \$250 billion in care could shift to telehealth, boosting research and infrastructural development” [44].

### Conflicts of Interest

None declared.

### References

1. Ryu S. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2). *Healthc Inform Res* 2012;18(2):153. [doi: [10.4258/hir.2012.18.2.153](https://doi.org/10.4258/hir.2012.18.2.153)]
2. Ryu S. History of Telemedicine: Evolution, Context, and Transformation. *Healthc Inform Res* 2010;16(1):65. [doi: [10.4258/hir.2010.16.1.65](https://doi.org/10.4258/hir.2010.16.1.65)]
3. Zundel K. Telemedicine: history, applications, and impact on librarianship. *Bull Med Libr Assoc* 1996 Jan;84(1):71-79 [FREE Full text] [Medline: [8938332](https://pubmed.ncbi.nlm.nih.gov/8938332/)]
4. World Health Organization. Global diffusion of eHealth: Making universal health coverage achievable. Report of the third global survey on eHealth Global Observatory for eHealth. 2016. URL: [http://who.int/goe/publications/global\\_diffusion/en/](http://who.int/goe/publications/global_diffusion/en/) [accessed 2020-09-18]
5. Hess D. Science studies: an advanced introduction. New York, NY: New York University Press; 1997.
6. Fatehi F, Wootton R. Telemedicine, telehealth or e-health? A bibliometric analysis of the trends in the use of these terms. *J Telemed Telecare* 2012 Dec;18(8):460-464. [doi: [10.1258/jtt.2012.gth108](https://doi.org/10.1258/jtt.2012.gth108)] [Medline: [23209265](https://pubmed.ncbi.nlm.nih.gov/23209265/)]
7. Groneberg DA, Rahimian S, Bundschuh M, Schwarzer M, Gerber A, Kloft B. Telemedicine - a scientometric and density equalizing analysis. *J Occup Med Toxicol* 2015;10:38 [FREE Full text] [doi: [10.1186/s12995-015-0076-3](https://doi.org/10.1186/s12995-015-0076-3)] [Medline: [26500688](https://pubmed.ncbi.nlm.nih.gov/26500688/)]



8. Gu D, Li T, Wang X, Yang X, Yu Z. Visualizing the intellectual structure and evolution of electronic health and telemedicine research. *Int J Med Inform* 2019 Oct;130:103947. [doi: [10.1016/j.ijmedinf.2019.08.007](https://doi.org/10.1016/j.ijmedinf.2019.08.007)] [Medline: [31450080](#)]
9. Askari A, Khodaie M, Bahaadinbeigy K. The 60 most highly cited articles published in the Journal of Telemedicine and Telecare and Telemedicine Journal and E-health. *J Telemed Telecare* 2014 Jan 10;20(1):35-43. [doi: [10.1177/1357633x13519899](https://doi.org/10.1177/1357633x13519899)]
10. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. *Telemed J E Health* 2007 Oct;13(5):573-590. [doi: [10.1089/tmj.2006.0073](https://doi.org/10.1089/tmj.2006.0073)] [Medline: [17999619](#)]
11. Chen C. The CiteSpace Manual v1.05. College of Computing and Informatics, Drexel University. 2015. URL: <http://cluster.ischool.drexel.edu/~cchen/citespace/manual/CiteSpaceManual.pdf> [accessed 2020-09-18]
12. Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inf Sci* 2006 Feb 01;57(3):359-377. [doi: [10.1002/asi.20317](https://doi.org/10.1002/asi.20317)]
13. Schwamm LH, Rosenthal ES, Hirshberg A, Schaefer PW, Little EA, Kvedar JC, et al. Virtual TeleStroke support for the emergency department evaluation of acute stroke. *Acad Emerg Med* 2004 Nov;11(11):1193-1197. [doi: [10.1197/j.aem.2004.08.014](https://doi.org/10.1197/j.aem.2004.08.014)] [Medline: [15528584](#)]
14. Lilly CM, Cody S, Zhao H, Landry K, Baker SP, McIlwaine J, University of Massachusetts Memorial Critical Care Operations Group. Hospital mortality, length of stay, and preventable complications among critically ill patients before and after tele-ICU reengineering of critical care processes. *JAMA* 2011 Jun 01;305(21):2175-2183. [doi: [10.1001/jama.2011.697](https://doi.org/10.1001/jama.2011.697)] [Medline: [21576622](#)]
15. Kahn JM, Hill NS, Lilly CM, Angus DC, Jacobi J, Rubenfeld GD, et al. The research agenda in ICU telemedicine: a statement from the Critical Care Societies Collaborative. *Chest* 2011 Jul;140(1):230-238 [FREE Full text] [doi: [10.1378/chest.11-0610](https://doi.org/10.1378/chest.11-0610)] [Medline: [21729894](#)]
16. Silva GS, Farrell S, Shandra E, Viswanathan A, Schwamm LH. The Status of Telestroke in the United States. *Stroke* 2012 Aug;43(8):2078-2085. [doi: [10.1161/strokeaha.111.645861](https://doi.org/10.1161/strokeaha.111.645861)]
17. Audebert HJ, Kukla C, Vatankeh B, Gotzler B, Schenkel J, Hofer S, et al. Comparison of Tissue Plasminogen Activator Administration Management Between Telestroke Network Hospitals and Academic Stroke Centers. *Stroke* 2006 Jul;37(7):1822-1827. [doi: [10.1161/01.str.0000226741.20629.b2](https://doi.org/10.1161/01.str.0000226741.20629.b2)]
18. Audebert HJ, Kukla C, Clarmann von Clarana S, Ku"hn J, Vatankeh B, Schenkel J, et al. Telemedicine for Safe and Extended Use of Thrombolysis in Stroke. *Stroke* 2005 Feb;36(2):287-291. [doi: [10.1161/01.str.0000153015.57892.66](https://doi.org/10.1161/01.str.0000153015.57892.66)]
19. Audebert HJ, Schenkel J, Heuschmann PU, Bogdahn U, Haberl RL. Effects of the implementation of a telemedical stroke network: the Telemed Pilot Project for Integrative Stroke Care (TEMPiS) in Bavaria, Germany. *The Lancet Neurology* 2006 Sep;5(9):742-748. [doi: [10.1016/s1474-4422\(06\)70527-0](https://doi.org/10.1016/s1474-4422(06)70527-0)]
20. May C, Harrison R, Finch T, MacFarlane A, Mair F, Wallace P, Telemedicine Adoption Study Group. Understanding the normalization of telemedicine services through qualitative evaluation. *J Am Med Inform Assoc* 2003;10(6):596-604 [FREE Full text] [doi: [10.1197/jamia.M1145](https://doi.org/10.1197/jamia.M1145)] [Medline: [12925553](#)]
21. Broens THF, Huis in't Veld RMHA, Vollenbroek-Hutten MMR, Hermens HJ, van Halteren AT, Nieuwenhuis LJM. Determinants of successful telemedicine implementations: a literature study. *J Telemed Telecare* 2007;13(6):303-309. [doi: [10.1258/135763307781644951](https://doi.org/10.1258/135763307781644951)] [Medline: [17785027](#)]
22. Dharmar M, Romano PS, Kuppermann N, Nesbitt TS, Cole SL, Andrada ER, et al. Impact of critical care telemedicine consultations on children in rural emergency departments. *Crit Care Med* 2013 Oct;41(10):2388-2395. [doi: [10.1097/CCM.0b013e31828e9824](https://doi.org/10.1097/CCM.0b013e31828e9824)] [Medline: [23921273](#)]
23. Hilty D, Ferrer D, Parish M, Johnston B, Callahan E, Yellowlees P. The effectiveness of telemental health: a 2013 review. *Telemed J E Health* 2013 Jun;19(6):444-454 [FREE Full text] [doi: [10.1089/tmj.2013.0075](https://doi.org/10.1089/tmj.2013.0075)] [Medline: [23697504](#)]
24. Sanders C, Rogers A, Bowen R, Bower P, Hirani S, Cartwright M, et al. Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. *BMC Health Serv Res* 2012 Jul 26;12:220 [FREE Full text] [doi: [10.1186/1472-6963-12-220](https://doi.org/10.1186/1472-6963-12-220)] [Medline: [22834978](#)]
25. Wootton R. Twenty years of telemedicine in chronic disease management--an evidence synthesis. *J Telemed Telecare* 2012 Jun;18(4):211-220. [doi: [10.1258/jtt.2012.120219](https://doi.org/10.1258/jtt.2012.120219)] [Medline: [22674020](#)]
26. Wade VA, Karnon J, Elshaug AG, Hiller JE. A systematic review of economic analyses of telehealth services using real time video communication. *BMC Health Serv Res* 2010 Aug 10;10:233. [doi: [10.1186/1472-6963-10-233](https://doi.org/10.1186/1472-6963-10-233)] [Medline: [20696073](#)]
27. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. *N Engl J Med* 2010 Dec 09;363(24):2301-2309 [FREE Full text] [doi: [10.1056/NEJMoa1010029](https://doi.org/10.1056/NEJMoa1010029)] [Medline: [21080835](#)]
28. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. *Int J Med Inform* 2010 Nov;79(11):736-771. [doi: [10.1016/j.ijmedinf.2010.08.006](https://doi.org/10.1016/j.ijmedinf.2010.08.006)] [Medline: [20884286](#)]
29. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2015 Sep 07;2015(9):CD002098 [FREE Full text] [doi: [10.1002/14651858.CD002098.pub2](https://doi.org/10.1002/14651858.CD002098.pub2)] [Medline: [26343551](#)]

30. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. *J Med Internet Res* 2014 Apr 16;16(4):e110 [FREE Full text] [doi: [10.2196/jmir.2790](https://doi.org/10.2196/jmir.2790)] [Medline: [24739471](https://pubmed.ncbi.nlm.nih.gov/24739471/)]
31. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. *World J Psychiatry* 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: [10.5498/wjp.v6.i2.269](https://doi.org/10.5498/wjp.v6.i2.269)] [Medline: [27354970](https://pubmed.ncbi.nlm.nih.gov/27354970/)]
32. Bashshur RL, Shannon GW, Smith BR, Alverson DC, Antoniotti N, Barsan WG, et al. The Empirical Foundations of Telemedicine Interventions for Chronic Disease Management. *Telemedicine and e-Health* 2014 Sep;20(9):769-800. [doi: [10.1089/tmj.2014.9981](https://doi.org/10.1089/tmj.2014.9981)]
33. Fierson WM, Capone A, American Academy of Pediatrics Section on Ophthalmology, American Academy of Ophthalmology, American Association of Certified Orthoptists. Telemedicine for evaluation of retinopathy of prematurity. *Pediatrics* 2015 Jan;135(1):e238-e254. [doi: [10.1542/peds.2014-0978](https://doi.org/10.1542/peds.2014-0978)] [Medline: [25548330](https://pubmed.ncbi.nlm.nih.gov/25548330/)]
34. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. *Health Aff (Millwood)* 2014 Feb;33(2):194-199. [doi: [10.1377/hlthaff.2013.0992](https://doi.org/10.1377/hlthaff.2013.0992)] [Medline: [24493760](https://pubmed.ncbi.nlm.nih.gov/24493760/)]
35. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: MAST. *Int J Technol Assess Health Care* 2012 Jan 23;28(1):44-51. [doi: [10.1017/s0266462311000638](https://doi.org/10.1017/s0266462311000638)]
36. Ashwood JS, Mehrotra A, Cowling D, Uscher-Pines L. Direct-To-Consumer Telehealth May Increase Access To Care But Does Not Decrease Spending. *Health Aff (Millwood)* 2017 Mar 01;36(3):485-491. [doi: [10.1377/hlthaff.2016.1130](https://doi.org/10.1377/hlthaff.2016.1130)] [Medline: [28264950](https://pubmed.ncbi.nlm.nih.gov/28264950/)]
37. Şenel E, Demir E. A global productivity and bibliometric analysis of telemedicine and teledermatology publication trends during 1980–2013. *Dermatologica Sinica* 2015 Mar;33(1):16-20. [doi: [10.1016/j.dsi.2014.10.003](https://doi.org/10.1016/j.dsi.2014.10.003)]
38. Ohannessian R, Yaghobian S, Chaleuil M, Salles N. Telemedicine in France: A review of registered clinical trials from 2000 to 2015. *European Research in Telemedicine / La Recherche Européenne en Télémédecine* 2016 Jun;5(2):29-36. [doi: [10.1016/j.eurtel.2016.04.001](https://doi.org/10.1016/j.eurtel.2016.04.001)]
39. Maia JX, de Sousa LAP, Marcolino MS, Cardoso CS, da Silva JLP, Alkmim MBM, et al. The Impact of a Clinical Decision Support System in Diabetes Primary Care Patients in a Developing Country. *Diabetes Technol Ther* 2016 Apr;18(4):258-263. [doi: [10.1089/dia.2015.0253](https://doi.org/10.1089/dia.2015.0253)] [Medline: [26840128](https://pubmed.ncbi.nlm.nih.gov/26840128/)]
40. Bagayoko CO, Traoré D, Thevoz L, Diabaté S, Pecoul D, Niang M, et al. Medical and economic benefits of telehealth in low- and middle-income countries: results of a study in four district hospitals in Mali. *BMC Health Serv Res* 2014 May 12;14(Suppl 1):S9 [FREE Full text] [doi: [10.1186/1472-6963-14-s1-s9](https://doi.org/10.1186/1472-6963-14-s1-s9)]
41. Permanent Workgroup Telemedicine and Telehealth of the Comunidade dos Países de Língua Portuguesa. CPLP. URL: [https://www.cplp.org/Admin/Public/DWSDownload.aspx?File=%2FFiles%2FFiler%2F1\\_CPLP%2FSAude%2FPECS%2FPECS-2018\\_2021\\_APPROVED\\_april-2018\\_English-Version.pdf](https://www.cplp.org/Admin/Public/DWSDownload.aspx?File=%2FFiles%2FFiler%2F1_CPLP%2FSAude%2FPECS%2FPECS-2018_2021_APPROVED_april-2018_English-Version.pdf)
42. Armfield NR, Edirippulige S, Caffery LJ, Bradford NK, Grey JW, Smith AC. Telemedicine--a bibliometric and content analysis of 17,932 publication records. *Int J Med Inform* 2014 Oct;83(10):715-725. [doi: [10.1016/j.ijmedinf.2014.07.001](https://doi.org/10.1016/j.ijmedinf.2014.07.001)] [Medline: [25066950](https://pubmed.ncbi.nlm.nih.gov/25066950/)]
43. Correia J, Lapao LV, Mingas R, Augusto H, Balo M, Maia M. Implementation of a telemedicine network in Angola: challenges and opportunities. *J Health Inform Dev Countries* 2018;12(1):171 [FREE Full text]
44. Albert Henry T. After COVID-19, \$250 billion in care could shift to telehealth. American Medical Association. URL: <https://www.ama-assn.org/practice-management/digital/after-covid-19-250-billion-care-could-shift-telehealth> [accessed 2020-09-17]

## Abbreviations

**COPD:** chronic obstructive pulmonary disease  
**ICT:** information and communication technology  
**ICU:** intensive care unit  
**LLR:** log likelihood ratio  
**LMIC:** low- and middle-income country  
**mHealth:** mobile health  
**PTSD:** posttraumatic stress disorder  
**RAFT:** Réseau en Afrique Francophone pour la Télémédecine  
**TF-IDF:** term frequency-inverse document frequency  
**WHO:** World Health Organization  
**WOS:** Web of Science

*Edited by G Eysenbach; submitted 22.03.20; peer-reviewed by N Piland, DM Kulhandjian; comments to author 28.06.20; accepted 26.07.20; published 02.10.20.*

*Please cite as:*

Waqas A, Teoh SH, Lapão LV, Messina LA, Correia JC

*Harnessing Telemedicine for the Provision of Health Care: Bibliometric and Scientometric Analysis*

*J Med Internet Res* 2020;22(10):e18835

URL: <https://www.jmir.org/2020/10/e18835>

doi: [10.2196/18835](https://doi.org/10.2196/18835)

PMID: [33006571](https://pubmed.ncbi.nlm.nih.gov/33006571/)

©Ahmed Waqas, Soo Huat Teoh, Luís Velez Lapão, Luiz Ary Messina, Jorge César Correia. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 02.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Evaluation of a Heart Failure Telemonitoring Program Through a Microsimulation Model: Cost-Utility Analysis

Chris Boodoo<sup>1</sup>, BSc, MSc; Qi Zhang<sup>1</sup>, BSc, MA; Heather J Ross<sup>2,3,4</sup>, MHS, MD; Ana Carolina Alba<sup>2,3,4</sup>, PhD, MD; Audrey Laporte<sup>1,5</sup>, MA, PhD; Emily Seto<sup>1,6</sup>, PEng, PhD

<sup>1</sup>Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

<sup>2</sup>Ted Rogers Centre for Heart Research, University Health Network, Toronto, ON, Canada

<sup>3</sup>Department of Medicine, University of Toronto, Toronto, ON, Canada

<sup>4</sup>Peter Munk Cardiac Centre, University Health Network, Toronto, ON, Canada

<sup>5</sup>Canadian Centre for Health Economics, Toronto, ON, Canada

<sup>6</sup>Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

**Corresponding Author:**

Chris Boodoo, BSc, MSc

Institute of Health Policy, Management and Evaluation

Dalla Lana School of Public Health

University of Toronto

HS Building, 4th Floor

Toronto, ON, M5T 3M6

Canada

Phone: 1 (416) 978 4326

Email: [c2boodoo@gmail.com](mailto:c2boodoo@gmail.com)

## Abstract

**Background:** Heart failure (HF) is a major public health issue in Canada that is associated with high prevalence, morbidity, and mortality rates and high financial and social burdens. Telemonitoring (TM) has been shown to improve all-cause mortality and hospitalization rates in patients with HF. The *Medly* program is a TM intervention integrated as standard of care at a large Canadian academic hospital for ambulatory patients with HF that has been found to improve patient outcomes. However, the cost-effectiveness of the *Medly* program is yet to be determined.

**Objective:** This study aims to conduct a cost-utility analysis of the *Medly* program compared with the standard of care for HF in Ontario, Canada, from the perspective of the public health care payer.

**Methods:** Using a microsimulation model, individual patient data were simulated over a 25-year time horizon to compare the costs and quality-adjusted life years (QALYs) between the *Medly* program and standard care for patients with HF treated in the ambulatory care setting. Data were sourced from a *Medly* Program Evaluation study and literature to inform model parameters, such as *Medly*'s effectiveness in reducing mortality and hospitalizations, health care and intervention costs, and model transition probabilities. Scenario analyses were conducted in relation to HF severity and TM deployment models. One-way deterministic effectiveness analysis and probabilistic sensitivity analysis were performed to explore the impact on the results of uncertainty in model parameters.

**Results:** The *Medly* program was associated with an average total cost of Can \$102,508 (US \$77,626) per patient and total QALYs of 5.51 per patient compared with the average cost of Can \$97,497 (US \$73,831) and QALYs of 4.95 per patient in the Standard Care Group. This led to an incremental cost of Can \$5011 (US \$3794) and incremental QALY of 0.566, resulting in an incremental cost-effectiveness ratio of Can \$8850 (US \$6701)/QALY. Cost-effectiveness improved in relation to patients with advanced HF and with deployment models in which patients used their own equipment. Baseline and alternative scenarios consistently showed probabilities of cost-effectiveness greater than 85% at a willingness-to-pay threshold of Can \$50,000 (US \$37,718). Although the results showed some sensitivity to assumptions about effectiveness parameters, the intervention was found to remain cost-effective.

**Conclusions:** The *Medly* program for patients with HF is cost-effective compared with standard care using commonly reported willingness-to-pay thresholds. This study provides evidence for decision makers on the use of TM for HF, supports the use of a

nurse-led model of TM that embeds clinically validated algorithms, and informs the use of economic modeling for future evaluations of early-stage health informatics technology.

(*J Med Internet Res* 2020;22(10):e18917) doi:[10.2196/18917](https://doi.org/10.2196/18917)

## KEYWORDS

cost utility analysis; cost effectiveness; telemedicine; heart failure; microsimulation; mobile phone

## Introduction

### Background

Heart failure (HF) is a major public health issue with a worldwide prevalence of 26 million and 669,600 in Canada [1,2]. Half of those with a diagnosis of HF will die within 5 years, and up to 80% will die within 10 years [3-5]. Furthermore, flare-ups of HF symptoms occur often and can result in frequent hospitalizations, with more than 50% of individuals being readmitted within 6 months of discharge [6,7]. Reasons for rehospitalizations include incomplete treatment in hospitals, poor coordination of services or poor communication of care plans at discharge, inadequate access to services, poor patient education, failure to optimize therapies, and lack of long-term monitoring for early signs of worsening health [7]. HF-associated hospital admissions have been estimated to cost the Canadian health care system Can \$482 (US \$364) million in 2013, and the cost is expected to increase to Can \$720 (US \$543.14) million by 2030 [8].

It has been recommended that disease management interventions that enable patient empowerment, education, and clinical follow-up should be integrated within the system of care for patients with HF because these interventions have been associated with reduced hospitalization rates and improved quality of life (QoL) and survival [9]. In response, the *Medly* telemonitoring (TM) system and program, which was deployed to augment the existing standard of care at the Ted Rogers Centre for Heart Research at the University Health Network (UHN), was designed to shift traditional episodic care of HF to a more continuous paradigm where care is extended into the daily lives of patients rather than confined to health care institutions. This program enables patients to record symptoms and physiological measurements, including weight, blood pressure, and heart rate, which are then transmitted to a registered nurse coordinator who reviews and responds to alerts and serves as the first resource for patients. Overall, meta-analyses have shown that TM programs similar to *Medly* reduce all-cause mortality and hospitalizations when compared with the standard of care without TM [10-14]. However, other studies have also shown null or mixed results for TM [15-18]. Some of this uncertainty in effectiveness can be attributed to the heterogeneity of the studies, such as patient demographics, characteristic differences between the evaluated interventions, and quality of the trial [19].

The decision to implement interventions are often dependent on cost-effectiveness. However, there is a lack of economic evidence for TM stemming from the challenges associated with conducting economic evaluations owing to the heterogeneity and complexity of TM [20,21]. Specifically, diversity stemming from clinical conditions under study, technology, applications,

objectives, and context makes comparisons between telemedicine interventions difficult [21]. That said, a number of studies have been conducted that included evaluations of the financial impact of TM for patients with HF, with many reporting savings. However, most studies did not conduct a full economic evaluation [22-33], such as a cost-effectiveness or cost-utility analysis (CUA) [34]. Studies that included a full economic analysis did not evaluate the long-term effects of TM because time horizons of 18 months or less were used [35-37], or the studies were conducted outside of Canada [38-40]. This is owing to the lack of long-term Canadian economic evaluations of TM interventions for patients with HF, information on the life expectancy of this patient population, the possible fluctuations of health status over time, and the nuances of the Canadian population and health care system.

### Study Objectives

The objective of this study was to evaluate the long-term cost-effectiveness of TM for patients with HF within a Canadian context from a public health care payer perspective, referencing costing data and concepts from a TM program, *Medly*, implemented at a large academic hospital in Ontario, and data from the literature. Specifically, the central research question is as follows: What is the cost utility of the *Medly* program for patients with HF compared with the standard of care in Ontario? This question will be explored through the application of a microsimulation model.

## Methods

The methods used in this study follow the Consolidated Health Economic Evaluation Reporting Standards [41].

### Type of Economic Evaluation

A CUA was performed. A CUA was chosen because it allows for the effectiveness outcome to be comparable with that of other disease groups and across interventions, making it the gold standard for economic evaluations. Furthermore, there is utility evidence available for patients with HF, allowing for the use of quality-adjusted life years (QALYs) [34].

### Target Population

The target population was a cohort of ambulatory patients with HF from the UHN in Toronto, Canada, enrolled in the *Medly Program Evaluation* study. Table 1 presents information on the baseline patient characteristics and the missing data. Owing to the limitations associated with manual extraction of data from patients' clinical notes and inconsistent laboratory testing orders, there were considerable missing data. The nature of the missingness was not quantitatively analyzed, and it was assumed that the missing data were randomly distributed. The data reported correspond to the set of variables that were used in the



Seattle Heart Failure Model (SHFM), which is a multivariate Cox hazard model used to predict mortality [42-45]. As of June 31, 2019, 315 patients were enrolled in the program, based on

a joint decision between the patient and the cardiologist at either a follow-up outpatient visit or after an inpatient hospital stay.

**Table 1.** Baseline patient characteristics of the Medly Program Evaluation cohort (n=315; number of patients unless specified otherwise).

Characteristics	Overall	Missing, n
Age (years), mean (SD)	58.23 (15.43)	2
Proportion of females, n (%)	69 (22.0)	2
Proportion of ischemic etiology, n (%)	65 (28.5)	87
Proportion of beta blockers, n (%)	270 (89.4)	13
Proportion of aldosterone blockers, n (%)	215 (71.2)	13
Proportion of ARBs <sup>a</sup> , n (%)	82 (27.2)	13
Proportion of ACE <sup>b</sup> inhibitors, n (%)	137 (45.5)	13
Proportion of allopurinol, n (%)	41 (13.6)	13
Percentage LVEF <sup>c</sup> , mean (SD)	32.07 (13.62)	7
New York Health Association (average class), mean (SD)	2.36 (0.59)	13
Systolic blood pressure (mm Hg), mean (SD)	110.36 (17.91)	53
Percentage of lymphocytes, mean (SD)	22.18 (9.07)	52
Sodium (mEq/L), mean (SD)	137.73 (3.06)	33
Cholesterol (mg/dL), mean (SD)	154.77 (52.71)	83
Hemoglobin (g/dL), mean (SD)	13.33 (1.99)	52
Urate (mg/dL), mean (SD)	7.97 (2.70)	86
Weight (kg), mean (SD)	83.39 (20.04)	44
Furosemide-equivalent dose (mg/day), mean (SD)	99.57 (123.93)	16
Proportion of implantable cardioverter-defibrillators, n (%)	165 (56.5)	23

<sup>a</sup>ARB: angiotensin II receptor blocker.

<sup>b</sup>ACE: angiotensin-converting-enzyme.

<sup>c</sup>LVEF: left ventricular ejection fraction.

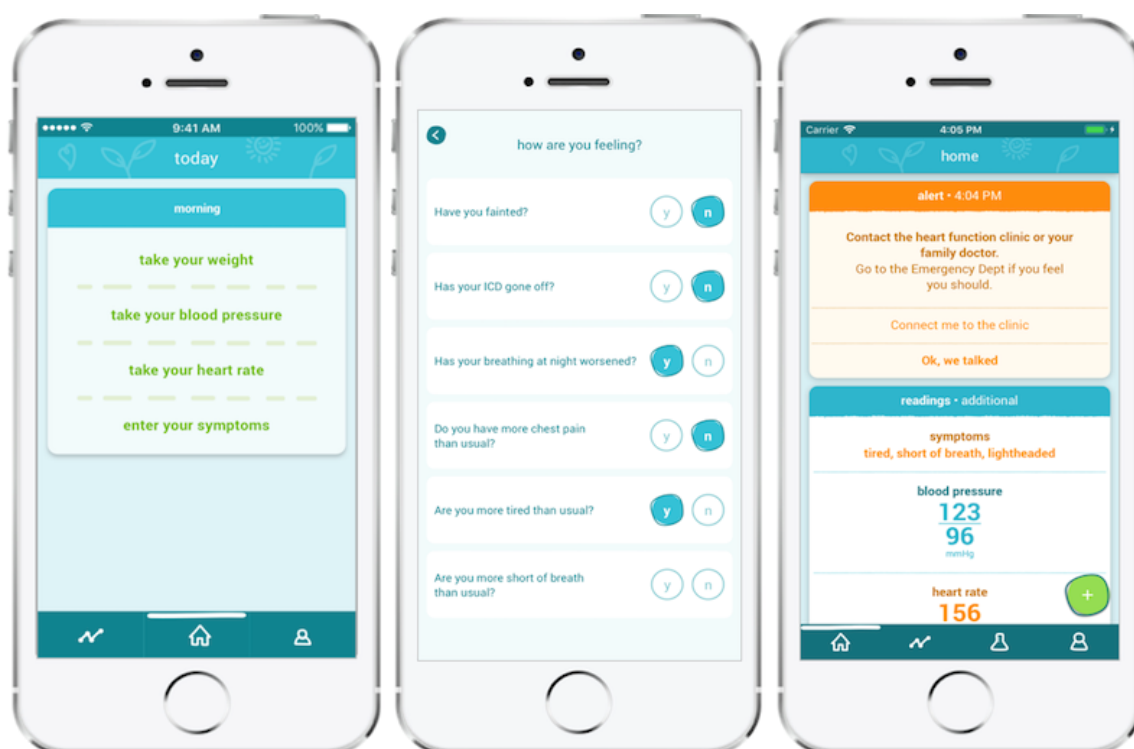
## The Intervention-Medly

In August 2016, the *Medly* program was deployed to augment the existing standard of care at the Ted Rogers Centre for Heart Research at the UHN. Patients are trained to use the technology, and the importance of taking daily readings is emphasized. Patients then use the intervention as long as there are clinical benefits as determined by both the clinician and the patient. The program is led by a registered nurse coordinator who reviews and responds to alerts and serves as the first resource for patients' clinical concerns or technical troubleshooting.

The main component of the program is the *Medly* smartphone app. Patients use the app to record their body weight, blood pressure, and heart rate and to answer a short yes or no questionnaire about their symptoms. Patients are asked to take these readings daily immediately after they wake up. These data are then processed by a clinically validated algorithm to interpret the readings relative to the patient's target thresholds set by the

most responsible HF physician [46]. If the algorithm determines that the recordings are within the target range, patients are presented with a prompt stating their HF is in a stable condition. If the algorithm deems that the readings are outside the target range and/or identify an abnormal trend in weight gain, the patient is prompted with self-care feedback such as taking an additional dose of their diuretic medication and to contact their care provider or to visit the emergency department (ED). [Figure 1](#) shows screenshots of some of the various interfaces with which patients interact. In addition to the self-care messages, the registered nurse coordinator receives the alert via email and triages the event. The nurse also responds to technical troubleshooting. A full-time nurse is projected to be able to manage approximately 500 patients through the *Medly* program. Alerts and all patient TM data can be viewed on the *Medly* clinical dashboard. Other features of the *Medly* app include graphical trends of specific measurements and an automated phone call to remind patients to take their daily measurements if it is past 10 AM (can be disabled per patient's request).

**Figure 1.** Medly app showing instructions for required readings (screen 1), the symptoms questionnaire (screen 2), and personalized self-care feedback (screen 3).



## Key Data Source

This study referenced the ongoing *Medly* program evaluation, which included multiple pre- and posttest analyses on patient-level impacts, patient adherence, and cost. Quantitative data analyses leverage data that were collected at the 6-month follow-up as a part of the standard of care, such as health care utilization and laboratory results from electronic patient records, while also using data from the TM system. The results of this study have been published in the study by Ware et al (2020) [47], and further details on the study can also be found therein.

## Comparators

In this analysis, the intervention group was a cohort of patients with HF enrolled in the *Medly* program. The comparator group consisted of patients with HF who received standard care, which does not include the use of TM. It was assumed that standard care was conducted according to the typical care practices in Ontario, which involves specialized multidisciplinary HF clinics, although care models may vary among clinics [48].

## Perspective

This analysis was conducted from the perspective of the public payer, namely the Ontario Ministry of Health because *Medly* is currently implemented in a publicly funded health care system.

## Time Horizon and Discounting

A time horizon of 25 years was adopted to determine the long-term cost and outcomes associated with *Medly* for the patient population with HF. Costs and outcomes were discounted

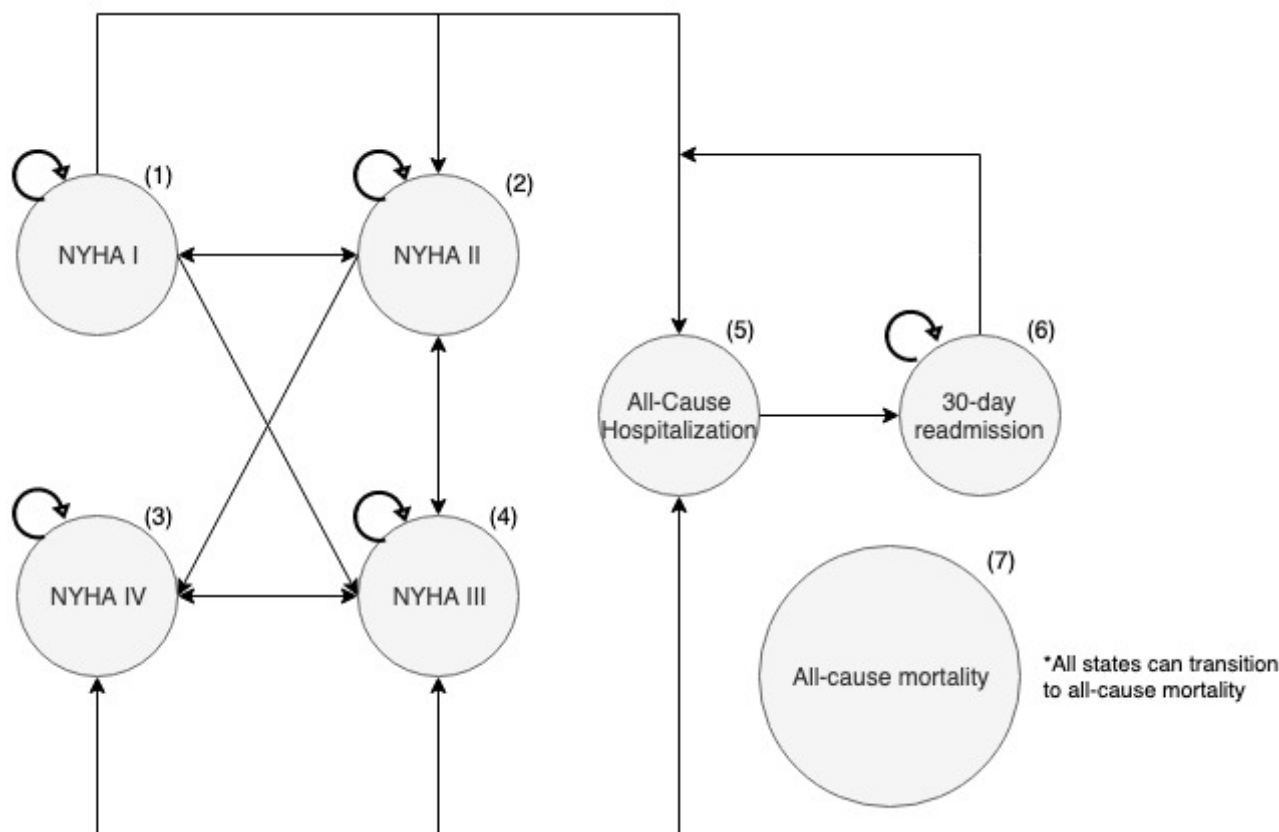
at an annual rate of 1.5%, as recommended by the Canadian Agency for Drugs and Technologies in Health [49].

## Model Framework, Conceptualization, and Technique

All analysis and model construction were conducted using RStudio [50]. The model consists of 7 mutually exclusive states: 4 states specific to HF severity, 2 states for hospitalization events, and 1 absorbing state for death. Patients with HF can alternate between a state of decompensation (or symptom exacerbation or functional capacity impairment) and a state of clinical stability. To capture this, the model was stratified by, and allowed for, transitions between New York Health Association (NYHA) functional classes. The NYHA functional classification is a common measure used by clinicians to classify the severity of symptoms in patients with HF, in which a higher class indicates worse health [51]. Hospitalizations mark a fundamental change in the natural history of HF with subsequent increased rehospitalizations and higher mortality rates as the patient's disease progresses [52,53]. A cycle length of 1 month was chosen to account for 30-day readmission rates common in the HF population [54]. As recommended by Naimark et al [55] for models that are relatively simple and have a cycle length of a month or less, a half-cycle correction was omitted.

The modeling technique chosen was a patient-level state-transition model, also known as a first-order Monte Carlo microsimulation. This model was appropriate because it can capture patient heterogeneity that is common in patients with HF and is also the preferred option for modeling chronic disease [56]. Figure 2 shows the conceptualization of the Markov model that was developed to represent an individual's progression of HF.

**Figure 2.** Conceptual representation of the microsimulation model structure. States 1 to 4 represent the transitions between New York Health Association classes. States 5 and 6 show transitions into and between hospitalization states. State 7 is an absorbing state representing death, where all states can transition to. NYHA: New York Health Association.



### Parameter Estimates

The values used in the model were based on a literature review. The values inputted into the model are conditional on patient characteristics. Patients with a more limited functional capacity by NYHA class have a higher risk of hospitalization [57-61]. In addition, the risk for readmission is highest within 30 days of hospital discharge [52]. As a patient's NYHA functional class can change over time, the probability of transitioning between classes was derived from the large-scale international SENIORS study [62,63]. All-cause mortality during hospitalization was based on the study by Yeung et al (2012) [64]. The SHFM was used to derive a survival curve for each patient over their lifetime, which is a multivariate Cox hazard model that has been validated on multiple cohorts of patients with HF [42-45].

### Generating Virtual Patient Profiles

To generate virtual patient profiles, a Cholesky decomposition was performed on a correlation matrix that describes the interdependence between patient characteristics [56]. The matrix was derived from a consolidated representative sample of 7125 patients with HF from the University of Washington, Prospective Randomized Amlodipine Survival Evaluation, Valsartan Heart Failure Trial, and Italian Heart Failure Registry [65] (Multimedia Appendix 1). Values for each patient characteristic were sampled from a multivariate normal distribution, defined by the *Medly Program Evaluation* cohort's mean and SD in Table 1 using the R-package "PoisBinOrdNonNor" [66,67].

Patient characteristics included clinical, pharmacological, device, and laboratory data, based on the SHFM developed by Levy et al (2006) [43].

### Effectiveness

The 2 primary outcomes that TM interventions for patients with HF aim to improve are all-cause mortality and all-cause hospitalization rates. The risk of all-cause hospitalization was based on evidence from the *Medly* program evaluation, which reported a relative risk (RR) of 0.753 (95% CI 0.634-0.879) for patients using *Medly* (Multimedia Appendix 1; All-cause hospitalization). Owing to the lack of an interdependent comparative group and small sample size of the *Medly* program evaluation, it was not possible to evaluate its effectiveness in reducing mortality. Therefore, the estimate was based on a meta-analysis of the effectiveness of TM for patients with HF by Yun et al (2018) [11]. This meta-analysis only included randomized controlled trials that defined TM as the transmission of biological information, such as body weight, heart rate, and blood pressure via telecommunication technologies. Owing to this strict inclusion criteria, it was deemed comparable evidence for the expected benefits that *Medly* users could experience. It was reported that TM users had an RR of 0.81 (0.70-0.94) for all-cause mortality compared with 416 of 3724 patients in the TM group to 483 of 3733 patients in the control group [11]. The study follow-up periods ranged from 3 months to 15 months, with 1 study having a 4 year follow-up. The meta-analysis also presented a subgroup analysis of only asynchronous interventions (ie, removal of real-time and

teleconferencing technologies), where an RR of 0.79 (0.66-0.94) was reported for all-cause mortality, which is similar to the primary analysis.

### Measurement and Evaluation of Health

Each state in the model has an associated utility value between 0 and 1. Utility values for each health state were derived on the basis of values from other health economic evaluations of HF

interventions. NYHA classes are commonly used to categorize patients with HF based on severity of symptoms, and studies have estimated utility values for each class [60,62,68,69]. All utilities and the sources of the values are presented in Table 2. To adjust for the decrease in the QoL patients experience when hospitalized [70], the patient utility value in the model was decreased by 0.059 in the hospitalization state, consistent with Sandhu et al (2016) [71].

**Table 2.** Model parameters conditional on New York Health Association (NYHA) class including living with heart failure costs (in Can currency), utilities, probability of hospitalization, and transitions between NYHA classes.

Description	NYHA <sup>a</sup> I	NYHA II	NYHA III	NYHA IV <sup>b</sup>	Source	Distribution
<b>Health care costs</b>						
Emergency department costs, Can \$ (US \$)	0.00	0.00	62.83 (47.20)	62.83 (47.20)	Medly Program Evaluation	Gamma
General practitioner visit costs, Can \$ (US \$)	0.00	0.00	12.87 (9.67)	12.87 (9.67)	Medly Program Evaluation	Fixed
Drug costs (only if patient age is ≥65), Can \$ (US \$)	52.00 (39.06)	52.00 (39.06)	79.43 (59.67)	208.16 (156.38)	Kaul et al (2011) [72]	Gamma
Outpatient costs, Can \$ (US \$)	97.00 (72.87)	97.00 (72.87)	97.00 (72.87)	97.00 (72.87)	Medly Program Evaluation	Gamma
Total monthly cost of living with heart failure, Can \$ (US \$)	149.00 (111.94)	149.00 (111.94)	252.13 (189.42)	380.86 (286.14)	OCCI <sup>c</sup> [73], SOB <sup>d</sup> [74], Kaul et al (2011) [72]	N/A <sup>e</sup>
<b>Utilities (range)</b>						
Living with heart failure	0.81 (0.81-0.90)	0.72 (0.72-0.83)	0.59 (0.59-0.74)	0.508 (0.508-0.59)	Yao et al (2008) [62]	Beta
Probability of all-cause hospitalization	0.0152 (0.008-0.023)	0.024 (0.012-0.036)	0.024 (0.012-0.036)	0.154 (0.077-0.231)	Ford et al (2012) [68], Borisenko et al (2015) [69]	Beta
<b>Transition probabilities between NYHA classes (probability)</b>						
NYHA I	0.977	0.019	0.004	0	Flather et al (2005) [63], Yao et al (2008) [62]	Dirichlet
NYHA II	0.008	0.981	0.01	0.001	Flather et al (2005) [63], Yao et al (2008) [62]	Dirichlet
NYHA III	0	0.034	0.96	0.006	Flather et al (2005) [63], Yao et al (2008) [62]	Dirichlet
NYHA IV	0	0	0.055	0.945	Flather et al (2005) [63], Yao et al (2008) [62]	Dirichlet

<sup>a</sup>NYHA: New York Health Association.

<sup>b</sup>NYHA IV assumed same as NYHA III, except drug cost.

<sup>c</sup>OCCI: Ontario Case Costing Initiative.

<sup>d</sup>SOB: Schedule of Benefits for Physician Services.

<sup>e</sup>N/A: not applicable.

### Resource Use and Costs

Health care utilization was based on data from the *Medly* program evaluation, and hospitalization and ED visit costs were derived from the Ontario Case Costing Initiative (OCCI) report for 2017-2018 using diagnosis codes I500, I509, and I501 [73]. The unit cost per outpatient visit was based on a paper outlining health care utilization for patients with HF over the last 6 months of their lives by Kaul et al (2006) [72] in a comparable health care system in Alberta. This was based on the provincial ambulatory care case mix group, which captures direct and

indirect functional center costs [75]. Physician fees for general practitioner (GP) visits were based on billing code A005 in Ontario's Schedule of Benefits [74]. The unit costs were multiplied by utilization data from the *Medly* program evaluation to calculate the monthly costs (Table 3). Median values for utilization were used because the distribution of health care utilization is typically left-skewed [76]. Monthly drug costs were incurred on the part of the public payer under the Ontario Drug Benefit program for patients aged 65 years and older. The

monthly drug costs were calculated on the basis of the costs reported in Kaul et al (2006).

Of note, the number of outpatient visits for both the intervention and comparator-simulated cohorts was limited to those that occurred at UHN because information outside of UHN's services

was unavailable at the time of the study. Furthermore, self-reported ED visits were used because UHN patient records may underreport the true number of ED visits as patients may live at some distance from UHN and may instead visit a community hospital for an emergency. Self-reported GP visits were used because UHN data do not record this information.

**Table 3.** Median health care utilization over 6 months before using Medly, unit costs per service (in Can currency), and associated distribution stratified by New York Health Association (NYHA) classes. N is the number of patients in each NYHA class.

Type of resource	Unit cost, mean (SD), Can \$	Source for unit cost	NYHA I health care utilization (n=44)	NYHA II health care utilization (n=166)	NYHA III health care utilization (n=93)	NYHA IV health care utilization (n=1)	Distribution
Emergency department (self-reported)	377.00 (374.00)	Ontario Case Costing Initiative	0	0	1	— <sup>a</sup>	Negative binomial
Outpatient visit	291.33 (161.11)	Kaul et al (2011) [72]	2	1	2	—	Negative binomial
General practitioner visit (self-reported)	77.20	Schedule of Benefits	0	1	2	—	Negative binomial
Drug costs over 6 months	1248.96 (2233.52)	Kaul et al (2011) [72]	—	—	—	—	Gamma

<sup>a</sup>No data available.

## Medly Costs and Deployment Models

Costs related to implementation and maintenance of *Medly* were provided by the *Medly* project management and development team (Table 4). The fixed costs associated with implementation were based on a system that delivers care to 1000 patients. The operational cost per patient included costs associated with asset management (technical and application support) and on-site frontline support for patients and clinicians, which was delivered via 2 registered nurse coordinators hired by the *Medly* program. Two registered nurses were included according to the *Medly* project management team's cost projections for 1000 patients. This is similar to what was reported by Ware et al (2020) [47], where approximately 300 patients were managed by 1 registered nurse coordinator. The variable cost per patient included the cost of the device and equipment, depending on the equipment that was loaned to the patient. The costs of the device and the equipment were based on a rental model. These costs were

adjusted for monthly costs according to the cycle length of the model.

The variable cost was based on a mix of models where users can fall into 1 of the 3 categories: Full Kit (FK), Bring Your Own Phone (BYOP), and Bring Your Own Everything (BYOE). An FK user refers to a user who was provided with all necessary equipment for the technology, which is currently funded by the *Medly* program, including a smartphone with a data plan, blood pressure monitor, weight scale, and licensing fee. A BYOP user brings their own phone and pays for their own data plan, but the blood pressure monitor, weight scale, and *licensing fee are provided by the Medly* program. The BYOE user brings their own equipment and is provided with just the licensing fee by the program. The reference case analysis uses a ratio of 2 FK:1 BYOP:2 BYOE, which was based on the number of each category of users in *Medly*'s current implementation. All costs were converted to 2019 Canadian dollars using Statistics Canada's Consumer Price Index to adjust for inflation [77].



**Table 4.** Parameter estimates not conditional on the New York Health Association class including hospitalization costs (in Can currency) and disutility, readmission rates, Medly costs (in Can currency), and Medly effectiveness estimates.

Parameters	Value	Source	Distribution
<b>Costs</b>			
Hospitalization cost per admission (Can \$), mean (SD)	8908 (16,867)	OCCI <sup>a</sup>	Gamma
Hospitalization length of stay days, mean (SD)	5.9 (11.2)	OCCI	Log normal
Medly fixed costs for site implementation	102,500	Medly program	Fixed
Medly operational cost per patient (cost per month), Can \$	44.67	Medly program	Fixed
Medly Full Kit cost per patient (cost per month), Can \$	67.56	Medly program	Fixed
Medly Bring-Your-Own-Phone cost per patient (cost per month), Can \$	18.87	Medly program	Fixed
Medly Bring-Your-Own-Everything cost per patient (cost per month), Can \$	3.80	Medly program	Fixed
<b>Hospitalization, probability</b>			
30-day readmission probability (95% CI)	0.159 (0.089-0.159)	Yeung et al (2012) [64]	Beta
Disutility for hospitalization (95% CI)	0.059 (0-0.11)	Sandhu et al (2015) [71]	Beta
<b>Medly treatment effect, disutility</b>			
RR <sup>b</sup> for hospitalization	0.857 (0.703-1.014)	Medly	Log normal
RR for morality	0.81 (0.70-0.94)	Yun et al (2018) [13]	Log normal

<sup>a</sup>OCCI: Ontario Case Costing Initiative.

<sup>b</sup>RR: relative risk.

## Reference Case Analysis

The expected values for all model parameters were used for the deterministic analysis. The cohort size was assumed to be 1000 patients. Each simulated patient progressed through the model twice until death; once as a patient using *Medly* and again as a patient receiving standard care. Each patient incurred costs and QALYs depending on the health state they were in. Total costs and QALYs were summed for both the *Medly* simulations and standard care simulations. From this, the average incremental cost-effectiveness ratio (ICER) per patient was computed using the following formula:



Monte Carlo standard errors were also reported to show how the results vary owing to patient heterogeneity and randomness introduced from patients transitioning to each state.

A second-order probabilistic analysis was also conducted to characterize the uncertainty in the deterministic results. Each parameter in the model was assigned a distribution based on the nature of the input parameter [56]. R-package “fitdistrplus” was used to fit negative binomial distributions for the health care utilization data from the *Medly* program evaluation via maximum likelihood estimation [78]. Details regarding how distributions were chosen based on Akaike Information Criterion and Bayesian Information Criterion scores are available in [Multimedia Appendix 1](#) ([11,54]; Curve fitting). Values were then randomly selected from the respective distributions and assigned as the input parameter. This process was repeated 1000 times. The results for each iteration were plotted on a cost-utility plane to visualize whether *Medly* was cost-effective, cost-saving,

cheaper, or dominated. The simulations were also plotted onto a cost-effectiveness acceptability curve (CEAC), where the proportion of simulations that resulted in an ICER under a range of willingness-to-pay (WTP) thresholds are plotted. A commonly used WTP threshold is Can \$50,000 (US \$37,718)/QALY [79,80].

## Scenario Analyses

### NYHA Functional Classes

Currently, most patients enrolled in the *Medly* program have HF that corresponds to NYHA functional classes II or III. For this scenario analysis, cohorts were generated for NYHA functional classes I, II, and III and simulated deterministically and probabilistically. NYHA functional class IV was not included because only 1 functional class IV patient was enrolled in the *Medly* program. Average ICERs per NYHA functional class were calculated in both the deterministic and probabilistic models, and CEAC curves were produced on a plot to visualize for which classes *Medly* was most likely to be cost-effective.

### Deployment Model

As mentioned, the *Medly* program currently offers 3 types of kits where the ratio of types of user is 2:1:2 for FK, BYOP, and BYOE, respectively. As the *Medly* program expands, understanding how the ICER changes when costs are shifted from the public dollar to the individual is informative to decision makers. Therefore, this analysis explored various proportions of user types. Specifically, each patient in the reference case cohort was randomly assigned FK, BYOP, and BYOE according to predefined ratios. The ratios of interest were 1:0:0 (100% FK), 1:4:5 (40% BYOP, 50% BYOE), and 0:0:1 (100% BYOE).

These were identified as all FK, mixed deployment, and all BYOE, respectively.

### Effectiveness

One-way deterministic analyses were also conducted to address the uncertainty associated with the estimates for *Medly*'s effectiveness in reducing all-cause hospitalizations (RR 0.63-0.88) and mortality (RR 0.70-0.94). Specifically, 95% CI for each estimate were inputted into the model, and the range of ICERs was presented.

## Results

### Reference Case Analysis

#### Deterministic Results

Over a 25-year time horizon, the average total costs were Can \$97,497 (US \$73,547.84) for the comparator group and Can \$102,508 (US \$77,327.93) for patients using *Medly*. The average total QALYs gained were 4.95 and 5.51 for the comparator group and *Medly* patients, respectively. When comparing the 2 groups, there was an incremental cost of Can \$5011 (US \$3780.10) with an incremental QALY gained of 0.566. This resulted in an ICER of Can \$8850 (US \$6676.09)/QALY (Table 5).

**Table 5.** Deterministic results of the reference case analysis.

Reference	Costs, Can \$ (US \$)	MCSE <sup>a</sup> , Can \$ (US \$)	QALYs <sup>b</sup>	MCSE	Incremental cost, Can \$ (US \$)	MCSE, Can \$ (US \$)	Incremental QALYs	MCSE	ICER <sup>c</sup> , Can \$ (US \$; \$/QALY)
Comparator	97,497 (73,831)	3948 (2989)	4.95	0.12	N/A <sup>d</sup>	N/A	N/A	N/A	N/A
Medly	102,508 (77,626)	3592 (2720)	5.51	0.13	5011 (3794)	2014 (1525)	0.57	0.05	8850 (6701)

<sup>a</sup>MCSE: Monte Carlo standard error.

<sup>b</sup>QALY: quality-adjusted life year.

<sup>c</sup>ICER: incremental cost-effectiveness ratio.

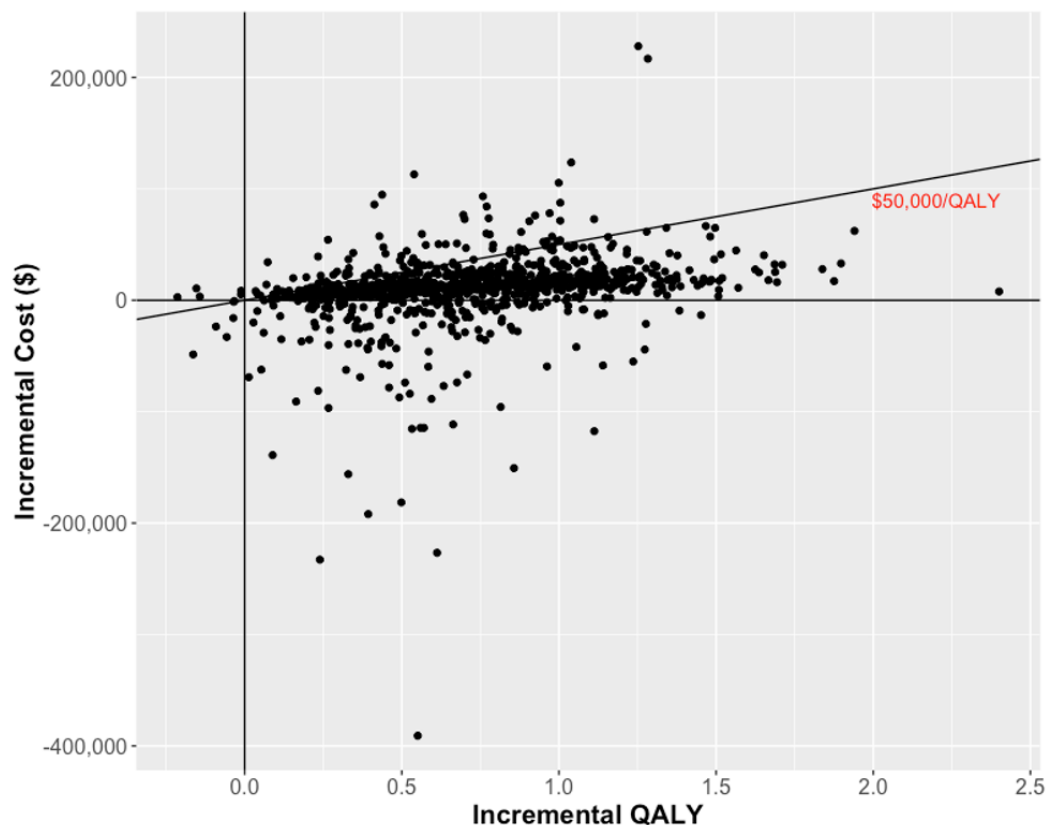
<sup>d</sup>N/A: not applicable.

### Probabilistic Results

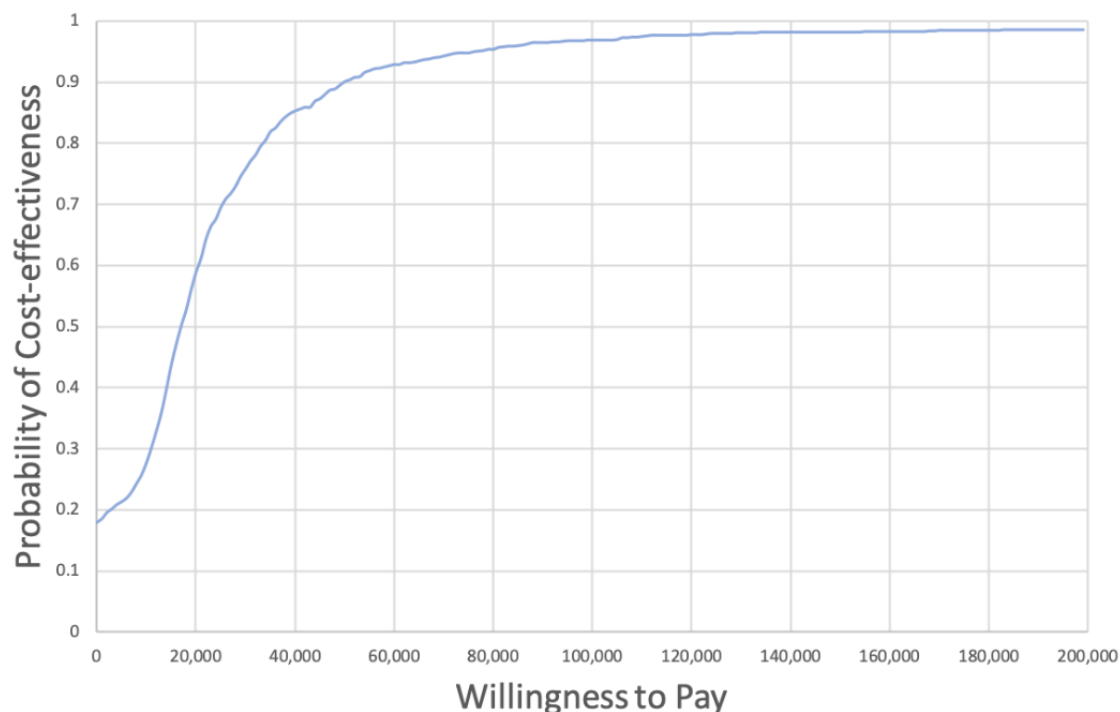
On the basis of 1000 simulations of the reference case scenario in which each parameter was sampled from their respective distribution, 81.6% (816/1000) of the simulations showed that

*Medly* was costlier and more effective, whereas 17.3% (173/1000) showed that *Medly* was less costly and more effective (Figure 3). The CEAC in Figure 4 shows that 90.1% (901/1000) of the simulations resulted in an ICER below the Can \$50,000 (US \$37,718)/QALY threshold.

**Figure 3.** Cost-utility plane of 1000 iterations from the reference case probabilistic analysis. QALY: quality-adjusted life years.



**Figure 4.** Cost-effectiveness acceptability curve of reference case probabilistic analysis.



## Scenario Analyses

### NYHA Functional Classes

Table 6 presents the results of the NYHA functional class scenario analyses. As NYHA functional class increased, the

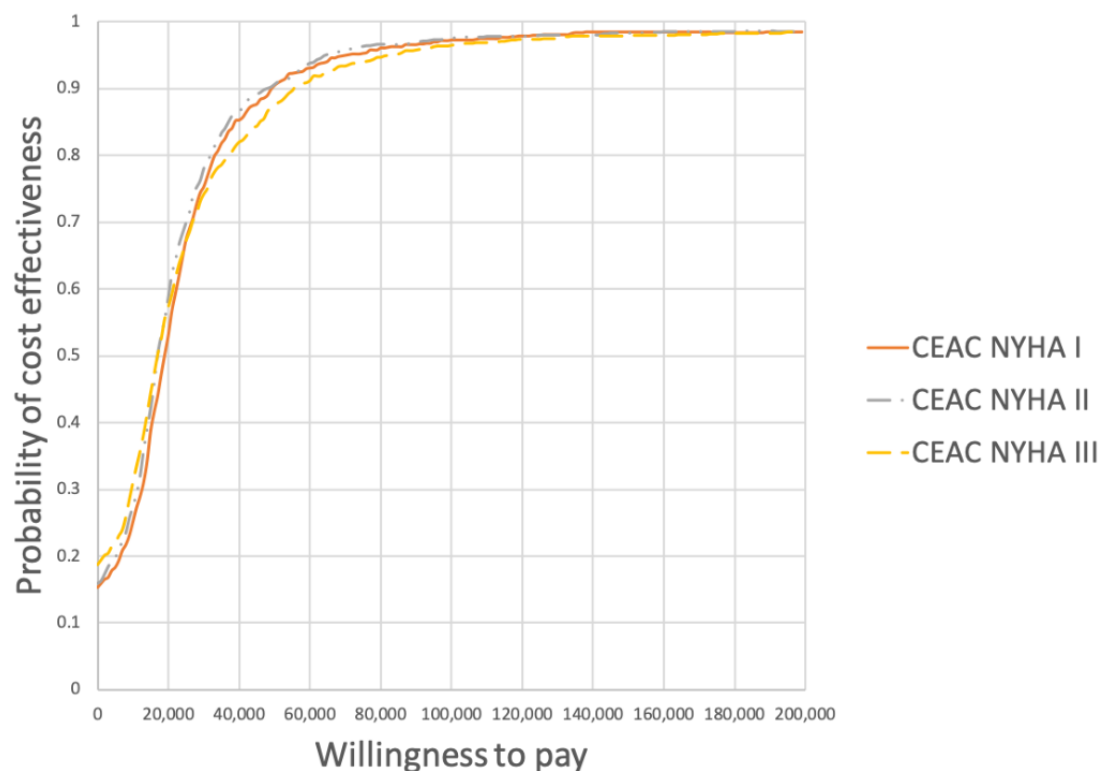
average total costs and incremental costs increased. In addition, as NYHA functional class increased, total QALYs per population decreased. This led to a decreasing trend in ICERs with increasing NYHA functional class.

**Table 6.** Deterministic results for New York Health Association classes I, II, and III.

NYHA <sup>a</sup> classes	Costs, Can \$ (US \$)	MCSE <sup>b</sup> , Can \$ (US \$)	QALYs <sup>c</sup>	MCSE	Incremental cost, Can \$ (US \$)	MCSE, Can \$ (US \$)	Incremental QALYs	MCSE	ICER <sup>d</sup> , Can \$ (US \$; \$/QALY)
<b>NYHA I</b>									
Comparator	81,714 (61,641)	3417 (2577)	6.89	0.13	N/A <sup>e</sup>	N/A	N/A	N/A	N/A
Medly	88,016 (66,395)	3215 (2425)	7.48	0.14	6302 (4753)	1806 (1362)	0.60	0.05	10,567 (7971)
<b>NYHA II</b>									
Comparator	88,405 (66,689)	3821 (2882)	5.65	0.12	N/A	N/A	N/A	N/A	N/A
Medly	94,335 (71,162)	3521 (2656)	6.35	0.13	5930 (4473)	2014 (1519)	0.70	0.06	8510 (6419)
<b>NYHA III</b>									
Comparator	104,421	4356	4.12	0.11	N/A	N/A	N/A	N/A	N/A
Medly	107,803	3929	4.69	0.11	3382	2134	0.57	0.05	5931

<sup>a</sup>NYHA: New York Health Association.<sup>b</sup>MCSE: Monte Carlo standard error.<sup>c</sup>QALY: quality-adjusted life year.<sup>d</sup>ICER: incremental cost-effectiveness ratio.<sup>e</sup>N/A: not applicable.

The CEAC curves for NYHA functional classes I, II, and III are shown in Figure 5. At a WTP threshold of \$50,000 (US \$37,718), the probability of cost-effectiveness for NYHA functional classes I, II, and III was 90.5% (905/1000), 90.6% (906/1000), and 87.5% (875/1000), respectively.

**Figure 5.** Cost-effectiveness acceptability curve of the New York Health Association functional class scenario analyses. CEAC: cost-effectiveness acceptability curve; NYHA: New York Health Association.

## Deployment Models

Table 7 shows the results of the deployment model scenario analyses. As the only difference between scenarios was the total cost incurred by the *Medly* group, all comparator groups had the same average total costs and average total QALYs. The average total QALYs for patients using *Medly* were also the same for all scenarios. Moreover, as expected, when the proportion of FK users increased, so did the average total costs

for patients using *Medly*. This led to ICERs following the same trend.

The CEAC curves for each deployment model are shown in Figure 6. At a WTP threshold of Can \$50,000 (US \$37,718), the probability of cost-effectiveness for all BYOE, mixed model, and all FK was 92.9% (929/1000), 91.7% (917/1000), and 85.4% (854/1000), respectively.

**Table 7.** Deterministic results for each deployment model of *Medly*.

Deployment models	Costs (Can \$)	MCSE <sup>a</sup> (Can \$)	QALYs <sup>b</sup>	MCSE	Incremental cost (Can \$)	MCSE (Can \$)	Incremental QALYs	MCSE	ICER <sup>c</sup> (Can \$/QALY)
<b>All Bring Your Own Entertainment</b>									
Comparator	97,497	3948	4.95	0.12	N/A <sup>d</sup>	N/A	N/A	N/A	N/A
Medly	99,393	3542	5.51	0.13	1896	2006	0.57	0.05	3349
<b>Mixed model</b>									
Comparator	97,497	3948	4.947	0.12	N/A	N/A	N/A	N/A	N/A
Medly	100,769	3567	5.51	0.13	3273	2007	0.57	0.05	5780
<b>All Full Kit</b>									
Comparator	97,497	3948	4.947	0.12	N/A	N/A	N/A	N/A	N/A
Medly	106,194	3646	5.51	0.13	8697	2015	0.57	0.05	15,362

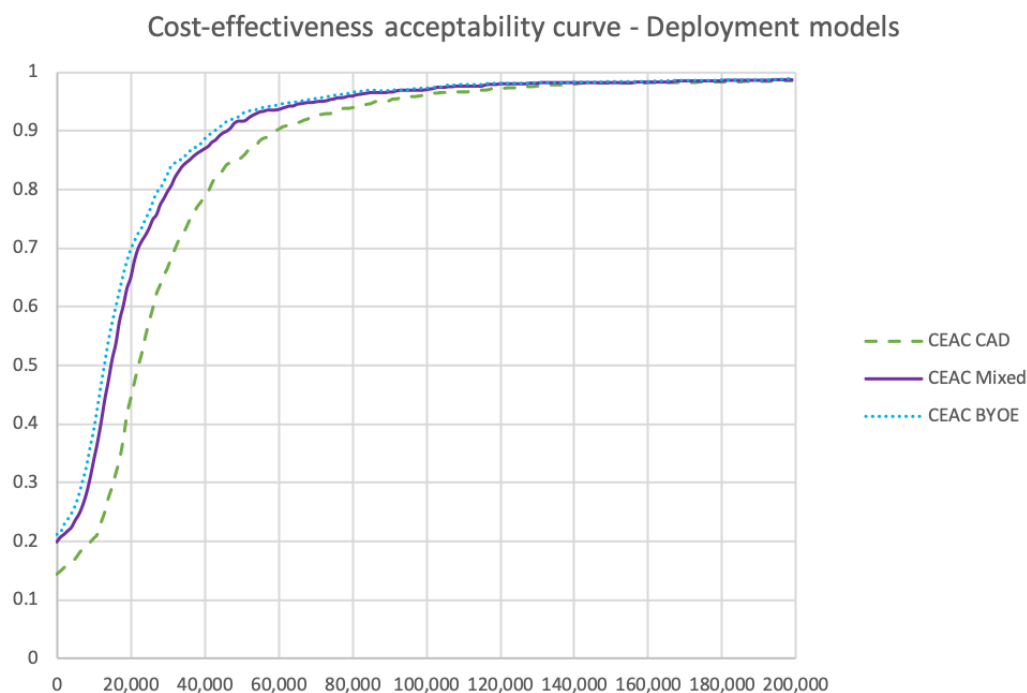
<sup>a</sup>MCSE: Monte Carlo standard error.

<sup>b</sup>QALY: quality-adjusted life year.

<sup>c</sup>ICER: incremental cost-effectiveness ratio.

<sup>d</sup>N/A: not applicable.

**Figure 6.** Cost-effectiveness acceptability curves for each deployment model in the scenario analyses. BYOE: Bring Your Own Everything; CEAC: cost-effectiveness acceptability curve; FK: Full Kit.





## Effectiveness Uncertainty

When the RR for mortality was set to its lower range, the ICER increased to Can \$18,556 (US \$13,997.90)/QALY (Table 8). When the RR for mortality was set to its upper range, *Medly*

became dominant (Table 8). When the RR for hospitalization was set to its lower range, *Medly* became dominant. When the RR for hospitalization was set to its upper range, the ICER increased to Can \$29,240 (US \$22,057.49)/QALY.

**Table 8.** Deterministic results for the upper and lower limits of effectiveness in reducing mortality and hospitalization rates.

Effectiveness	Costs (Can \$)	MCSE <sup>a</sup> (Can \$)	QALYs <sup>b</sup>	MCSE	Incremental cost (Can \$)	MCSE (Can \$)	Incremental QALYs	MCSE	ICER <sup>c</sup> (Can \$/QALY)
<b>RR<sup>d</sup> for mortality</b>									
<b>RR=0.70</b>									
Comparator	97,497	3948	4.95	0.12	— <sup>e</sup>	—	—	—	—
Medly	114,682	3995	5.87	0.13	17,186	2660	0.93	0.07	18,556
<b>RR=0.94</b>									
Comparator	97,497	3948	4.95	0.12	—	—	—	—	—
Medly	91,542	3410	5.14	0.12	−5955	1339	0.19	0.03	−30,806
<b>RR for hospitalization</b>									
<b>RR=0.63</b>									
Comparator	97,497	3948	4.95	0.12	—	—	—	—	—
Medly	92,107	3176	5.55	0.13	−5390	2126	0.61	0.06	−8895
<b>RR=0.88</b>									
Comparator	97,497	3948	4.95	0.12	—	—	—	—	—
Medly	113,763	4052	5.50	0.13	16,267	2104	0.56	0.028	29,240

<sup>a</sup>MCSE: Monte Carlo standard error.

<sup>b</sup>QALY: quality-adjusted life year.

<sup>c</sup>ICER: incremental cost-effectiveness ratio.

<sup>d</sup>RR: relative risk.

<sup>e</sup>N/A: not applicable.

## Discussion

### Principal Findings

The purpose of this study was to assess the cost utility of the *Medly* program for patients with HF compared with the standard of care from a public payer perspective. The analysis showed that *Medly* had a high probability (90.1%) of being cost-effective at a WTP threshold of Can \$50,000 (US \$37,718)/QALY. The results also showed that cost-effectiveness improved in cohorts with more advanced HF. This is attributable to the higher health care utilization rates experienced in higher NYHA functional classes. Deployment models with larger proportions of patients bringing their own equipment to the *Medly* program were also shown to be more cost-effective owing to lower costs incurred by the program itself. Furthermore, the model itself was sensitive to the effectiveness parameters that informed the magnitude of the decrease in all-cause hospitalizations and mortality. However, the results of these analyses showed that *Medly* remains cost-effective even in scenarios with smaller clinical benefit, assuming a WTP threshold of Can \$50,000 (US \$37,718)/QALY [79,80].

### Study Implications

The significance of the study findings are 3-fold: (1) providing evidence for health care decision makers on the use of TM for HF, (2) supporting the use of a nurse-led model of TM using clinically validated algorithms within HF clinics, and (3) informing the use of economic modeling for future evaluation of early-stage health informatics technology.

Our study provided the *Medly* program with its first evaluation, where an economic perspective was considered. This added to the growing body of evidence associated with the program's value not only for patients and health care professionals but also for the health care system. As discussions about implementing the *Medly* program at other sites in Ontario continue among decision makers and stakeholders, this study directly contributes to their understanding of *Medly*'s cost-effectiveness. It enables a new perspective on the upfront costs involved with implementing the TM infrastructure and purchasing necessary equipment, relative to the total costs a patient with HF incurs over a lifetime. Such evidence alleviates some of the uncertainty around the risks in introducing a new model of care for patients with HF.

A key factor contributing to the cost-effectiveness of *Medly* was the high number of patients (500) that a single nurse could

manage, which is possible owing to the clinically validated algorithms that generate automatic clinical alerts and self-care messages. Other studies have reported a concern regarding increased clinical workload associated with incompatibility of the TM program with existing workflows, including management of and responding to alerts [81-83]. To mitigate the increased physician workload, the *Medly* program relies on a registered nurse coordinator and a rule-based algorithm [46] as a patient's first point of contact, decreasing dependency on cardiologists. The registered nurse had the necessary skills to manage patient concerns and involved cardiologists as required. This nurse-led strategy presents a model of care that could be scalable to other hospitals.

Our study provides a case study on the use of multiple data sources and methods to develop a decision model for an early-stage health informatics intervention where knowledge gaps existed. As the purpose of this study was to evaluate the potential long-term effects of the *Medly* program in the management of patients with HF, the use of various data sources and modeling techniques are indispensable. This study was successful in developing a flexible algorithm based on the Cholesky decomposition method that was able to generate representative hypothetical cohorts of patients with HF according to the needs of the analysis [56]. An example of the algorithm's flexibility to adapt to the needs of the analyses was the ability to generate hypothetical patient cohorts for NYHA classes I, II, and III while maintaining the individual differences between patients within each class. Our study also successfully implemented a highly validated multivariate Cox model, the SHFM [42-45], within our algorithm to predict the survival of each generated patient over their lifetime. As the purpose of this study was to understand the *long-term effects of the Medly* program, the inclusion of the most validated predictive model for HF survival was logical [84]. The use of the SHFM provided a link for the survival probabilities derived in our model to a larger body HF research around predictive modeling. As mentioned, this was similar to the study by Reed et al (2015) [42], in which the SHFM was used as its underlying prognosis model and correlated health care costs and utility values via regression techniques [42,85,86].

### Comparison With Other Work

Other studies have investigated the cost-effectiveness of other types of TM in HF, where data are transmitted to medical staff. The study by Thokala et al (2013) [38] compared TM with usual care from the public payer perspective and found TM to be cost-effective at £11,873/QALY gained in 2011 (equivalent to Can \$19,996 (US \$15,084.18)/QALY gained in 2018) using a two-state (alive or dead) cohort-based Markov model over a 30-year time horizon. This relatively higher ICER than ICER in our study, which could be attributed to various factors, such as the lower cost of hospitalizations (£1529.97-£2514.49), led to less cost-saving per hospitalization. Furthermore, model structures differed between the study by Thokala et al [38] and our study, in which different health states and transition probabilities were used. A study by Liu et al (2016) [40] also broadly compared TM with standard care from the payer perspective in the United States and found cost-savings for specific scenarios. This included patients who were intermediate

and high-risk over 1- to 5-year time horizons. These results differ from our study owing to different model structures and transition probability parameters. The health states in the model developed by Liu et al (2016) were based on the number of past hospitalizations. In addition, hospitalization rates were conditional on both NYHA functional class and number of past hospitalizations, where the associated monthly probabilities for hospitalizations were much higher than those used in our study. This increased rate of hospitalizations in combination with the larger treatment effect size in reducing mortality and hospitalizations can be attributed to the cost-saving results [40]. A study by Grustam et al (2018) [39] compared TM with usual care for patients with HF from the public payer perspective within the Trans-European Network-Home-Care Management System using data from its original publication and other sources. This resulted in an ICER of €2,479/QALY gained in 2015 (Can \$18,145 [US \$13,687.86] in 2018), which was relatively higher than that reported in our study. The difference in our results can be attributed to the different methods used to model HF and measure effectiveness. The effectiveness of TM in the study was not defined by the risk for all-cause mortality and hospitalization events. Rather, effectiveness was measured by the decrease in probabilities of transitioning to more severe NYHA functional classes and the dead state based on an extrapolation from their database of patients with HF using TM [39].

### Limitations

As with any modeling exercise, it is important to understand the limitations around data availability and assumptions. First, owing to the lack of long-term studies, the trajectory of the effectiveness of TM was unknown and was assumed constant over the patient's lifetime. It is not known if effectiveness changes over time, which may affect the results of this study. Another limitation was the assumption that patients used *Medly* over the entirety of the model. It has been reported that clinicians have not established a generalizable duration of enrollment into the program [87]. Patients may be enrolled into the program for a period of time and be off-boarded after they have learned the necessary self-care behaviors for HF and no longer require the assistance of the technology. This could decrease the costs of the intervention. In addition, caution should be exercised when interpreting data from the *Medly* program evaluation because patients are enrolled in the program based on a joint decision between the clinician and the patient, which could lead to selection bias. Without strict enrollment criteria and end time points, patients may be less sick than the average patients with HF, making results less generalizable. Another limitation was that it was assumed that the *QoL of patients using Medly* was same as that of patients in the comparator group. However, evidence from the *Medly* program evaluation [47] and past literature [88] indicates that QoL improves when patients use *Medly*. As QoL in these studies was measured using HF-specific scoring tools, translation of the improved QoL to utility values used in this study was not feasible because the QALY is derived from the EuroQol-5 Dimension instrument [89]. This likely underestimated *Medly's* total QALYs gained. In addition, health care utilization data used to inform parameters in this study were based on the *Medly* program evaluation, which relied on

a relatively small sample of patients, self-reported ED and GP visits, and a database that was limited to events that occurred at 1 hospital. This early-stage evidence on baseline health care utilization in patients with HF may underestimate or overestimate the actual health care utilization of an HF population, which could alter the results of the study.

## Conclusions

The *Medly* program was found to be a cost-effective solution given the widely cited WTP thresholds of Can \$50,000 (US \$37,718)/QALY for patients with HF when implemented within a multidisciplinary HF clinic in Ontario. This is the first

Canadian economic evaluation of TM for HF using a cost-utility approach, and one of the few studies to use a long-term time horizon. The significance of this study includes providing evidence for health care decision makers on the use of TM for HF, supporting the use of a nurse-led model of TM within HF clinics, and informing the use of economic modeling for future evaluations of early-stage health informatics technology. Given the substantial impact of HF on patients' QoL and burden on health care resources, expanding access to TM programs may be an important mechanism to improve HF disease management and patient outcomes.

## Conflicts of Interest

HR and ES are considered inventors of the *Medly* system under the intellectual property policies of the UHN and may benefit from future commercialization of the technology by UHN.

## Multimedia Appendix 1

Additional methodology supplementary material.

[DOCX File, 62 KB - [jmir\\_v22i10e18917\\_app1.docx](#)]

## References

1. Savarese G, Lund LH. Global public health burden of heart failure. *Card Fail Rev* 2017 Apr;3(1):7-11 [FREE Full text] [doi: [10.1542/cfr.2016.25.2](#)] [Medline: [28785469](#)]
2. Heart disease in Canada: Highlights from the Canadian Chronic Disease Surveillance System. Government of Canada. 2020. URL: <https://www.canada.ca/en/public-health/services/publications/diseases-conditions/heart-disease-canada-fact-sheet.html> [accessed 2020-09-15]
3. Taylor CJ. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. *Br Med J* 2019 Oct 8;367:15840 [FREE Full text] [doi: [10.1136/bmj.15840](#)] [Medline: [31594775](#)]
4. Taylor CJ, Ryan R, Nichols L, Gale N, Hobbs FR, Marshall T. Survival following a diagnosis of heart failure in primary care. *Fam Pract* 2017 Apr 1;34(2):161-168 [FREE Full text] [doi: [10.1093/fampra/cmw145](#)] [Medline: [28137979](#)]
5. Gerber Y, Weston SA, Redfield MM, Chamberlain AM, Manemann SM, Jiang R, et al. A contemporary appraisal of the heart failure epidemic in Olmsted County, Minnesota, 2000 to 2010. *JAMA Intern Med* 2015 Jun;175(6):996-1004 [FREE Full text] [doi: [10.1001/jamainternmed.2015.0924](#)] [Medline: [25895156](#)]
6. Chun S, Tu JV, Wijeyesundera HC, Austin PC, Wang X, Levy D, et al. Lifetime analysis of hospitalizations and survival of patients newly admitted with heart failure. *Circ Heart Fail* 2012 Jul 1;5(4):414-421 [FREE Full text] [doi: [10.1161/CIRCHEARTFAILURE.111.964791](#)] [Medline: [22556322](#)]
7. Desai AS, Stevenson LW. Rehospitalization for heart failure: predict or prevent? *Circulation* 2012 Jul 24;126(4):501-506. [doi: [10.1161/CIRCULATIONAHA.112.125435](#)] [Medline: [22825412](#)]
8. Tran DT, Ohinmaa A, Thanh NX, Howlett JG, Ezekowitz JA, McAlister FA, et al. The current and future financial burden of hospital admissions for heart failure in Canada: a cost analysis. *CMAJ Open* 2016;4(3):E365-E370 [FREE Full text] [doi: [10.9778/cmajo.20150130](#)] [Medline: [27730101](#)]
9. Virani SA, Bains M, Code J, Ducharme A, Harkness K, Howlett JG, Board and Membership of the Canadian Heart Failure Society. The need for heart failure advocacy in Canada. *Can J Cardiol* 2017 Nov;33(11):1450-1454. [doi: [10.1016/j.cjca.2017.08.024](#)] [Medline: [29111108](#)]
10. Lin M, Yuan W, Huang T, Zhang H, Mai J, Wang J. Clinical effectiveness of telemedicine for chronic heart failure: a systematic review and meta-analysis. *J Investig Med* 2017 Jun;65(5):899-911. [doi: [10.1136/jim-2016-000199](#)] [Medline: [28330835](#)]
11. Yun JE, Park J, Park H, Lee H, Park D. Comparative effectiveness of telemonitoring versus usual care for heart failure: a systematic review and meta-analysis. *J Card Fail* 2018 Jan;24(1):19-28. [doi: [10.1016/j.cardfail.2017.09.006](#)] [Medline: [28939459](#)]
12. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. *J Med Internet Res* 2015 Mar 12;17(3):e63 [FREE Full text] [doi: [10.2196/jmir.4174](#)] [Medline: [25768664](#)]
13. Nakamura N, Koga T, Iseki H. A meta-analysis of remote patient monitoring for chronic heart failure patients. *J Telemed Telecare* 2014 Jan;20(1):11-17. [doi: [10.1177/1357633X13517352](#)] [Medline: [24352899](#)]

14. Xiang R, Li L, Liu SX. Meta-analysis and meta-regression of telehealth programmes for patients with chronic heart failure. *J Telemed Telecare* 2013 Jul;19(5):249-259. [doi: [10.1177/1357633X13495490](https://doi.org/10.1177/1357633X13495490)] [Medline: [24163234](#)]
15. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition-Heart Failure (BEAT-HF) Research Group. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: the better effectiveness after transition -- heart failure (BEAT-HF) randomized clinical trial. *JAMA Intern Med* 2016 Mar;176(3):310-318 [FREE Full text] [doi: [10.1001/jamainternmed.2015.7712](https://doi.org/10.1001/jamainternmed.2015.7712)] [Medline: [26857383](#)]
16. Chaudhry SI, Matterna JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. *N Engl J Med* 2010 Dec 9;363(24):2301-2309. [doi: [10.1056/nejmoa1010029](https://doi.org/10.1056/nejmoa1010029)]
17. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. *Lancet* 2018 Sep;392(10152):1047-1057. [doi: [10.1016/s0140-6736\(18\)31880-4](https://doi.org/10.1016/s0140-6736(18)31880-4)]
18. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, Telemedical Interventional Monitoring in Heart Failure Investigators. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. *Circulation* 2011 May 3;123(17):1873-1880. [doi: [10.1161/CIRCULATIONAHA.111.018473](https://doi.org/10.1161/CIRCULATIONAHA.111.018473)] [Medline: [21444883](#)]
19. Ware P, Seto E, Ross HJ. Accounting for complexity in home telemonitoring: a need for context-centred evidence. *Can J Cardiol* 2018 Jul;34(7):897-904. [doi: [10.1016/j.cjca.2018.01.022](https://doi.org/10.1016/j.cjca.2018.01.022)] [Medline: [29861204](#)]
20. Dávalos ME, French MT, Burdick AE, Simmons SC. Economic evaluation of telemedicine: review of the literature and research guidelines for benefit-cost analysis. *Telemed J E Health* 2009 Dec;15(10):933-948. [doi: [10.1089/tmj.2009.0067](https://doi.org/10.1089/tmj.2009.0067)] [Medline: [19954346](#)]
21. Bergmo TS. Approaches to economic evaluation in telemedicine. *J Telemed Telecare* 2012 Jun;18(4):181-184. [doi: [10.1258/jtt.2012.111112](https://doi.org/10.1258/jtt.2012.111112)] [Medline: [22619374](#)]
22. Noel HC, Vogel DC, Erdos JJ, Cornwall D, Levin F. Home telehealth reduces healthcare costs. *Telemed J E Health* 2004;10(2):170-183. [doi: [10.1089/tmj.2004.10.170](https://doi.org/10.1089/tmj.2004.10.170)] [Medline: [15319047](#)]
23. Myers S, Grant RW, Lugn NE, Holbert B, Kvedar JC. Impact of home-based monitoring on the care of patients with congestive heart failure. *Home Health Care Manag Pract* 2016 Jun 30;18(6):444-451. [doi: [10.1177/1084822306289991](https://doi.org/10.1177/1084822306289991)]
24. Hudson LR, Hamar GB, Orr P, Johnson JH, Neftzger A, Chung RS, et al. Remote physiological monitoring: clinical, financial, and behavioral outcomes in a heart failure population. *Dis Manag* 2005 Dec;8(6):372-381. [doi: [10.1089/dis.2005.8.372](https://doi.org/10.1089/dis.2005.8.372)] [Medline: [16351555](#)]
25. Giordano A, Scalvini S, Zanelli E, Corrà U, Longobardi GL, Ricci VA, et al. Multicenter randomised trial on home-based telemanagement to prevent hospital readmission of patients with chronic heart failure. *Int J Cardiol* 2009 Jan 9;131(2):192-199. [doi: [10.1016/j.ijcard.2007.10.027](https://doi.org/10.1016/j.ijcard.2007.10.027)] [Medline: [18222552](#)]
26. Scalvini S, Capomolla S, Zanelli E, Benigno M, Domenighini D, Paletta L, et al. Effect of home-based telecardiology on chronic heart failure: costs and outcomes. *J Telemed Telecare* 2005;11(Suppl 1):16-18. [doi: [10.1258/1357633054461688](https://doi.org/10.1258/1357633054461688)] [Medline: [16035980](#)]
27. Vaccaro J, Cherry J, Harper A, O'Connell M. Utilization reduction, cost savings, and return on investment for the pacificare chronic heart failure program, 'taking charge of your heart health'. *Disease Manag* 2001 Sep;4(3):131-142. [doi: [10.1089/10935070152596052](https://doi.org/10.1089/10935070152596052)]
28. Galbreath AD, Krasuski RA, Smith B, Stajduhar KC, Kwan MD, Ellis R, et al. Long-term healthcare and cost outcomes of disease management in a large, randomized, community-based population with heart failure. *Circulation* 2004 Dec 7;110(23):3518-3526. [doi: [10.1161/01.CIR.0000148957.62328.89](https://doi.org/10.1161/01.CIR.0000148957.62328.89)] [Medline: [15531765](#)]
29. Schwarz KA, Mion L, Hudock D, Litman G. Telemonitoring of heart failure patients and their caregivers: a pilot randomized controlled trial. *Prog Cardiovasc Nurs* 2008;23(1):18-26. [doi: [10.1111/j.1751-7117.2008.06611.x](https://doi.org/10.1111/j.1751-7117.2008.06611.x)] [Medline: [18326990](#)]
30. Soran OZ, Feldman AM, Piña IL, Lamas GA, Kelsey SF, Selzer F, et al. Cost of medical services in older patients with heart failure: those receiving enhanced monitoring using a computer-based telephonic monitoring system compared with those in usual care: the heart failure home care trial. *J Card Fail* 2010 Nov;16(11):859-866. [doi: [10.1016/j.cardfail.2010.05.028](https://doi.org/10.1016/j.cardfail.2010.05.028)] [Medline: [21055649](#)]
31. Tompkins C, Orwat J. A randomized trial of telemonitoring heart failure patients. *J Healthc Manag* 2010;55(5):312-22; discussion 322. [Medline: [21077581](#)]
32. Kenealy TW, Parsons MJ, Rouse AP, Doughty RN, Sheridan NF, Hindmarsh JK, et al. Telecare for diabetes, CHF or COPD: effect on quality of life, hospital use and costs. A randomised controlled trial and qualitative evaluation. *PLoS One* 2015;10(3):e0116188 [FREE Full text] [doi: [10.1371/journal.pone.0116188](https://doi.org/10.1371/journal.pone.0116188)] [Medline: [25768023](#)]
33. Paré G, Poba-Nzaou P, Sicotte C. Home telemonitoring for chronic disease management: an economic assessment. *Int J Technol Assess Health Care* 2013 Apr;29(2):155-161. [doi: [10.1017/S0266462313000111](https://doi.org/10.1017/S0266462313000111)] [Medline: [23514722](#)]
34. Methods for the Economic Evaluation of Health Care Programmes. Oxford, UK: Oxford University Press; 2005.
35. Smith B, Hughes-Cromwick P, Forkner E, Galbreath A. Cost-effectiveness of telephonic disease management in heart failure. *Am J Manag Care* 2008 Feb;14(2):106-115 [FREE Full text] [Medline: [18269306](#)]



36. Klersy C, de Silvestri A, Gabutti G, Raisaro A, Curti M, Regoli F, et al. Economic impact of remote patient monitoring: an integrated economic model derived from a meta-analysis of randomized controlled trials in heart failure. *Eur J Heart Fail* 2011 Apr;13(4):450-459 [FREE Full text] [doi: [10.1093/eurjhf/hfq232](https://doi.org/10.1093/eurjhf/hfq232)] [Medline: [21193439](#)]
37. Cui Y, Doupe M, Katz A, Nyhof P, Forget E. Economic evaluation of Manitoba Health Lines in the management of congestive heart failure. *Healthc Policy* 2013 Nov;9(2):36-50 [FREE Full text] [Medline: [24359716](#)]
38. Thokala P, Baalbaki H, Brennan A, Pandor A, Stevens JW, Gomersall T, et al. Telemonitoring after discharge from hospital with heart failure: cost-effectiveness modelling of alternative service designs. *BMJ Open* 2013 Sep 18;3(9):e003250 [FREE Full text] [doi: [10.1136/bmjopen-2013-003250](https://doi.org/10.1136/bmjopen-2013-003250)] [Medline: [24048626](#)]
39. Grustam AS, Severens JL, de Massari D, Buyukkaramikli N, Koymans R, Vrijhoef HJ. Cost-effectiveness analysis in telehealth: a comparison between home telemonitoring, nurse telephone support, and usual care in chronic heart failure management. *Value Health* 2018 Jul;21(7):772-782 [FREE Full text] [doi: [10.1016/j.jval.2017.11.011](https://doi.org/10.1016/j.jval.2017.11.011)] [Medline: [30005749](#)]
40. Liu SX, Xiang R, Lagor C, Liu N, Sullivan K. Economic modeling of heart failure telehealth programs: when do they become cost saving? *Int J Telemed Appl* 2016;2016:3289628 [FREE Full text] [doi: [10.1155/2016/3289628](https://doi.org/10.1155/2016/3289628)] [Medline: [27528868](#)]
41. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, ISPOR Health Economic Evaluation Publication Guidelines-CHEERS Good Reporting Practices Task Force. Consolidated health economic evaluation reporting standards (cheers)--explanation and elaboration: a report of the ISOPR health economic evaluation publication guidelines good reporting practices task force. *Value Health* 2013;16(2):231-250 [FREE Full text] [doi: [10.1016/j.jval.2013.02.002](https://doi.org/10.1016/j.jval.2013.02.002)] [Medline: [23538175](#)]
42. Reed SD, Neilson MP, Gardner M, Li Y, Briggs AH, Polsky DE, et al. Tools for economic analysis of patient management interventions in heart failure cost-effectiveness model: a web-based program designed to evaluate the cost-effectiveness of disease management programs in heart failure. *Am Heart J* 2015 Nov;170(5):951-960 [FREE Full text] [doi: [10.1016/j.ahj.2015.08.015](https://doi.org/10.1016/j.ahj.2015.08.015)] [Medline: [26542504](#)]
43. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, et al. The Seattle heart failure model: prediction of survival in heart failure. *Circulation* 2006 Mar 21;113(11):1424-1433. [doi: [10.1161/CIRCULATIONAHA.105.584102](https://doi.org/10.1161/CIRCULATIONAHA.105.584102)] [Medline: [16534009](#)]
44. Kalogeropoulos AP, Georgiopoulos VV, Giamouzis G, Smith AL, Agha SA, Waheed S, et al. Utility of the Seattle Heart Failure Model in patients with advanced heart failure. *J Am Coll Cardiol* 2009 Jan 27;53(4):334-342 [FREE Full text] [doi: [10.1016/j.jacc.2008.10.023](https://doi.org/10.1016/j.jacc.2008.10.023)] [Medline: [19161882](#)]
45. Ketchum ES, Moorman AJ, Fishbein DP, Mokadam NA, Verrier ED, Aldea GS, et al. Predictive value of the Seattle Heart Failure Model in patients undergoing left ventricular assist device placement. *J Heart Lung Transplant* 2010 Sep;29(9):1021-1025. [doi: [10.1016/j.healun.2010.05.002](https://doi.org/10.1016/j.healun.2010.05.002)] [Medline: [20558086](#)]
46. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Developing healthcare rule-based expert systems: case study of a heart failure telemonitoring system. *Int J Med Inform* 2012 Aug;81(8):556-565. [doi: [10.1016/j.jimedinf.2012.03.001](https://doi.org/10.1016/j.jimedinf.2012.03.001)] [Medline: [22465288](#)]
47. Ware P, Ross HJ, Cafazzo JA, Boodoo C, Munnery M, Seto E. Outcomes of a heart failure telemonitoring program implemented as the standard of care in an outpatient heart function clinic: pretest-posttest pragmatic study. *J Med Internet Res* 2020 Feb 8;22(2):e16538 [FREE Full text] [doi: [10.2196/16538](https://doi.org/10.2196/16538)] [Medline: [32027309](#)]
48. Wijesundera HC, Trubiani G, Abrahamyan L, Mitsakakis N, Witteman W, Paulden M, et al. Specialized multi-disciplinary heart failure clinics in Ontario, Canada: an environmental scan. *BMC Health Serv Res* 2012 Aug 3;12:236 [FREE Full text] [doi: [10.1186/1472-6963-12-236](https://doi.org/10.1186/1472-6963-12-236)] [Medline: [22863276](#)]
49. Guidelines for the Economic Evaluation of Health Technologies: Canada. Canadian Agency for Drugs and Technologies in Health (CADTH). 2017. URL: <https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada> [accessed 2020-09-15]
50. Krijkamp EM, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MG, Pechlivanoglou P. Microsimulation modeling for health decision sciences using R: a tutorial. *Med Decis Making* 2018 Apr;38(3):400-422 [FREE Full text] [doi: [10.1177/0272989X18754513](https://doi.org/10.1177/0272989X18754513)] [Medline: [29587047](#)]
51. Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels. Boston, UK: Martin Dolgin; 1994.
52. Desai AS. The three-phase terrain of heart failure readmissions. *Circ Heart Fail* 2012 Jul 1;5(4):398-400. [doi: [10.1161/CIRCHEARTFAILURE.112.968735](https://doi.org/10.1161/CIRCHEARTFAILURE.112.968735)] [Medline: [22811548](#)]
53. Giamouzis G, Kalogeropoulos A, Georgiopoulos V, Laskar S, Smith AL, Dunbar S, et al. Hospitalization epidemic in patients with heart failure: risk factors, risk prediction, knowledge gaps, and future directions. *J Card Fail* 2011 Jan;17(1):54-75. [doi: [10.1016/j.cardfail.2010.08.010](https://doi.org/10.1016/j.cardfail.2010.08.010)] [Medline: [21187265](#)]
54. Canadian Institute for Health Information (CIHI). All-Cause Readmission to Acute Care and Return to the Emergency Department. 2020. URL: [https://secure.cihi.ca/free\\_products/Readmission\\_to\\_acutecare\\_en.pdf](https://secure.cihi.ca/free_products/Readmission_to_acutecare_en.pdf) [accessed 2020-09-15]
55. Naimark DM, Kabboul NN, Krahn MD. The half-cycle correction revisited: redemption of a kludge. *Med Decis Making* 2013 Oct;33(7):961-970. [doi: [10.1177/0272989X13501558](https://doi.org/10.1177/0272989X13501558)] [Medline: [24048350](#)]
56. Decision Modelling for Health Economic Evaluation. Oxford, UK: Oxford University Press; 2006.



57. Beta-Blocker Evaluation of Survival Trial Investigators, Eichhorn EJ, Domanski MJ, Krause-Steinrauf H, Bristow MR, Lavori PW. A trial of the beta-blocker bucindolol in patients with advanced chronic heart failure. *N Engl J Med* 2001 May 31;344(22):1659-1667. [doi: [10.1056/NEJM200105313442202](https://doi.org/10.1056/NEJM200105313442202)] [Medline: [11386264](https://pubmed.ncbi.nlm.nih.gov/11386264/)]
58. Chan DC, Heidenreich PA, Weinstein MC, Fonarow GC. Heart failure disease management programs: a cost-effectiveness analysis. *Am Heart J* 2008 Feb;155(2):332-338. [doi: [10.1016/j.ahj.2007.10.001](https://doi.org/10.1016/j.ahj.2007.10.001)] [Medline: [18215605](https://pubmed.ncbi.nlm.nih.gov/18215605/)]
59. Delea TE, Vera-Llonch M, Richner RE, Fowler MB, Oster G. Cost effectiveness of carvedilol for heart failure. *Am J Cardiol* 1999 Mar;83(6):890-896 [FREE Full text] [doi: [10.1016/s0002-9149\(98\)01066-2](https://doi.org/10.1016/s0002-9149(98)01066-2)]
60. Göhler A, Geisler BP, Manne JM, Kosiborod M, Zhang Z, Weintraub WS, et al. Utility estimates for decision-analytic modeling in chronic heart failure--health states based on New York Heart Association classes and number of rehospitalizations. *Value Health* 2009;12(1):185-187 [FREE Full text] [doi: [10.1111/j.1524-4733.2008.00425.x](https://doi.org/10.1111/j.1524-4733.2008.00425.x)] [Medline: [18647251](https://pubmed.ncbi.nlm.nih.gov/18647251/)]
61. Ahmed A, Aronow WS, Fleg JL. Higher New York Heart Association classes and increased mortality and hospitalization in patients with heart failure and preserved left ventricular function. *Am Heart J* 2006 Feb;151(2):444-450 [FREE Full text] [doi: [10.1016/j.ahj.2005.03.066](https://doi.org/10.1016/j.ahj.2005.03.066)] [Medline: [16442912](https://pubmed.ncbi.nlm.nih.gov/16442912/)]
62. Yao B, Freemantle N, Tharmanathan P, Coats A, Poole-Wilson PA. Long-term cost-effectiveness analysis of nebivolol compared with standard care in elderly patients with heart failure: an individual patient-based simulation model. *Pharmacoecon Open* 2008;26(10):879-889 [FREE Full text] [doi: [10.1186/1471-2105-5-99](https://doi.org/10.1186/1471-2105-5-99)] [Medline: [15272935](https://pubmed.ncbi.nlm.nih.gov/15272935/)]
63. Flather MD, Shibata MC, Coats AJ, van Veldhuisen DJ, Parkhomenko A, Borbola J, SENIORS Investigators. Randomized trial to determine the effect of nebivolol on mortality and cardiovascular hospital admission in elderly patients with heart failure (SENIORS). *Eur Heart J* 2005 Feb;26(3):215-225. [doi: [10.1093/eurheartj/ehi115](https://doi.org/10.1093/eurheartj/ehi115)] [Medline: [15642700](https://pubmed.ncbi.nlm.nih.gov/15642700/)]
64. Yeung DF, Boom NK, Guo H, Lee DS, Schultz SE, Tu JV. Trends in the incidence and outcomes of heart failure in Ontario, Canada: 1997 to 2007. *Can Med Assoc J* 2012 Oct 2;184(14):E765-E773 [FREE Full text] [doi: [10.1503/cmaj.111958](https://doi.org/10.1503/cmaj.111958)] [Medline: [22908143](https://pubmed.ncbi.nlm.nih.gov/22908143/)]
65. Reed SD, Neilson MP, Gardner M, Li Y, Briggs AH, Polsky DE, et al. Tools for economic analysis of patient management interventions in heart failure cost-effectiveness model: a web-based program designed to evaluate the cost-effectiveness of disease management programs in heart failure. *Am Heart J* 2015 Nov;170(5):951-960 [FREE Full text] [doi: [10.1016/j.ahj.2015.08.015](https://doi.org/10.1016/j.ahj.2015.08.015)] [Medline: [26542504](https://pubmed.ncbi.nlm.nih.gov/26542504/)]
66. Amatya A, Demirtas H. Simultaneous generation of multivariate mixed data with Poisson and normal marginals. *J Stat Comput Sim* 2014 Sep;85(15):3129-3139. [doi: [10.1080/00949655.2014.953534](https://doi.org/10.1080/00949655.2014.953534)]
67. Demirtas H, Doganay B. Simultaneous generation of binary and normal data with specified marginal and association structures. *J Biopharm Stat* 2012;22(2):223-236. [doi: [10.1080/10543406.2010.521874](https://doi.org/10.1080/10543406.2010.521874)] [Medline: [22251171](https://pubmed.ncbi.nlm.nih.gov/22251171/)]
68. Ford E, Adams J, Graves N. Development of an economic model to assess the cost-effectiveness of hawthorn extract as an adjunct treatment for heart failure in Australia. *BMJ Open* 2012;2(5):- [FREE Full text] [doi: [10.1136/bmjopen-2012-001094](https://doi.org/10.1136/bmjopen-2012-001094)] [Medline: [22942231](https://pubmed.ncbi.nlm.nih.gov/22942231/)]
69. Borisenko O, Haude M, Hoppe UC, Siminiak T, Lipiecki J, Goldberg SL, et al. Cost-utility analysis of percutaneous mitral valve repair in inoperable patients with functional mitral regurgitation in German settings. *BMC Cardiovasc Disord* 2015 May 14;15:43 [FREE Full text] [doi: [10.1186/s12872-015-0039-8](https://doi.org/10.1186/s12872-015-0039-8)] [Medline: [25971307](https://pubmed.ncbi.nlm.nih.gov/25971307/)]
70. Jaagosild P, Dawson NV, Thomas C, Wenger NS, Tsevat J, Knaus WA, et al. Outcomes of acute exacerbation of severe congestive heart failure: quality of life, resource use, and survival. Support investigators. The study to understand prognosis and preferences for outcomes and risks of treatments. *Arch Intern Med* 1998 May 25;158(10):1081-1089. [doi: [10.1001/archinte.158.10.1081](https://doi.org/10.1001/archinte.158.10.1081)] [Medline: [9605779](https://pubmed.ncbi.nlm.nih.gov/9605779/)]
71. Sandhu AT, Goldhaber-Fiebert JD, Owens DK, Turakhia MP, Kaiser DW, Heidenreich PA. Cost-effectiveness of implantable pulmonary artery pressure monitoring in chronic heart failure. *JACC Heart Fail* 2016 May;4(5):368-375 [FREE Full text] [doi: [10.1016/j.jchf.2015.12.015](https://doi.org/10.1016/j.jchf.2015.12.015)] [Medline: [26874380](https://pubmed.ncbi.nlm.nih.gov/26874380/)]
72. Kaul P, McAlister FA, Ezekowitz JA, Bakal JA, Curtis LH, Quan H, et al. Resource use in the last 6 months of life among patients with heart failure in Canada. *Arch Intern Med* 2011 Feb 14;171(3):211-217. [doi: [10.1001/archinternmed.2010.365](https://doi.org/10.1001/archinternmed.2010.365)] [Medline: [20937918](https://pubmed.ncbi.nlm.nih.gov/20937918/)]
73. Costing Analysis Tool. Ontario Case Costing. 2019. URL: [https://www.engenderhealth.org/files/pubs/qi/toolbook/cost\\_analysis\\_tool.pdf](https://www.engenderhealth.org/files/pubs/qi/toolbook/cost_analysis_tool.pdf) [accessed 2020-09-15]
74. Schedule of Benefits: Physician Services Under the Health Insurance Act. Ministry of Health and Long Term Care. URL: <http://govdocs.ourontario.ca/node/2924> [accessed 2020-09-15]
75. Health Costing in Alberta 2006: Annual Report. Alberta Health and Wellness. 2006. URL: [https://www.assembly.ab.ca/lao/library/egovdocs/2006/alhw/129693\\_06.pdf](https://www.assembly.ab.ca/lao/library/egovdocs/2006/alhw/129693_06.pdf) [accessed 2020-09-15]
76. Mihaylova B, Briggs A, O'Hagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Econ* 2011 Aug;20(8):897-916 [FREE Full text] [doi: [10.1002/hec.1653](https://doi.org/10.1002/hec.1653)] [Medline: [20799344](https://pubmed.ncbi.nlm.nih.gov/20799344/)]
77. The Canadian Consumer Price Index Reference Paper. Statistics Canada. 2019. URL: <https://www150.statcan.gc.ca/n1/pub/62-553-x/62-553-x2015001-eng.pdf> [accessed 2020-09-15]
78. Delignette-Muller ML, Dutang C. An R package for fitting distributions. *J Stat Soft* 2015;64(4):- [doi: [10.18637/jss.v064.i04](https://doi.org/10.18637/jss.v064.i04)]

79. Valuing Health in Canada: Who, How, and How Much. Canada 2020. 2013. URL: <http://canada2020.ca/wp-content/uploads/2013/06/Canada-2020-Analytical-Commentary-No-3-Valuing-Health-in-Canada-FINAL.pdf> [accessed 2020-09-15]
80. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Can Med Assoc J* 1992 Feb 15;146(4):473-481 [[FREE Full text](#)] [Medline: [1306034](#)]
81. Seto E, Leonard KJ, Masino C, Cafazzo JA, Barnsley J, Ross HJ. Attitudes of heart failure patients and health care providers towards mobile phone-based remote monitoring. *J Med Internet Res* 2010 Nov 29;12(4):e55 [[FREE Full text](#)] [doi: [10.2196/jmir.1627](#)] [Medline: [21115435](#)]
82. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Perceptions and experiences of heart failure patients and clinicians on the use of mobile phone-based telemonitoring. *J Med Internet Res* 2012 Feb 10;14(1):e25 [[FREE Full text](#)] [doi: [10.2196/jmir.1912](#)] [Medline: [22328237](#)]
83. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. Evaluating the implementation of a mobile phone-based telemonitoring program: longitudinal study guided by the consolidated framework for implementation research. *JMIR Mhealth Uhealth* 2018 Jul 31;6(7):e10768 [[FREE Full text](#)] [doi: [10.2196/10768](#)] [Medline: [30064970](#)]
84. Alba AC, Agoritsas T, Jankowski M, Courvoisier D, Walter SD, Guyatt GH, et al. Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review. *Circ Heart Fail* 2013 Sep 1;6(5):881-889. [doi: [10.1161/CIRCHEARTFAILURE.112.000043](#)] [Medline: [23888045](#)]
85. Li Y, Neilson MP, Whellan DJ, Schulman KA, Levy WC, Reed SD. Associations between Seattle Heart Failure Model scores and health utilities: findings from HF-ACTION. *J Card Fail* 2013 May;19(5):311-316 [[FREE Full text](#)] [doi: [10.1016/j.cardfail.2013.03.008](#)] [Medline: [23663813](#)]
86. Reed SD, Li Y, Ellis SJ, Whellan DJ, Schulman KA, O'Connor CM, et al. Seattle heart failure model scores significantly predict medical resource use and costs in HF-action. *J Card Fail* 2011 Aug;17(8):S80. [doi: [10.1016/j.cardfail.2011.06.272](#)]
87. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. User-centered adaptation of an existing heart failure telemonitoring program to ensure sustainability and scalability: qualitative study. *JMIR Cardio* 2018 Dec 6;2(2):e11466 [[FREE Full text](#)] [doi: [10.2196/11466](#)] [Medline: [31758774](#)]
88. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. *J Med Internet Res* 2012 Feb 16;14(1):e31 [[FREE Full text](#)] [doi: [10.2196/jmir.1909](#)] [Medline: [22356799](#)]
89. Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, Canadian EQ-5D-5L Valuation Study Group. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care* 2016 Jan;54(1):98-105 [[FREE Full text](#)] [doi: [10.1097/MLR.0000000000000447](#)] [Medline: [26492214](#)]

## Abbreviations

**BYOE:** Bring Your Own Everything  
**BYOP:** Bring Your Own Phone  
**CEAC:** cost-effectiveness acceptability curve  
**CUA:** cost-utility analysis  
**ED:** emergency department  
**FK:** Full Kit  
**GP:** general practitioner  
**HF:** heart failure  
**ICER:** incremental cost-effectiveness ratio  
**NYHA:** New York Heart Association  
**QALY:** quality-adjusted life year  
**QoL:** quality of life  
**RR:** relative risk  
**SHFM:** Seattle Heart Failure Model  
**TM:** telemonitoring  
**UHN:** University Health Network  
**WTP:** willingness-to-pay

*Edited by G Eysenbach; submitted 27.03.20; peer-reviewed by B Chan, K Wei; comments to author 12.06.20; revised version received 05.08.20; accepted 06.08.20; published 06.10.20.*

*Please cite as:*

Boodoo C, Zhang Q, Ross HJ, Alba AC, Laporte A, Seto E

*Evaluation of a Heart Failure Telemonitoring Program Through a Microsimulation Model: Cost-Utility Analysis*

*J Med Internet Res* 2020;22(10):e18917

URL: <https://www.jmir.org/2020/10/e18917>

doi: [10.2196/18917](https://doi.org/10.2196/18917)

PMID: [33021485](https://pubmed.ncbi.nlm.nih.gov/33021485/)

©Chris Boodoo, Qi Zhang, Heather J Ross, Ana Carolina Alba, Audrey Laporte, Emily Seto. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Evaluating the Implementation of a Remote-Monitoring Program for Chronic Obstructive Pulmonary Disease: Qualitative Methods from a Service Design Perspective

Florence van Lieshout<sup>1</sup>, BSc; Rebecca Yang<sup>2</sup>, MPH; Vess Stamenova<sup>2</sup>, PhD; Payal Agarwal<sup>2</sup>, MD; Daniel Cornejo Palma<sup>2</sup>, MD; Aman Sidhu<sup>2</sup>, MD; Katrina Engel<sup>3</sup>, MSc; Adam Erwood<sup>3</sup>, MSc; R Sacha Bhatia<sup>2</sup>, MD; Onil Bhattacharyya<sup>2</sup>, MD, PhD; James Shaw<sup>2,4</sup>, PhD

<sup>1</sup>University of Amsterdam, Amsterdam, Netherlands

<sup>2</sup>Women's College Hospital Institute for Health System Solutions and Virtual Care, Toronto, ON, Canada

<sup>3</sup>Markham Stouffville Hospital, Markham, ON, Canada

<sup>4</sup>University of Toronto Joint Centre for Bioethics, Toronto, ON, Canada

**Corresponding Author:**

Florence van Lieshout, BSc

University of Amsterdam

Spui 21

Amsterdam, 1012 WX

Netherlands

Phone: 31 4163236400

Email: [floorvanlieshout@gmail.com](mailto:floorvanlieshout@gmail.com)

## Abstract

**Background:** Implementing digital health technologies is complex but can be facilitated by considering the features of the tool that is being implemented, the team that will use it, and the routines that will be affected.

**Objective:** The goal of this study was to assess the implementation of a remote-monitoring initiative for patients with chronic obstructive pulmonary disease in Ontario, Canada using the Tool+Team+Routine framework and to refine this approach to conceptualize the adoption of technologies in health care.

**Methods:** This study was a qualitative research project that took place alongside a randomized controlled trial comparing a technology-enabled self-monitoring program with a technology-enabled self- and remote-monitoring program in patients with chronic obstructive pulmonary disease and with standard care. This study included interviews with 5 remote-monitoring patients, 3 self-monitoring patients, 2 caregivers, 5 health care providers, and 3 hospital administrators. The interview questions were structured around the 3 main concepts of the Tool+Team+Routine framework.

**Results:** Findings emphasized that (1) technologies can alter relationships between providers and patients, and that these relationships drove the development of a new service arising from the technology, in our case, and (2) technologies can create additional work that is not visible to management as a result of not being considered within the scope of the service.

**Conclusions:** Literature on the implementation of digital health technologies has still not reconciled the importance of interpersonal relationships to conventional implementation strategies. By acknowledging the centrality of such relationships, implementation teams can better plan for the adaptations required in order to make new technologies work for patients and health care providers. Further work will need to address how specific individuals administering a remote-monitoring program work to build relationships, and how these relationships and other sources of activity might lead to technological scope creep—an unanticipated expanding scope of work activities in relation to the function of the tool.

(*J Med Internet Res* 2020;22(10):e18148) doi:[10.2196/18148](https://doi.org/10.2196/18148)

**KEYWORDS**

service design; digital health; innovation; implementation science; remote monitoring; telemedicine

## Introduction

The digital health market continues to expand, with projected growth from US \$79 billion in 2015 to an estimated value of US \$206 billion in 2020 [1]. This growth is driven, in part, by the fact that many patient-facing digital health technologies, such as remote monitoring, have shown potential for positive effects on patient health and health care provider performance, such as enhanced chronic disease management and enhanced access to care [2,3]. However, despite some demonstrated successes with digital health technologies, positive results remain sporadic and difficult to achieve [4].

The implementation of digital health technologies remains an extremely complex area of health system improvement, and remote monitoring has served as an important example of the challenges associated with encouraging the meaningful use of technologies in routine care [5,6]. The need for patients to feel comfortable with a given technology in addition to the health care provider team engaging in remote care delivery creates additional barriers to successful adoption [7,8]. Furthermore, many technologies are advertised to be “plug and play,” suggesting there should be no barriers to their immediate use when, in fact, much local adaptation is required for them to be incorporated into users’ everyday lives [9]. These observations illustrate the persistent and unresolved issues associated with determining both the mechanisms through which remote monitoring works for users and the optimal strategies to promote implementation.

Unsuccessful implementation of technologies such as remote monitoring is a result of several complex influences [10], including poor alignment with the main objective of the health care organization, poor training of staff, nonalignment of procurement models, or noninteroperability of the technology with existing systems [11,12]. Over the past 10 years especially, methods and frameworks have been developed to support better implementation of technological tools in health care settings [10,13,14]. Many of these approaches have been rooted specifically in theories of implementation science, including the Consolidated Framework for Implementation Research [15] and the Normalization Process Theory [16]. Theories of implementation science have been important in clarifying the conceptual dimensions requiring consideration in technology implementation initiatives. However, we suggest they have done less to clearly articulate the most important links between the adoption of a technology and the redesign of the service into which a technology is being implemented.

This basic point formed the motivation for previous work by our research team to present an approach to the adoption of digital health technologies that emphasizes the links between technological innovations and service innovations, informed by a service design perspective [17]. Service design is an approach to the improvement or establishment of services that takes the entire service as the primary focus, emphasizing the importance of user experience and the realities of service delivery for health care providers and other staff [17]. In our past work, we formulated the objective of service design: to “carefully plan and promote the coordinated action required to

execute a high quality health care service [17].” In this way, service design considers a technology for its contributions to a service more generally by focusing less on the adoption of the technology itself and more on its contributions to the work required to deliver an excellent service overall.

The recently developed Tool+Team+Routine framework [17] outlines a service design approach to innovation adoption that includes (1) identifying the value propositions of a given tool, (2) considering the implications of the tool for team relationships, and (3) explicitly addressing the new routines required in the technology-informed reconfiguration of the service. Although this approach is based on past empirical research examining this alternative approach to technology adoption, the Tool+Team+Routine framework is a simple heuristic that has not yet been subject to critical examination [17].

The goals of this study were two-fold. First, to apply the Tool+Team+Routine framework to interpret the results of the implementation of a remote-monitoring initiative for patients with chronic obstructive pulmonary disease (COPD) in Ontario, Canada, in order to generate insights that inform potential improvements to the Tool+Team+Routine framework. Second, to identify insights pertaining to the influences that drive adoption of remote-monitoring technologies specifically.

## Methods

### Study Setting

This is a qualitative substudy of a randomized controlled trial [18]; both the randomized controlled trial and this qualitative study were conducted at a community-based hospital in a large city in Ontario, Canada, within a respiratory center that treats patients with COPD. The randomized controlled trial investigated the effectiveness of implementing a technology-enabled self-monitoring program or a technology enabled self- and remote-monitoring program in a COPD patient population compared to standard care. Both the self-monitoring and the remote-monitoring groups used the Cloud DX Connected Health Kit as the monitoring tool [19]. This tool consisted of a tablet and multiple smaller devices to measure physiological vitals and to complete surveys on objective COPD symptoms. The tablet notified patients whenever their readings exceeded their personal thresholds.

All patients from both groups were also provided with a paper-based action plan, which they could follow whenever they received such notifications. The vitals and questionnaire answers from the remote-monitoring group were monitored by the project lead who is also a respiratory therapist. If it was thought to be necessary by the respiratory therapist, based on the vitals or the questionnaire answers, the patient was contacted directly. In addition, the project lead called each remote-monitoring patient once a week.

The study was approved by Markham Stouffville Hospital and Women’s College Hospital research ethics boards in Ontario, Canada.



## Participants

We included 8 patients with COPD (5 from the remote-monitoring group and 3 from the self-monitoring group), 5 health care providers, and 3 hospital administrators (herein referred to as hospital managers). One remote-monitoring patient and 1 self-monitoring patient were interviewed together with their caregiver. Participants were purposefully sampled in order to capture a meaningful and varied range of views on the program from the parties most intensively involved in its implementation. The patients were recruited by a respiratory therapist at the host site and were selected from the randomized controlled trial's group of participants. These patients' contact details were passed onto the research team. The patients were contacted for participation by the principal author (FvL), either over the phone or through email. Prior to the interview, all participants provided informed consent to participate and were informed that they could decline or withdraw from the study at any time without any consequence to their health care. Written or verbal consent was obtained from all participants by the principal author (FvL) at the start of the interview. All patients had moderate (Global Initiative for Chronic Obstructive Lung Disease level 2) to very severe (Global Initiative for Chronic Obstructive Lung Disease level 4) COPD [20]. Patients' demographics were obtained from the randomized controlled trial's onboarding questionnaires.

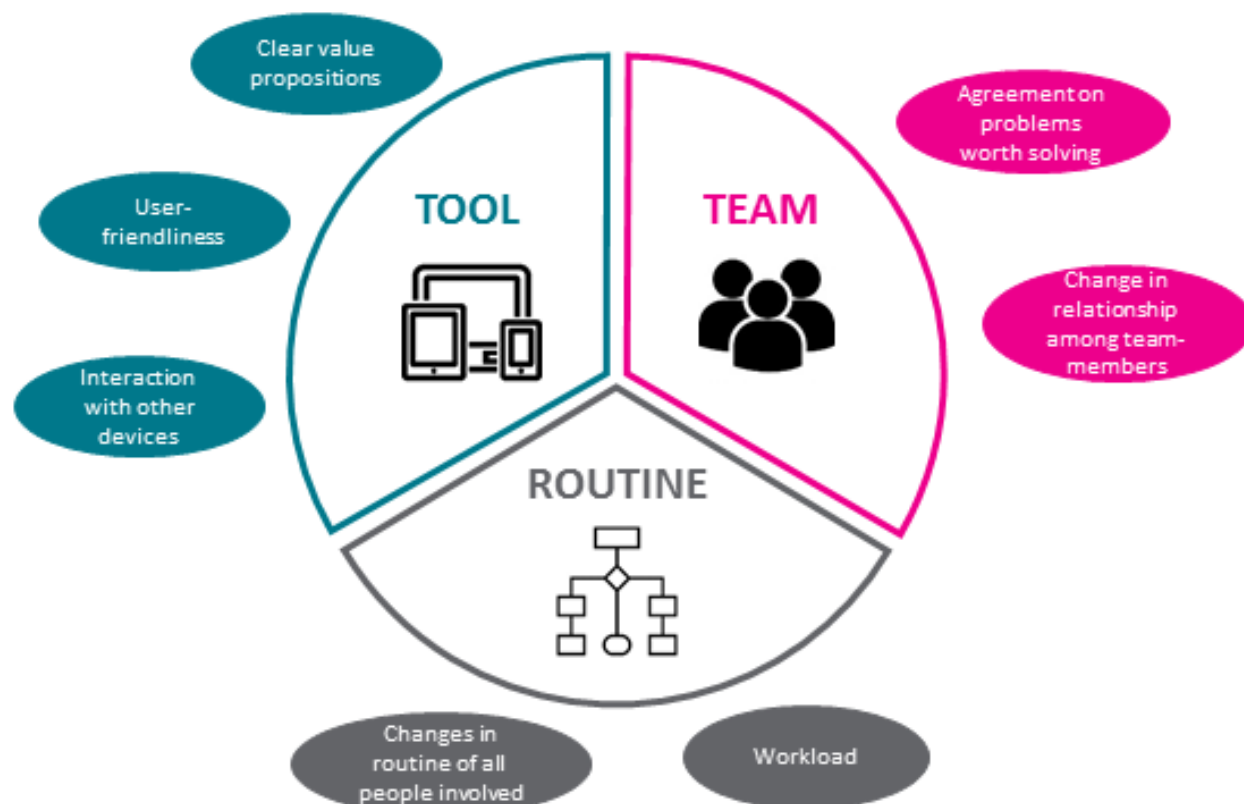
## The Framework

The framework used in this research study was the Tool+Team+Routine framework [17]. Unlike other implementation frameworks, this framework focuses explicitly on service design. Some key ideas that characterize service design in health care are

1. Focus on all the changes to services that are required when introducing a new digital health tool in a health care delivery process.
2. A technology is not immutable; it may need to be modified in order to be of most use to the health care delivery process.
3. There needs to be clear value to all stakeholders, so-called *value propositions* [21]. Ideally, the new service improves one or more factors of the health care quadruple aim—enhancing patient experience, improving population health, reducing costs, and improving the work life of health care providers [22].

Implementing these principles of service design in health care, the Tool+Team+Routine framework is a simple, heuristic approach for introducing new digital health tools to a health care delivery process, or re-designing the delivery process overall and focuses on 3 main factors (Figure 1; Table 1).

**Figure 1.** Visual representation of the main factors in the Tool+Team+Routine framework.



**Table 1.** Tool+Team+Routine framework.

Factor	Description	Questions
Tool	Determines how useful and acceptable the technology is	Are the value propositions clear for all the stakeholders that will use the tool? Is the tool user-friendly? In what way does the tool interact with other devices?
Team	Investigates what changes will occur to the whole team and the way they interact with each other	Does the whole team agree on the problem to be solved by the technology? How does the introduction of the new system change the relationships among team-members, including changed relationships between patient and health care provider?
Routine	Considers in what way people's routines change when the new system is introduced	What other changes will occur in the routines and day-to-day lives of all the people involved—intentionally or unintentionally—when the new system is introduced?

## Data Collection Methods

Participants were interviewed by the principal author (FvL). The interviews were semistructured and contained questions pertaining to the 3 main concepts of the framework. Each group of participants (patients, health care providers, and hospital managers) were interviewed using a different interview guide. Interviews lasted 30 to 60 minutes. All patient interviews were conducted over the phone. The interviews with the health care providers and hospital managers were conducted in person, except for 2 that were conducted by phone. All interviews were audiorecorded and transcribed verbatim. The transcripts were anonymized and deidentified.

## Data Analysis

The interview data were analyzed using thematic analysis methods [23,24]. The first 2 transcripts were analyzed and coded individually by the principal author and 2 other members of the research team. The codes were then compared and discussed which informed the development of a code book that was used to code 3 other transcripts. Thereafter, the code book was

discussed again between the same 3 researchers, which resulted in the revised code book. This method was used to ensure the members of the research team contributing to the analysis agreed on the relevance and definitions of the codes being used. The remaining transcripts were analyzed and coded by the principal author using the revised code book. Any necessary changes to the coding of the first 5 transcripts that resulted from the revised code book were made as well, through regularly occurring analysis meetings between 4 authors on the study team.

## Results

### Overview

**Table 2** provides an overview of the patients and caregivers who participated.

**Table 3** provides an overview of the health care providers and hospital managers who participated. The tables provide the overall attitude toward the program (negative, neutral, mixed, or positive) as gleaned from participant interviews.

**Table 2.** Patients and caregivers who participated.

Participant	Study arm (remote or self-monitoring)	Gender	Age (years)	Interviewed with caregiver	Overall attitude toward program
P1	Remote monitoring	Male	73	No	Positive
P2	Remote monitoring	Female	89	No	Positive
P3	Self-monitoring	Female	69	No	Neutral
P4	Remote monitoring	Female	72	No	Neutral
P5	Remote monitoring	Male	65	No	Positive
P6	Self-monitoring	Female	75	No	Positive
P7	Remote monitoring	Male	73	Yes	Positive (patient); positive (caregiver)
P8	Self-monitoring	Female	81	Yes	Neutral (patient); positive (caregiver)

**Table 3.** Overview of the health care providers and hospital managers who participated.

Participant label	Job title	Level of interaction with Cloud DX <sup>a</sup>	Overall attitude toward program
HCP 1	Physician	3	Positive
HCP 2	Physician	3	Neutral
HCP 3	Allied health professional	5	Mixed
HCP 4	Allied health professional	5	Positive
HCP 5	Allied health professional	4	Positive
HM 1	Senior manager	2	Positive
HM 2	Manager	2	Positive
HM 3	Senior manager	1	Positive

<sup>a</sup>Levels—1 (very little): no interaction with the Cloud DX data on a daily basis and little to moderate understanding of the tool; 2 (little): no interaction with the Cloud DX data on a daily basis but moderate to high understanding of the tool; 3 (moderate): regular indirect interaction with the Cloud DX data but not involved in daily monitoring of patients; 4 (high): daily direct interaction with the Cloud DX data but not involved in daily monitoring of patients; 5 (very high): daily monitoring of patients with the Cloud DX tool and data.

In the sections below, qualitative descriptions of the results that emerged from the interviews are given. These results are structured around the 3 main themes of the framework: *Tool*, *Team*, and *Routine*. Verbatim quotations are labeled with the role of the participant and their study arm (for patient and caregiver participants) or their role and level of interaction with the Cloud DX tool (for health care provider and hospital manager participants).

### Attitude Toward the Project

Within all stakeholder groups, the attitude toward the project was mainly positive. Of the remote-monitoring patients, 4 out of 5 viewed the program as positive; the fifth had a neutral attitude toward the program. Of the self-monitoring patients, 1 out of 3 viewed the program as positive; the other 2 had neutral feelings about it. The 2 caregivers were both positive about the program. Within health care provider and hospital manager groups, attitudes toward the program were more nuanced. All 3 hospital managers showed positive attitudes toward the program. Of the health care providers, 3 out of 5 showed positive attitudes as well; the other 2 had mixed feelings about the program. These are elaborated.

### Tool: Value Propositions in a Technology-Push Project

Stakeholders' perceptions of the value of the tool varied and were characterized by the fact that the implementation project was structured using a *technology-push* approach. In this case, a team of managers and providers from 2 different health care organizations established a project agreement with the vendor, securing funding to support an evaluation of the tool in a hospital environment. One manager explained expectations for the project:

*I think [the technology] is allowing patients to catch their symptoms, uh sooner so they can start treatment before they get really sick. I think it's keeping them away from the emergency department and it should be decreasing the number of admissions... So, decreasing admissions to hospital, overall, it's gotta be improving their quality of life. [Senior manager, very little interaction]*

Only after the funding was in place and the agreement to complete the project was secured did the project team seek to engage the health care providers who would be using the tool in the newly established model of care. As a result, not all health care providers saw value in the tool, leading some to resist engaging with the technology in a meaningful way.

Resistance to engaging with the technology and its outputs was reported by some of the respiratory physicians who were involved in the care of patients enrolled in the program. One physician stated:

*I was very skeptical when the study was first started because uh, a lot of the times, these sort of home monitoring or self-monitoring or other programs kind of invent the technology first before asking a lot of important clinical questions of whether it's actually gonna be of benefit or uh, examining why patients actually have exacerbations or the mechanisms and then it's sort of just implemented into this study and then we see what happens, so, it's uh, a bit of the uh, cart before the horse kind of scenario [Physician, moderate interaction]*

This quotation illustrates the general lack of engagement with the tool among some physicians primarily as a result of feeling that the tool does not solve a direct clinical need (although the lack of time physicians have for such initiatives is also clearly acknowledged). The primary example of this was that vitals data, such as blood oxygenation saturation, which are a primary data output of the tool for health care providers are known to be a poor predictor of exacerbations of COPD. Physicians were not satisfied with the effort to establish a meaningful use case for the tool after the decision to move forward with the project was made. This created a scenario where physicians had widely varying levels of engagement with the lead respiratory therapist who provided them with information about the outputs of the tool in order to inform their approach to care.

In this initiative, the physicians were not actually required to interact directly with the tool. Instead, a respiratory therapist was responsible for communicating with patients by phone regarding the tool and managing the inputs arising from patient

data being collected at home. This respiratory therapist would communicate with physicians when remote-monitoring data from patients under their care suggested that some form of physician follow-up would be appropriate. Although the respiratory therapist had reservations about the workload implications of the tool, this person believed there would be a benefit for certain patients:

*So the program I think it's helping patients, ehm, better understand their condition, or at least some of the patients. Perhaps not all of them. And helping them to recognize when they're getting sick so that they can get treatment, either sooner or more appropriately... I would say that some patients did, or were good self-recognizers [already] and those are patients who have come to the hospital one or two times before because they had experienced [an exacerbation]... However those who haven't had many experiences with exacerbations, it helps them better recognize it. [Allied Health Professional, very high interaction]*

The patients and caregivers interviewed for the study reported some technology-related challenges that might be expected when adopting new digital tools for the first time. However, many participants reported generally positive sentiments about the tool and its adoption into their everyday lives. One patient, a woman in her 70s who lived alone, stated:

*The tablet gave me, just that confidence because I could see that my blood pressure was in normal range though I kind of figured that. And then my oxygen levels, I didn't know about my oxygen level and then, once I, I got to reading my oxygen levels, everything seems to be fine. So that, that was just, I don't know, just a report back to myself saying "Hey, you're not too bad at all" [Patient, remote monitoring]*

Another patient described an instance where her decision to seek out emergency services was directly informed by the tool. Although this example is not directly related to a patient's experience of COPD, people living with COPD often experience co-occurring conditions such as atrial fibrillation.

*[Respondent:] ... my heart was beating too fast and because I had the, eh, or it was getting, although it was indoors it seemed to be getting very dark and, I had ehm, and my blood pressure I don't know, I don't know what my blood pressure was but my pulse was, was going about a hundred and, at a hundred and seventy beats a minute and, I realised that was way too fast and I felt sort of dizzy so I went into emerge that time...*

*[Interviewer:] Alright. (.) So, did you feel like the Cloud DX tablet helped you at [that] incident?*

*[Respondent:] Yes, because like I didn't, I was, I guess I was sort of dizzy I didn't realize my pulse, when I saw my pulse was so high I knew I had to do something about it. Otherwise, I might not have gone in and I don't what would've happened then. [Patient, remote monitoring]*

In summary, although the value propositions of the tool were not immediately clear to all stakeholders, those who engaged with the tool directly did feel the tool was valuable for the management of COPD, at least in certain cases. However, the primary value that appeared to be offered by the program arose not from the tool per se but from the relationship between patients and the respiratory therapist responsible for administering the program. We explore this point in the next section.

### Team: The Role of Relationships in the Reconfigured Service

The relationships between members of the health care team did not change in meaningful ways as a result of the implementation of the tool and the establishment of the remote-monitoring service. The respiratory therapist who administered the program was the primary point of contact for patients and would notify physicians, when appropriate, regarding issues arising for patients under their care. This process was not substantially different from how these providers communicated previously, as respiratory therapists had opportunities to interact with some patients before the beginning of this project in the context of a community-based exercise program and as recurring patients at the Centre for Respiratory Health and would communicate with physicians under similar circumstances as during the remote-monitoring project.

The relationships that were most important were those between the respiratory therapist administering the program and the patients enrolled in the program. Specifically, the relationships between the respiratory therapist and patients in the remote-monitoring program deepened and became an important component of the patients' overall care. Many remote-monitoring patients spoke about the importance of this particular respiratory therapist for the value they experienced with the program, viewing the respiratory therapist not only as a source of information related to the tool itself, and COPD more generally, but as a person capable of supporting patients in solving health-related challenges of virtually any kind. The respiratory therapist corroborated this view by indicating that she had a good relationship especially with patients in the remote-monitoring group of the project:

*So a lot of the patients call me because they know I'll answer and it's just to help get a prescription filled or sort of some other issue that they're having... cause they know I'll answer and have access to their physicians. They'll call me over calling [the doctor's] receptionist... Which maybe isn't good, but, is also kind of good for [patients] because they feel more comfortable. Cause the relationship is there. [Allied Health Professional, very high interaction]*

Patients also reported the importance of being able to make contact with the respiratory therapist when questions arose regarding their health and health care. When asked what the most useful feature of the technology was, one patient stated:

*Well, the point that you can, ehm, call somebody whenever you have a question, like [the respiratory therapist] you know. It's very helpful because you*



*have somebody to talk to about whatever the problem is. Whether it's your breathing or your heart or what.*  
[Patient, remote monitoring]

These examples illustrate a fundamental mechanism by which the newly established service exhibited value for patients in the remote-monitoring group: through the relationship with the respiratory therapist administering the program. In this sense, although the tool was able to support the development of confidence and elicit appropriate responses to changing symptoms for patients across both self-monitoring and remote-monitoring groups, the value of the program extended beyond supporting the ability of patients to manage their COPD independently for the remote-monitoring group; it evolved to include a hidden care coordination function that was being performed by the respiratory therapist.

This care coordination function was not intended to be a part of the program from the outset and could be viewed as being a function of technological scope creep as a result of this respiratory therapist's strong relationships with patients. Specifically, we use *technological scope creep* to refer to the unanticipated expanding scope of the respiratory therapists' work activities in relation to the function of the tool. The positive relationship between the respiratory therapist and remote-monitoring patients perhaps boosted their engagement with the tool but also fundamentally changed the nature of the remote-monitoring program. It was no longer simply a matter of monitoring vital signs and symptoms related to COPD, but at least for some patients, became a strategy to coordinate and access health services well beyond the original boundaries of the program. This technological scope creep is one component of the ways in which routines evolved as the program was implemented, which is addressed in the final section of the findings.

### Routine: Routines of Care and Hidden Work

The demand for changes in routines of care was very different for the different people involved in the project. For physicians, they had to change virtually nothing, and independently chose whether or not to act upon any information provided by the respiratory therapist administering the project. This meant that some physicians would act directly on the information provided and reach out to patients to follow up, whereas others would simply note the information but proceed with their usual approach to care.

Patients underwent a learning process as they came to interact with the new tool that became a part of their home environment during the project period. Although patients suggested that the learning process was a challenge in some cases, ultimately, the everyday use of the tool was viewed as largely unproblematic. One patient explained:

*... I would make sure to get up a few minutes earlier than the time I needed to be up and kind of on my way out, to make sure I do have the time to do it. It's not a big deal.* [Patient, self-monitoring]

In contrast to the relative ease experienced by physicians and patients, the respiratory therapists directly involved in the new program experienced drastic changes to their routines. In a

sense, these changes were obvious and expected, as both respiratory therapists appointed to the program were fully aware that the program was just being established and represented net new work. However, the nature of the work they were ultimately required to carry out was not clear to them in advance. Two examples illustrate this point: extra documentation requirements and the growing communication demands arising from patient input through the tool.

In relation to documentation, the respiratory therapists were initially required to chart patient notes in the tool's record system related to any care-related activity that took place through the tool. However, it quickly became apparent that this meant double charting and would not be sustainable were the program to become a part of usual care. One respiratory therapist described this as follows:

*What we did was we actually charted in some of the patients' notes on the Cloud DX project, sort of things that we thought were relevant. But then we found out we were actually double charting. Ehm, because we would have to chart in, you know, the EMR system. And so, I think that, you know, if it could be integrated into the system, like at the hospital, ehm that would be awesome. But at, at present, during the study, we, you know, we had to sort of do extra work just cause it wasn't integrated.* [Allied health professional, very high interaction]

In this case, the known challenge of the lack of interoperability led to excessive workload on the part of the health care providers administering the program.

In relation to the communication demands arising from the program, these extended beyond the issue of patients calling the respiratory therapist for support in coordinating care more generally. The respiratory therapist also received calls related to technological challenges that were supposed to be directed to the vendor. Furthermore, the respiratory therapist would receive notifications indicating an exacerbation based on data that had been incorrectly entered and would need to call patients to follow up on those notifications only to discover that something had gone wrong in the data collection process (eg, they had responded to a survey incorrectly or the blood pressure cuff was not tightened properly). The respiratory therapist administering the program stated:

*I can tell you that, ehm, in the remote monitoring group I had 116 incoming calls in 6 months and the self-monitoring group I had like 134 incoming calls in six months. So the total is, what, 240? So the remote-monitoring called me a little less than the self-monitoring. Ehm, and then I further broke that down into the reason for them calling and I'd say it's like 77 were technological in the remote-monitoring and like 80 or so were technological [in self-monitoring]. And then the next highest one was just general health inquiries. Yeah, so, it's, it's mostly tech issues that they're calling about. And they are told to call [the vendor], they just feel more comfortable calling me. And I generally just push it over to [the vendor]. In addition to those incoming*



*calls, I had the notifications to respond to. So [sighs] I think in the three months for the remote monitoring group I had something, like, close to 300 notifications in six months, to respond to. [Allied health professional, very high interaction]*

The additional workload represented by these new routines was also not anticipated by managers. A manager responsible for considering the workload implication and sustainability of the program in the longer term commented:

*[An important question is] if it's something that's sustainable and how do we integrate it into the day to day clinic activities? This study did not address that. So that will be something we'll have to figure out. [Senior manager, little interaction]*

These instances of additional workload represent the establishment of new routines of care that were not anticipated at the outset of the program. These new routines quickly filled the time of the respiratory therapists responsible for the program, and added demands that would not be sustainable if the resources for staffing were held constant in the long-term. This finding represents the hidden work of technology implementation and remote-monitoring tools, which is not visible in advance of a program and might not be visible at all unless explicitly examined during the evaluation process.

## Discussion

### Principal Findings

Our findings direct attention toward two primary insights. Each insight helps to address the two primary objectives of our paper, which were to advance the Tool+Team+Routine heuristic for the adoption of digital health technologies and to clarify mechanisms by which remote monitoring may exert its value for patients. The first insight is the significance of interpersonal relationships for the functioning of this particular digital health technology in the context of delivering a comprehensive service of remote monitoring. The second is the expansion of routines of care and what we refer to as technological scope creep—the growing demands of health care providers' everyday work as the technology requires new actions to fulfil newly generated demands.

### The Role of Relationships in Remote-Monitoring Initiatives

In our study, we found that one of the features of the service that most drove the successful engagement of patients with the technology was the relationship between the respiratory therapist administering the program and the patients who were enrolled. This finding resonates strongly with previous research [10,25] on the mechanisms by which digital health services have their effects. Vassilev et al [26] completed a realist review of the literature examining the specific causal mechanisms influencing the effectiveness of telehealth programs and found relationships to be one of three crucial mechanisms. They explained that new technologies introduce “new sources of relationality [26],” and thereby, “restructure existing relationships [26].”

In support of previous literature [7], our findings outline the value of such interpersonal relationships between the provider administering the program and the patients using the technology for the successful adoption of the service overall. In relation to the Tool+Team+Routine framework, this is an important point to emphasize; it is not just that such relationships enhance the adoption of the new technology, but that the relationships drive the development of an actual service arising from the technology itself. The technology is one component of a new program but relies on many other interacting components to become a part of a successfully delivered service (including people, technologies, and newly established processes).

In the context of service design, the adoption of new technologies stands to influence interpersonal relationship in at least two distinct ways. First, the technology might influence relationships among health care providers as they are required to collaborate in different ways than they previously had. We did not find this to be the case in our study, likely because the expectations that different health care providers had for one another did not change throughout the study period. However, this change to relationships among health care providers remains a possibility [25]. Second, the mechanisms by which the technology has its effects for patients are themselves determined in part by the quality of the relationships established between patients and the provider administering the program, and in this case, mediated by the technology. Emphasizing these relationships represents an important conceptual development for the Tool+Team+Routine heuristic, placing relationships in a more prominent position when taking a service design approach to promoting the adoption of technology-enabled service innovations. This point also reinforces the finding in the literature that the interpersonal relationship between the patient and the health care providers is a primary consideration in the effort to understand the ways in which remote monitoring has its effects [25].

### Technological Scope Creep

The second important finding emerging as an important direction for future research in our study is the phenomenon of technological scope creep. This point relates specifically to the *routine* element of the Tool+Team+Routine heuristic in a service design context, as it refers to changes in the routines of providers who must interact with the technology in ways that were not predicted at the outset. Furthermore, these newly added routines, in our study, were not accounted for in the time allocations assigned to providers responsible for administering the program via the technology, even though they might add additional value for patients. This point raises an important issue for the Tool+Team+Routine heuristic: in the effort to promote adoption of technologies for new services or service re-design, explicit effort should be made to anticipate technological scope creep and the demand for new routine practices that were not expected of the technology at the outset.

This point resonates with past work outlining how routine work practices change when new technologies are adopted [16,27,28]. However, past work has primarily emphasized the point that technologies restructure service arrangements [29]. We acknowledged this point in our past work [17] but extend this

point now to acknowledge that even where work restructuring is explicitly anticipated in technology adoption projects, new technologies have a further potential to produce new work that is hidden from the view of management and evaluation as a result of not being considered as falling within the scope of the service itself. The technology evaluated in this study was intended to enable remote and self-monitoring of patients related specifically to COPD; however, patients ended up using the service as a general care coordination service. As a result of the relationships with the primary program administrator (the respiratory therapist), the scope of the service expanded to include this more general care coordination function. In effect, as a result of technological scope creep, the service itself expanded from its original function to take on a broader set of functions.

### Strengths and Limitations

This paper has applied a previously reported framework [17] to understand the implementation of a remote-monitoring

technology into a health care environment. Although the patient participants recruited for this study might not reflect the needs and opinions of the entire COPD population, the sample of patients included provided in-depth insight into the value of the tool in the context of the service.

### Conclusions

The observations arising from our findings advance the Tool+Team+Routine heuristic for use in service design activities that involve technology adoption in health care services. Deep thinking about how the specific individuals administering the program will work to build relationships and how those relationships and other sources of activity might lead to technological scope creep present two often unanticipated challenges when implementing digital technologies in health care. Future work can reconcile these important influences on technology adoption and will examine strategies to account for them in advance of initiating technology implementation projects using a service design approach.

### Acknowledgments

We gratefully acknowledge all the participants who took part in the study. This work was supported by Ontario Centres of Excellence Health Technologies Fund (grant number 27009). The grant supported personnel and materials related to the implementation of the technology into the clinic and the evaluation of the intervention.

### Conflicts of Interest

None declared.

### References

1. Global digital health market by major segment 2015-2020 | Statistic Internet. Statista. URL: <https://www.statista.com/statistics/387867/value-of-worldwide-digital-health-market-forecast-by-segment/> [accessed 2018-09-20]
2. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. *Health Aff (Millwood)* 2011 Mar;30(3):464-471. [doi: [10.1377/hlthaff.2011.0178](https://doi.org/10.1377/hlthaff.2011.0178)] [Medline: [21383365](https://pubmed.ncbi.nlm.nih.gov/21383365/)]
3. WHO. Compendium of new and emerging health technologies. 1st ed. Geneva: World Health Organization; 2011.
4. Huckvale K, Wang CJ, Majeed A, Car J. Digital health at fifteen: more human (more needed). *BMC Med* 2019 Mar 18;17(1):62 [FREE Full text] [doi: [10.1186/s12916-019-1302-0](https://doi.org/10.1186/s12916-019-1302-0)] [Medline: [30879466](https://pubmed.ncbi.nlm.nih.gov/30879466/)]
5. Fairbrother P, Pinnock H, Hanley J, McCloughan L, Sheikh A, Pagliari C, TELESCOT programme team. Continuity, but at what cost? The impact of telemonitoring COPD on continuities of care: a qualitative study. *Prim Care Respir J* 2012 Sep;21(3):322-328 [FREE Full text] [doi: [10.4104/pcrj.2012.00068](https://doi.org/10.4104/pcrj.2012.00068)] [Medline: [22875143](https://pubmed.ncbi.nlm.nih.gov/22875143/)]
6. Kenealy TW, Parsons MJG, Rouse APB, Doughty RN, Sheridan NF, Hindmarsh JKH, et al. Telecare for diabetes, CHF or COPD: effect on quality of life, hospital use and costs. A randomised controlled trial and qualitative evaluation. *PLoS One* 2015 Mar 13;10(3):e0116188 [FREE Full text] [doi: [10.1371/journal.pone.0116188](https://doi.org/10.1371/journal.pone.0116188)] [Medline: [25768023](https://pubmed.ncbi.nlm.nih.gov/25768023/)]
7. Nissen L, Lindhardt T. A qualitative study of COPD-patients' experience of a telemedicine intervention. *Int J Med Inform* 2017 Nov;107:11-17. [doi: [10.1016/j.ijmedinf.2017.08.004](https://doi.org/10.1016/j.ijmedinf.2017.08.004)] [Medline: [29029687](https://pubmed.ncbi.nlm.nih.gov/29029687/)]
8. Vatnøy TK, Thygesen E, Dale B. Telemedicine to support coping resources in home-living patients diagnosed with chronic obstructive pulmonary disease: Patients' experiences. *J Telemed Telecare* 2017 Jan;23(1):126-132. [doi: [10.1177/1357633X15626854](https://doi.org/10.1177/1357633X15626854)] [Medline: [26853926](https://pubmed.ncbi.nlm.nih.gov/26853926/)]
9. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Hinder S, Rouncefield M. What is quality in assisted living technology? The ARCHIE framework for effective telehealth and telecare services. *BMC Med* 2015 Apr 23;13:91 [FREE Full text] [doi: [10.1186/s12916-015-0279-6](https://doi.org/10.1186/s12916-015-0279-6)] [Medline: [25902803](https://pubmed.ncbi.nlm.nih.gov/25902803/)]
10. Greenhalgh T, Wherton J, Papoutsis C, Lynch J, Hughes G, A'Court C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. *BMC Med* 2018 May 14;16(1):66 [FREE Full text] [doi: [10.1186/s12916-018-1050-6](https://doi.org/10.1186/s12916-018-1050-6)] [Medline: [29754584](https://pubmed.ncbi.nlm.nih.gov/29754584/)]
11. Avgar A, Litwin A, Pronovost P. Drivers and barriers in health IT adoption: a proposed framework. *Appl Clin Inform* 2012;3(4):488-500 [FREE Full text] [doi: [10.4338/ACI-2012-07-R-0029](https://doi.org/10.4338/ACI-2012-07-R-0029)] [Medline: [23646093](https://pubmed.ncbi.nlm.nih.gov/23646093/)]

12. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). *Implement Sci* 2016 Oct 26;11(1):146-149 [[FREE Full text](#)] [doi: [10.1186/s13012-016-0510-7](#)] [Medline: [27782832](#)]
13. Acharya S, Werts N. Toward the design of an engagement tool for effective electronic health record adoption. *Perspect Health Inf Manag* 2019;16(Winter):1g [[FREE Full text](#)] [Medline: [30766458](#)]
14. Peek STM, Wouters EJ, Luijkx KG, Vrijhoef HJ. What it Takes to Successfully Implement Technology for Aging in Place: Focus Groups With Stakeholders. *J Med Internet Res* 2016 May 03;18(5):e98 [[FREE Full text](#)] [doi: [10.2196/jmir.5253](#)] [Medline: [27143097](#)]
15. Keith R, Crosson J, O'Malley AS, Crompton D, Taylor E. Using the Consolidated Framework for Implementation Research (CFIR) to produce actionable findings: a rapid-cycle evaluation approach to improving implementation. *Implement Sci* 2017 Feb 10;12(1):15-10 [[FREE Full text](#)] [doi: [10.1186/s13012-017-0550-7](#)] [Medline: [28187747](#)]
16. May C. Towards a general theory of implementation. *Implement Sci* 2013 Feb 13;8:18-13 [[FREE Full text](#)] [doi: [10.1186/1748-5908-8-18](#)] [Medline: [23406398](#)]
17. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond "implementation": digital health innovation and service design. *NPJ Digit Med* 2018;1:48 [[FREE Full text](#)] [doi: [10.1038/s41746-018-0059-8](#)] [Medline: [31304327](#)]
18. Stamenova V, Liang K, Yang R, Engel K, van Lieshout F, Lalingo E, et al. Technology-Enabled Self-Management of Chronic Obstructive Pulmonary Disease With or Without Asynchronous Remote Monitoring: Randomized Controlled Trial. *J Med Internet Res* 2020 Jul 30;22(7):e18598 [[FREE Full text](#)] [doi: [10.2196/18598](#)] [Medline: [32729843](#)]
19. Connected Health Kit with Telemedicine Internet. Cloud DX. URL: <https://www.cloudDX.com/#/healthkittelemedicine> [accessed 2018-10-03]
20. GOLD COPD Stages Internet. Lung Institute. 2016. URL: <https://lunginstitute.com/blog/gold-copd-stages/> [accessed 2018-11-22]
21. Osterwalder A, Pigneur Y, Bernarda G, Smith A. Value Proposition Design: How to Create Products and Services Customers Want. Hoboken: John Wiley & Sons; 2014.
22. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med* 2014 Nov 10;12(6):573-576. [doi: [10.1370/afm.1713](#)] [Medline: [25384822](#)]
23. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. *Health Serv Res* 2007 Aug;42(4):1758-1772 [[FREE Full text](#)] [doi: [10.1111/j.1475-6773.2006.00684.x](#)] [Medline: [17286625](#)]
24. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. *APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological*. Washington: American Psychological Association; 2012.
25. Shaw J, Shaw S, Wherton J, Hughes G, Greenhalgh T. Studying Scale-Up and Spread as Social Practice: Theoretical Introduction and Empirical Case Study. *J Med Internet Res* 2017 Jul 07;19(7):e244 [[FREE Full text](#)] [doi: [10.2196/jmir.7482](#)] [Medline: [28687532](#)]
26. Vassilev I, Rowsell A, Pope C, Kennedy A, O'Cathain A, Salisbury C. Assessing the implementability of telehealth interventions for self-management support: a realist review. *Implementation Science* 2015;1-9 [[FREE Full text](#)] [doi: [10.1186/s13012-015-0238-9](#)]
27. Nicolini D. The work to make telemedicine work: a social and articulative view. *Soc Sci Med* 2006 Jun;62(11):2754-2767. [doi: [10.1016/j.socscimed.2005.11.001](#)] [Medline: [16343724](#)]
28. Nicolini D. Stretching out and expanding work practices in time and space: The case of telemedicine. *Human Relations* 2016 Apr 22;60(6):889-920. [doi: [10.1177/0018726707080080](#)]
29. Cresswell K, Worth A, Sheikh A. Actor-Network Theory and its role in understanding the implementation of information technology developments in healthcare. *BMC Med Inform Decis Mak* 2010 Nov 01;10:67-10 [[FREE Full text](#)] [doi: [10.1186/1472-6947-10-67](#)] [Medline: [21040575](#)]

---

## Abbreviations

**COPD:** chronic obstructive pulmonary disease.

---

*Edited by G Eysenbach; submitted 06.02.20; peer-reviewed by L Friesen, E Widnall, E van der Velde; comments to author 23.04.20; revised version received 06.06.20; accepted 25.06.20; published 09.10.20.*

Please cite as:

van Lieshout F, Yang R, Stamenova V, Agarwal P, Cornejo Palma D, Sidhu A, Engel K, Erwood A, Bhatia RS, Bhattacharyya O, Shaw J

*Evaluating the Implementation of a Remote-Monitoring Program for Chronic Obstructive Pulmonary Disease: Qualitative Methods from a Service Design Perspective*

*J Med Internet Res* 2020;22(10):e18148

URL: <https://www.jmir.org/2020/10/e18148>

doi: [10.2196/18148](https://doi.org/10.2196/18148)

PMID: [33034565](https://pubmed.ncbi.nlm.nih.gov/33034565/)

©Florence van Lieshout, Rebecca Yang, Vess Stamenova, Payal Agarwal, Daniel Cornejo Palma, Aman Sidhu, Katrina Engel, Adam Erwood, R Sacha Bhatia, Onil Bhattacharyya, James Shaw. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 09.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Maximizing Telerehabilitation for Patients With Visual Loss After Stroke: Interview and Focus Group Study With Stroke Survivors, Carers, and Occupational Therapists

Stephen Dunne<sup>1</sup>, BSc, PhD; Helen Close<sup>2</sup>, PhD; Nicola Richards<sup>3</sup>, BSc, MBBS; Amanda Ellison<sup>4</sup>, BA (Mod), PhD; Alison R Lane<sup>4</sup>, BSc, MA, PhD

<sup>1</sup>School of Psychology, University of Sunderland, Sunderland, United Kingdom

<sup>2</sup>Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

<sup>3</sup>Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

<sup>4</sup>Department of Psychology, Durham University, Durham, United Kingdom

**Corresponding Author:**

Stephen Dunne, BSc, PhD

School of Psychology

University of Sunderland

Chester Road

City Campus

Sunderland, SR1 3SD

United Kingdom

Phone: 44 01915153070

Email: [stephen.dunne@sunderland.ac.uk](mailto:stephen.dunne@sunderland.ac.uk)

## Abstract

**Background:** Visual field defects are a common consequence of stroke, and compensatory eye movement strategies have been identified as the most promising rehabilitation option. There has been a move toward compensatory telerehabilitation options, such as the Durham Reading and Exploration (DREX) training app, which significantly improves visual exploration, reading, and self-reported quality of life.

**Objective:** This study details an iterative process of liaising with stroke survivors, carers, and health care professionals to identify barriers and facilitators to using rehabilitation tools, as well as elements of good practice in telerehabilitation, with a focus on how the DREX package can be maximized.

**Methods:** Survey data from 75 stroke survivors informed 12 semistructured engagement activities (7 focus groups and 5 interviews) with 32 stroke survivors, 10 carers, and 24 occupational therapists.

**Results:** Thematic analysis identified key themes within the data. Themes identified problems associated with poststroke health care from both patients' and occupational therapists' perspectives that need to be addressed to improve uptake of this rehabilitation tool and telerehabilitation options generally. This included identifying additional materials or assistance that were required to boost the impact of training packages. The acute rehabilitation setting was an identified barrier, and perceptions of technology were considered a barrier by some but a facilitator by others. In addition, 4 key features of telerehabilitation were identified: additional materials, the importance of goal setting, repetition, and feedback.

**Conclusions:** The data were used to try to overcome some barriers to the DREX training and are further discussed as considerations for telerehabilitation in general moving forward.

(*J Med Internet Res* 2020;22(10):e19604) doi:[10.2196/19604](https://doi.org/10.2196/19604)

**KEYWORDS**

telerehabilitation; vision; barriers; facilitators; technology



## Introduction

Visual impairments occur in about 65% of stroke survivors [1], with a large proportion of these experiencing partial visual field loss [2] (eg, blindness in a part of their visual field). For such individuals, visual restoration is limited [3]. Partial visual loss can be debilitating, with those who experience it having problems completing activities of daily living [4], an unwillingness to leave the home [5], and an increased risk of falling [6]. Individuals also have their driving licenses revoked, further reducing their independence, and can consequently become more isolated, reliant on support, and depressed [5,7-9]. Overall, people with visual field defects demonstrate poorer functional outcome [10], impaired vision-related quality of life [11], and reduced engagement in rehabilitation [12]. With improving stroke survival rates [13], an increasing number of stroke survivors are living with the long-term consequences of partial visual loss. Therefore, providing effective and accessible treatment for such individuals is important in order to significantly reduce their disability.

The uptake of technology among older people is relatively high, with 48% of those aged 55 years or older owning a tablet computer in 2017 and 8.1 million 55- to 64-year-olds in the United Kingdom using apps regularly [14]. Furthermore, telerehabilitation, the provision of rehabilitation services (therapeutic intervention, progress monitoring, education, training, etc) at a distance via electronic information and communication technologies [15], has been identified as a credible, potentially cost-effective future direction for health and social care [16,17]. Despite this, there is still a significant contingent of stroke survivors who distrust telerehabilitation, as they fear reduced quality of care [17]. Previous evidence has shown that motivation is a significant factor in improving engagement with home-based rehabilitation tools [18]. At present, only a limited number of studies exist regarding the effectiveness of telehealth-based digital interventions [19]. For these reasons, it is important to examine factors that could increase their uptake.

The most common form of telerehabilitation employed to address visual loss is computer-based compensatory training, which teaches stroke survivors adaptive eye movement strategies to cope more effectively with visual loss. Durham Reading and Exploration Training (DREX) is one such version of this training, taking the form of a free, multiplatform app and significantly improving visual exploration, reading, and quality of life [20] (see [Multimedia Appendix 1](#) for further details). We report findings from our 2-stage study with stroke survivors, carers, and health professionals, which explored potential barriers and facilitators to the use of DREX and perceptions about key features for telerehabilitation packages. Stage 1 described key findings from a survey identifying the key elements of patients' stroke experiences. In stage 2, these elements were addressed through a process of liaising with key stakeholders (stroke survivors, carers, and occupational therapists) to understand experiences of this specific tool (DREX) and adjusting the training package in response to the issues identified, thereby improving the potential for training uptake and user experience.

## Methods

### Recruitment

A snowball sample of 75 UK-based stroke survivors with partial visual loss completed the survey (mean age 63.77 years). The survey was distributed in an online format via prominent stroke Facebook groups (n=66; mean age 57.00 years) or as a paper version upon request (n=9; mean age 66.78 years).

Stroke survivors for interviews and focus groups were selected through opportunity sampling based on geographical location (region of North East England). A total of 32 stroke survivors (18 men; aged 43-83 years, mean age 62.28 years), 10 carers (7 women; 41-75 years, mean age 54.70 years), and 24 occupational therapists (19 women; 22-45 years, mean age 31.13 years) were involved in either focus groups (n=7) or interviews (n=5). Engagement activity characteristics are displayed in [Table 1](#).

**Table 1.** Characteristics of stakeholder engagement activity.

Activity	Stroke survivors, n	Carers, n	Occupational therapists, n	Total, n
Focus group A	4	1	N/A <sup>a</sup>	5
Focus group B	3	2	N/A	5
Interview A	1	1	N/A	2
Interview B	1	1	N/A	2
Focus group C	8	N/A	N/A	8
Focus group D	6	2	N/A	8
Focus group E	6	1	N/A	7
Focus group F	N/A	N/A	12	12
Focus group G	N/A	N/A	12	12
Interview C	1	N/A	N/A	1
Interview D	1	1	N/A	2
Interview E	1	1	N/A	2
Total	32	10	24	66

<sup>a</sup>N/A: not applicable.

## Design

This qualitative study was part of a larger study investigating barriers and facilitators in stroke rehabilitation in general. This data subset explores the interaction between DREX and specific issues in relation to poststroke visual field defects. Full project data and materials can be found on the Open Science Framework [21]. This study was divided into 2 stages. In stage 1, a short survey was used to identify the barriers and facilitators from the stroke survivors' experiences. Stage 1 data informed the second stage, which aimed to quantify the extent of the major themes identified. Interviews and focus groups were conducted with stroke survivors, carers, and occupational therapists in a semistructured manner to understand the specific impact of these barriers and facilitators and the factors perceived as important within telerehabilitation. Stroke survivors were recruited via Stroke Association groups in the northeast of England and could participate in either a focus group or an interview. Participants had to be at least 18 years old and able to give informed consent. Stroke survivors with communication difficulties were not excluded if they had capacity for consent; however, no stroke survivors reported communication difficulties. The information gathered from these activities was used to generate changes to the DREX package to increase accessibility. Ethical approval was provided by the Durham University Psychology Department Ethics Committee and the study was conducted in accordance with the Declaration of Helsinki.

### Stage 1: Survey

Full survey details can be found in [Multimedia Appendix 2](#). Survey data were analyzed to inform key points of interest for discussion within stage 2.

### Stage 2: Focus Groups and Interviews

Survey data permitted the construction of a semistructured interview protocol (see [Multimedia Appendix 3](#) for the interview

schedule), which probed further information from a number of identified areas. Specifically, stroke survivors and carers responded to questions regarding their everyday difficulties, rehabilitation journey, experiences with technology, motivation, and levels of support. Occupational therapists were also asked about their experiences, but with a greater focus on understanding the barriers and facilitators they experience in supporting stroke survivors and their views on the use of technology in poststroke visual rehabilitation.

Stroke survivor focus groups were carried out at one of 5 venues: Age UK Roundhouse Ashington Stroke Group, Gateshead Stroke Association Stroke Group, Newton Aycliffe Stroke Group, Teesside Stroke Club, and Stockton-On-Tees Happy Talk Stroke Club. Interviews were carried out in a quiet room in the survivor's home. The 2 focus groups with occupational therapists took place with the stroke teams at the University Hospital of North Tees and Northumbria Hospital. Discussions were audio recorded with consent.

Focus groups were facilitated by the primary author (an experienced qualitative researcher with clinical experience). Initially, the project and focus group aims were outlined and the participants introduced. The notion of facilitators and barriers were explained, and participants were given the opportunity to discuss their own expectations and clarify any miscommunications. Carers were encouraged to let the stroke survivors give their views first but were also given the opportunity to discuss their own experiences. All participants communicated verbally. The focus groups and interviews were semistructured, with 6 open questions used to stimulate discussion and follow-up questions used depending on responses. The 2 focus groups conducted with occupational therapists instead focused on the DREX training package and how it could be improved, current rehabilitation techniques implemented by therapists dealing with poststroke visual loss, and additional materials that could assist therapists. Focus groups and interviews lasted approximately 60 minutes.

Qualitative methods have proved particularly suitable for the analysis and interpretation of focus group data [22]. Thematic analysis was used [23] to work through the data until a small number of themes that described the data set were identified. Data were managed using Microsoft Word. Specific information regarding the data coding process can be found in [Multimedia Appendix 4](#).

## Results

### Stage 1: Survey Data

Stroke survivors identified family, carers, and occupational therapists as best placed to support them in their recovery. Therefore, these stakeholder groups were included as participants in stage 2. Additionally, confidence with technology and the internet was found to decline increasingly with age, while fear of making mistakes rose. Finally, face-to-face support

was identified as the most useful factor in rehabilitation. A full discussion of these results can be found in [Multimedia Appendix 5](#).

### Stage 2: Interview Data

#### *Qualification and Quantification of Barriers and Facilitators Identified by Participants*

There were 2 main themes and several subthemes that were identified by stroke survivors as factors that act as barriers to using telerehabilitation, while 1 main theme with 1 subtheme was identified as a facilitator. In addition, 4 key features were identified by participants for those designing and creating e-therapies for stroke survivors. These themes are presented in [Textbox 1](#) and discussed below, illustrated with verbatim transcripts. Interestingly, technology was seen as both a barrier and a facilitator, so these are discussed together.

**Textbox 1.** Factors identified by stroke survivors, carers, and occupational therapists as acting as barriers or facilitators in the use of telerehabilitation in stroke and the key features identified as important for telerehabilitation.

<b>Barriers</b>	
1. Acute ward rehabilitation	<ul style="list-style-type: none"> <li>Limited occupational therapy time</li> <li>Acute patient readiness</li> </ul>
2. Perceived disadvantages of technology	<ul style="list-style-type: none"> <li>Confidence</li> <li>Lack of interest</li> <li>Lack of personal contact</li> </ul>
<b>Facilitators</b>	
1. Technology and the internet as a perceived advantage	<ul style="list-style-type: none"> <li>YouTube</li> </ul>
<b>Key features for rehabilitation</b>	
1. Additional resources	
2. Goal setting	
3. Feedback	
4. Repetition	

### *Acute Ward Rehabilitation as a Barrier*

Occupational therapists discussed the limited time they have with stroke survivors in acute settings as a barrier in their rehabilitation. Therapists praised the idea of having a form of therapy they could leave with a stroke survivor to use as and when they want without requiring therapist supervision and time:

*It's also good as well because like obviously we can only provide so much time as well with rehab but if we can give the patient that to do when we're not there then if they're doing even an hour every other day then its brilliant. Sometimes we give them things*

*to do and think, "Are they going to do it or not?" but something like that they know it's there, they do it, log on.*

However, occupational therapists also identified that even with prior technological experience, the use of technology on acute wards was often too big a jump for many stroke survivors:

*I think one of the reasons our use of DREX hasn't been as much as we'd like or anticipate has partly been the whole problem of how ready they are for some of this in the acute phase, and I think that can run over into the acute phase at home as well. That applies to DREX but also applies to some of the other rehab activities we apply to patients. And it's about,*

*if possible, increasing the option to grade patients into it. I often find with my patients on the ward, DREX is just too big of a jump for them. That even if they've got some tech knowledge beforehand the assessment is a big jump for them straight away.*

### Perceptions of Technology

A key factor discussed across all focus groups was technology, specifically if and how it is used by stroke survivors. There was a clear dichotomy between those who perceived technology as a positive aspect of rehabilitation and those who viewed it negatively.

### Technology as a Barrier to Telerehabilitation

Within this overarching theme of technology as a barrier, 3 key reasons were identified by stakeholders as to why they felt technology was not useful. The primary subtheme was confidence with technology; a large number of the stroke survivors talked about loss of confidence post stroke, which seriously impacted their life. Those who were previously technophobic reported having little confidence to learn something new:

*Facilitator: Would you ever want to use the internet or need to use the internet?*

*Participant: I don't think so, no, I don't really think so. It would be handy for some things but the amount of learning it would take...*

In addition to this, even those who were previously comfortable using technology reported becoming less confident after their stroke:

*I use it. Far less than I used to though. I used to be really comfortable on the internet but now I just stick to basics of where I'm comfortable and I wouldn't go exploring online at all now. I just use Amazon, Facebook, that's it. I haven't got the wherewithal that I used to. I used to be very switched on. Now I'm only half switched on.*

One aspect of confidence that stroke survivors discussed was fear, for example, fear of making a mistake while online. This was also echoed in a carer's responses, who similarly felt fearful letting their partner online because of the potential ramifications of errors:

*Participant: I'm just frightened of making a mistake. I think that's my biggest worry. One time I would just go on the internet and do my shopping on the internet. Now I'm frightened if I order 19 of something.*

*Carer: I haven't got the confidence to let him [use the internet] because if he gets into trouble on it I'm [in trouble]!*

A second subtheme was level of interest; some stroke survivors were not interested in accessing technology. This was encapsulated in one participant's response:

*I didn't have an interest then, I haven't got an interest now.*

It is important to understand that individuals may be reluctant to use technology purely because they have previously not had to use it and do not see a need now.

A third subtheme relating to disadvantages of technology was stroke survivors not appreciating the lack of personal contact. Some stroke survivors viewed the emergence of everything being done online, including rehabilitation, as a barrier in their rehabilitation journey:

*The other thing as well is everything is done on the internet, there's no personal contact. I think sometimes you need to break that barrier down.*

### Technology as a Facilitator for Telerehabilitation

Although some stroke survivors viewed technology as a barrier, others spoke about the benefits of technology and its positive impact:

*Technology has been my saviour.*

Specific positives were attributed to the accessibility of certain technology. Particular mention was made of YouTube and the advantage of being able to watch an online video repeatedly:

*YouTube videos would help teach me stuff because the beauty of YouTube is I can play it again and again and again.*

### Key Features of Telerehabilitation Packages

#### Additional Resources

After discussion with occupational therapists, it became apparent that there was no fixed activity provided for stroke survivors with visual loss to complete on acute wards, resulting in the time-consuming process of therapists creating their own materials. This led to therapists discussing the specific tools and materials they could use:

*So that's something we're already doing, not in the same format. So it's more paper based so it doesn't move up and it's not...we just make them ourselves as well. It's not as content heavy, it's not moving through space the way the app is with having to keep up and that.*

One theme that emerged was a need for stroke survivors to be provided with opportunities to better understand the requirements of the training packages used in telerehabilitation and the need for tools that could assist occupational therapists in delivering effective therapy in acute wards.

#### Goal Setting

Stroke survivors described the positive nature of setting goals in their rehabilitation and how this aided their rehabilitation in general:

*[I] think to myself I'm going to improve...again it was that 4%, 8%, 10%, 12%.*

Generally, goal setting is a good motivator and its inclusion in telerehabilitation provides stroke survivors with further motivation to undertake exercises above and beyond their rehabilitative benefit.



## Feedback

Feedback was highlighted by occupational therapists as critically important to rehabilitation, with stroke survivors feeling more positive and having a sense of achievement after receiving progress feedback:

*[Feedback] gives you that boost, doesn't it. They can say, "I'm improving." Because you like to know if you're improving.*

## Repetition

Stroke survivors discussed the positive aspects of repetition in their rehabilitation. One patient discussed the benefits of being able to do things repeatedly, as with his memory issues, this was the only way he could function independently:

*Facilitator: So for you, do you think the repetition...because obviously the tasks themselves are extremely repetitive. You're doing the same kind of training over and over again. Is that something that works really well for you?*

*Patient: Yeah because it takes me a while to get a hang of it and then I'm ahead of the game. Repetition is the secret of running my life.*

## Applications of These Findings to DREX

Participant responses highlighted a range of issues in the application of telerehabilitation for stroke survivors with visual field defects. We outline alterations made to the DREX training package in order to directly address these issues.

In liaison with occupational therapists, we created a dossier of paper-based tasks for use by stroke survivors who are not currently able to access the training. The tasks are paper-based versions adapted from the training exercises in the DREX app and provide an intermediary form of compensatory training for acute ward use. This tackles several identified issues, including alleviating some of the time pressure on therapists, since they do not have to do this task anymore, and addressing patient readiness for telerehabilitation by providing a simplified form. It could also be a starting point for those lacking confidence or interest in technology by familiarizing them with the tasks in a different format first. Therapists have begun trialing these resources at several local National Health Service trusts and have found them helpful:

*I can imagine a lot of people that it will be a real lifeline for because, you know, we run to the photocopy [sic], get our bits of paper, we make stuff up ourselves for them to do.*

Furthermore, working with stroke survivors and therapists, we updated user guides and website information to provide a more accessible service to key stakeholders. These user guides have also been printed and provided to therapists in acute settings to avoid having to use technology to access them, which may be a barrier to some. Furthermore, in accordance with an identified facilitator, a YouTube channel was created that included task demonstration videos. Such additional resources were designed to help alleviate some concerns that users experience regarding unknown technology, and we know that the videos viewed more often are those that correspond to some of the more difficult

aspects of navigating the training, indicating that they are a tool that can aid this.

Given the importance of feedback in rehabilitation, the delivery of this within the DREX training was improved; users now receive a speed and accuracy score after every training block. This allows them to effectively track their progress over the course of training. However, we are continuing to investigate ways to improve the feedback provided. Although stroke survivors have responded well to receiving feedback, therapists believed the feedback delivery could be improved. Therapists felt one-to-one monitoring of stroke survivors' feedback would further help stroke survivors understand how they are progressing and challenge them to continually improve.

*I thought it might be quite helpful to sit down with someone and go through it afterwards to say, "How did you find this," and "Well this is what it's shown." "This is what you've done really well" or "This part shows that's what you're saying is really difficult." For some people it might be really helpful to have those conversations in some more detail.*

Delivery of feedback is an extremely important factor in the success of rehabilitation in clinical populations, and this factor was no less understated in the responses of occupational therapists. This provides an interesting point in the move toward telerehabilitation, as this personalized delivery may be lost.

## Discussion

### Principal Findings

Through interviews with key stakeholder groups involved in stroke survivor rehabilitation, this study identified several facilitators and barriers to telerehabilitation. A general lack of confidence with technology, the perceived fear of using telerehabilitation, and the reduced face-to-face contact that a move toward more technological solutions in rehabilitation would bring were highlighted as the main issues stroke survivors faced. Initial in-person visits are one way to alleviate initial fear, as they can facilitate stroke survivor engagement and motivation in the rehabilitation process [24].

Telerehabilitation has been shown to have positive outcomes in motor consequences of stroke, health outcomes, carer support, and health professional satisfaction, albeit evidence quality was recognized to be poor [25]. The present study, therefore, thoroughly addresses the issues that stroke survivors with visual impairment face when using telerehabilitation.

On the basis of the work presented here, we recommend that several core issues be considered in the development and delivery of telerehabilitation. Compensatory training for poststroke visual field rehabilitation is repetitive in nature, which is why it works. It is important to ensure that such tools provide repetition in the most accessible way to ensure that repetition does not result in disengagement. Feedback and goal setting are important considerations when creating telerehabilitation, as they can improve engagement and motivation, but it is also important to provide a tangible progress update. In DREX, goals are set by the individuals based on feedback received when training. Previous studies have highlighted the motivational



benefit of individual goal commitment and goal setting on successful performance in other tasks [26]. As such, introducing goal-setting elements and feedback mechanisms into future telerehabilitation for stroke survivors would be beneficial, especially tailored programs that allow final goals (eg, overall aim) and interim goals (eg, daily target) to be determined.

In order to make telerehabilitation as accessible as possible for the majority of individuals, it would be valuable to consider intermediary steps and additional resources that can aid in intervention delivery. Using a multimodal approach (eg, paper, text, audio, visual elements, etc) enables these resources to be understood and used by all stakeholder populations. Engaging with key users prior to and during the development of telerehabilitation packages can ensure the creation of tools that are of maximum effectiveness and accessibility. For example, if during the acute phase of rehabilitation, perceptions of technology are identified as a barrier, then intermediary steps can be put in place to alleviate concerns and assist in the eventual uptake of the telerehabilitation tool. Additionally, training and education of users and key stakeholders involved in the application of telerehabilitation is an important

consideration previously found to lead to more effective treatment [27].

## Conclusions

In summary, this study identified several facilitators and barriers to telerehabilitation from the perspectives of the key stakeholders involved in stroke survivor rehabilitation, and it gathered opinions regarding key features required in telerehabilitation. This process has not only given a voice to the key populations directly involved in the rehabilitation process but also developed an understanding of the most important factors in developing a successful and engaging telerehabilitation tool to improve accessibility and usability. More generally, the issues highlighted by all of the populations engaged with show the need to address the disparity between acute care and the mountainous step to what comes next. An intermediary support phase may negate many of the identified barriers in stroke populations and deflect the fear that telerehabilitation reduces quality of care. Using the mitigations outlined will result in a more personalized recovery program for all stroke survivors.

---

## Acknowledgments

This work was supported by The Dunhill Medical Trust (grant number R455/1115). We thank the stroke survivors and carers who informed this research from Age UK Roundhouse Ashington Stroke Group, Gateshead Stroke Association Stroke Group, Newton Aycliffe Stroke Group, Teesside Stroke Club, and Stockton-On-Tees Happy Talk Stroke Club and the Stroke Association for their assistance. We also thank the stroke teams at the University Hospital of North Tees and Northumbria Hospital for their time and participation.

---

## Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

What is DREX: Compensatory eye-movement training following stroke.

[PDF File (Adobe PDF File), 527 KB - [jmir\\_v22i10e19604\\_app1.pdf](#)]

---

### Multimedia Appendix 2

Full survey.

[PDF File (Adobe PDF File), 267 KB - [jmir\\_v22i10e19604\\_app2.pdf](#)]

---

### Multimedia Appendix 3

Focus group and interview protocol.

[PDF File (Adobe PDF File), 512 KB - [jmir\\_v22i10e19604\\_app3.pdf](#)]

---

### Multimedia Appendix 4

The process of thematic analysis.

[PDF File (Adobe PDF File), 504 KB - [jmir\\_v22i10e19604\\_app4.pdf](#)]

---

### Multimedia Appendix 5

Supplementary survey data.

[PDF File (Adobe PDF File), 563 KB - [jmir\\_v22i10e19604\\_app5.pdf](#)]

---

## References

1. Hepworth L, Rowe F, Walker M, Rockliffe J, Noonan C, Howard C, et al. Post-stroke Visual Impairment: A Systematic Literature Review of Types and Recovery of Visual Conditions. *OR* 2016 Jan 10;5(1):1-43. [doi: [10.9734/or/2016/21767](https://doi.org/10.9734/or/2016/21767)]

2. Rowe F, Wright D, Brand D, Jackson C, Harrison S, Maan T, et al. A prospective profile of visual field loss following stroke: prevalence, type, rehabilitation, and outcome. *Biomed Res Int* 2013;2013:719096 [[FREE Full text](#)] [doi: [10.1155/2013/719096](https://doi.org/10.1155/2013/719096)] [Medline: [24089687](#)]
3. Zhang X, Kedar S, Lynn MJ, Newman NJ, Biousse V. Homonymous hemianopia in stroke. *J Neuroophthalmol* 2006 Sep;26(3):180-183. [doi: [10.1097/01.wno.0000235587.41040.39](https://doi.org/10.1097/01.wno.0000235587.41040.39)] [Medline: [16966935](#)]
4. Warren M. Pilot study on activities of daily living limitations in adults with hemianopsia. *Am J Occup Ther* 2009;63(5):626-633. [doi: [10.5014/ajot.63.5.626](https://doi.org/10.5014/ajot.63.5.626)] [Medline: [19785262](#)]
5. Hazelton C, Pollock A, Taylor A, Brady M. Visual field loss after stroke: exploring the impact on daily life. *Clin Rehabil* 2015 Jan;29(10):1016.
6. Ramrattan RS, Wolfs RC, Panda-Jonas S, Jonas JB, Bakker D, Pols HA, et al. Prevalence and causes of visual field loss in the elderly and associations with impairment in daily functioning: the Rotterdam Study. *Arch Ophthalmol* 2001 Dec;119(12):1788-1794. [doi: [10.1001/archophth.119.12.1788](https://doi.org/10.1001/archophth.119.12.1788)] [Medline: [11735788](#)]
7. Rowe FJ. Stroke survivors' views and experiences on impact of visual impairment. *Brain Behav* 2017 Sep;7(9):e00778 [[FREE Full text](#)] [doi: [10.1002/brb3.778](https://doi.org/10.1002/brb3.778)] [Medline: [28948075](#)]
8. MacIntosh C. Review articles-Stroke re-visited: Visual problems following stroke and their effect on rehabilitation. *BIOS* 2003 Jan;60:10-16.
9. Zihl J, Kennard C. Disorder of higher visual function. In: Brandt T, Caplan LR, Dichgans F, Diener HC, Kennard C, editors. *Neurological Disorders: Course and Treatment*. San Diego, CA: Academic Press; 1996.
10. Ali M, Hazelton C, Lyden P, Pollock A, Brady M, VISTA Collaboration. Recovery from poststroke visual impairment: evidence from a clinical trials resource. *Neurorehabil Neural Repair* 2013 Feb;27(2):133-141. [doi: [10.1177/1545968312454683](https://doi.org/10.1177/1545968312454683)] [Medline: [22961263](#)]
11. Chen CS, Lee AW, Clarke G, Hayes A, George S, Vincent R, et al. Vision-related quality of life in patients with complete homonymous hemianopia post stroke. *Top Stroke Rehabil* 2009;16(6):445-453. [doi: [10.1310/tsr1606-445](https://doi.org/10.1310/tsr1606-445)] [Medline: [20139047](#)]
12. Jones S, Shinton R. Improving outcome in stroke patients with visual problems. *Age Ageing* 2006 Nov;35(6):560-565. [doi: [10.1093/ageing/af074](https://doi.org/10.1093/ageing/af074)] [Medline: [16820528](#)]
13. Feigin VL, Lawes CM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *The Lancet Neurology* 2009 Apr;8(4):355-369. [doi: [10.1016/s1474-4422\(09\)70025-0](https://doi.org/10.1016/s1474-4422(09)70025-0)]
14. Penetration of tablet computer ownership in the United Kingdom (UK) in 2013-2017, by age. Statista - The Statistics Portal. 2017. URL: <https://www.statista.com/statistics/271882/tablet-owners-in-the-united-kingdom-uk-by-age/> [accessed 2020-07-03]
15. Rosen MJ. Telerehabilitation. *Telemed J E Health* 2004;10(2):115-117. [doi: [10.1089/tmj.2004.10.115](https://doi.org/10.1089/tmj.2004.10.115)] [Medline: [15319039](#)]
16. White J, Janssen H, Jordan L, Pollack M. Tablet technology during stroke recovery: a survivor's perspective. *Disabil Rehabil* 2015;37(13):1186-1192. [doi: [10.3109/09638288.2014.958620](https://doi.org/10.3109/09638288.2014.958620)] [Medline: [25212736](#)]
17. Edgar MC, Monsees S, Rhebergen J, Waring J, Van der Star T, Eng JJ, et al. Telerehabilitation in Stroke Recovery: A Survey on Access and Willingness to Use Low-Cost Consumer Technologies. *Telemed J E Health* 2017 May;23(5):421-429. [doi: [10.1089/tmj.2016.0129](https://doi.org/10.1089/tmj.2016.0129)] [Medline: [27705095](#)]
18. Vourganis I, Stankovic V, Stankovic L, Kerr A. Factors That Contribute to the Use of Stroke Self-Rehabilitation Technologies: A Review. *JMIR Biomed Eng* 2019 Aug 15;4(1):e13732 [[FREE Full text](#)] [doi: [10.2196/13732](https://doi.org/10.2196/13732)]
19. Zhou L, Parmanto B. Reaching People With Disabilities in Underserved Areas Through Digital Interventions: Systematic Review. *J Med Internet Res* 2019 Oct 25;21(10):e12981 [[FREE Full text](#)] [doi: [10.2196/12981](https://doi.org/10.2196/12981)] [Medline: [31654569](#)]
20. Aimola L, Lane AR, Smith DT, Kerkhoff G, Ford GA, Schenk T. Efficacy and feasibility of home-based training for individuals with homonymous visual field defects. *Neurorehabil Neural Repair* 2014;28(3):207-218. [doi: [10.1177/1545968313503219](https://doi.org/10.1177/1545968313503219)] [Medline: [24048623](#)]
21. Maximising telerehabilitation following visual loss after stroke: Interviews and focus groups with stroke survivors, carers and occupational therapists. Open Science Framework. URL: <https://osf.io/drxpf/> [accessed 2020-10-05]
22. Morgan D. Focus groups as qualitative research. 2nd ed. London, United Kingdom: Sage Publications; 1997.
23. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp0630a](https://doi.org/10.1191/1478088706qp0630a)]
24. Kairy D, Tousignant M, Leclerc N, Côté AM, Levasseur M, Researchers The Telage. The patient's perspective of in-home telerehabilitation physiotherapy services following total knee arthroplasty. *Int J Environ Res Public Health* 2013 Aug 30;10(9):3998-4011 [[FREE Full text](#)] [doi: [10.3390/ijerph10093998](https://doi.org/10.3390/ijerph10093998)] [Medline: [23999548](#)]
25. Johansson T, Wild C. Telerehabilitation in stroke care--a systematic review. *J Telemed Telecare* 2011;17(1):1-6. [doi: [10.1258/jtt.2010.100105](https://doi.org/10.1258/jtt.2010.100105)] [Medline: [21097560](#)]
26. Landers RN, Bauer KN, Callan RC. Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior* 2017 Jun;71:508-515. [doi: [10.1016/j.chb.2015.08.008](https://doi.org/10.1016/j.chb.2015.08.008)]

27. Peretti A, Amenta F, Tayebati SK, Nittari G, Mahdi SS. Telerehabilitation: Review of the State-of-the-Art and Areas of Application. *JMIR Rehabil Assist Technol* 2017 Jul 21;4(2):e7 [[FREE Full text](#)] [doi: [10.2196/rehab.7511](https://doi.org/10.2196/rehab.7511)] [Medline: [28733271](https://pubmed.ncbi.nlm.nih.gov/28733271/)]

## Abbreviations

**DREX:** Durham Reading and Exploration

*Edited by G Eysenbach; submitted 24.04.20; peer-reviewed by MR Golomb, L Fernandes; comments to author 23.06.20; revised version received 28.07.20; accepted 29.07.20; published 23.10.20.*

*Please cite as:*

Dunne S, Close H, Richards N, Ellison A, Lane AR

*Maximizing Telerehabilitation for Patients With Visual Loss After Stroke: Interview and Focus Group Study With Stroke Survivors, Carers, and Occupational Therapists*

*J Med Internet Res* 2020;22(10):e19604

URL: <http://www.jmir.org/2020/10/e19604/>

doi: [10.2196/19604](https://doi.org/10.2196/19604)

PMID: [33095179](https://pubmed.ncbi.nlm.nih.gov/33095179/)

©Stephen Dunne, Helen Close, Nicola Richards, Amanda Ellison, Alison R Lane. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Barriers and Enablers to Using a Patient-Facing Electronic Questionnaire: A Qualitative Theoretical Domains Framework Analysis

Janet Yamada<sup>1</sup>, RN, PhD; Andrew Kouri<sup>2,3</sup>, MD, FRCPC; Sarah-Nicole Simard<sup>1</sup>, RN, MSc; Stephanie A Segovia<sup>3</sup>, PhD; Samir Gupta<sup>2,3,4</sup>, MD, MSc, FRCPC

<sup>1</sup>Daphne Cockwell School of Nursing, Faculty of Community Services, Ryerson University, Toronto, ON, Canada

<sup>2</sup>Department of Medicine, University of Toronto, Toronto, ON, Canada

<sup>3</sup>Division of Respiriology, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada

<sup>4</sup>Keenan Research Centre for Biomedical Science, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada

**Corresponding Author:**

Samir Gupta, MD, MSc, FRCPC

Department of Medicine

University of Toronto

Division of Respiriology, St. Michael's Hospital, Unity Health Toronto, Bond Wing, Suite 6042

30 Bond St

Toronto, ON, M5B 1W8

Canada

Phone: 1 416 864 6026

Email: [samir.gupta@unityhealth.to](mailto:samir.gupta@unityhealth.to)

## Abstract

**Background:** Electronic patient questionnaires are becoming ubiquitous in health care. To address care gaps that contribute to poor asthma management, we developed the Electronic Asthma Management System, which includes a previsit electronic patient questionnaire linked to a computerized clinical decision support system.

**Objective:** This study aims to identify the determinants (barriers and enablers) of patient uptake and completion of a previsit mobile health questionnaire.

**Methods:** We conducted semistructured interviews with adult patients with asthma in Toronto, Canada. After demonstrating the questionnaire, participants completed the questionnaire using their smartphones and were then interviewed regarding perceived barriers and enablers to using and completing the questionnaire. Interview questions were based on the Theoretical Domains Framework to identify the determinants of health-related behavior. We generated themes that addressed the enablers and barriers to the uptake and completion of the questionnaire.

**Results:** In total, 12 participants were interviewed for saturation. Key enablers were as follows: the questionnaire was easy to complete without additional knowledge or skills and was perceived as a priority and responsibility for patients, use could lead to more efficient and personalized care, completion on one's own time would be convenient, and uptake and completion could be optimized through patient reminders. Concerns about data security, the usefulness of questionnaire data, the stress of completing it accurately and on time, competing priorities, and preferences to complete the questionnaire on other devices were the main barriers.

**Conclusions:** The barriers and enablers identified by patients should be addressed by developing implementation strategies to enhance e-questionnaire use and completion by patients. As the use of e-questionnaires grows, our findings will contribute to implementation efforts across settings and diseases.

(*J Med Internet Res* 2020;22(10):e19474) doi:[10.2196/19474](https://doi.org/10.2196/19474)

**KEYWORDS**

asthma; electronic questionnaire; patients; barriers; enablers; mobile phone

## Introduction

### Background

Asthma affects approximately 7.7% of adults in the United States and carries an annual economic burden exceeding US \$50 billion [1,2]. As in many other common chronic diseases, large gaps between guideline-recommended care and real-world care continue to exist in asthma [3,4].

Patient-facing electronic health questionnaires represent a promising strategy to address clinical care gaps, as they engage patients in taking a more active role in their care, increase efficiency by reducing data acquisition burdens on clinicians, and collect data that can be directly processed through computerized decision support system (CDSS) algorithms and then integrated into point-of-care electronic health records (EHRs) for clinicians [5]. With the dramatic rise in smartphone ownership in the United States [6] and a shrinking *digital divide* [7], mobile health (mHealth) previsit questionnaires have emerged as a commonly used strategy across primary and specialty care settings [8-11]. However, little is known about patient perceptions of these questionnaires and their actual uptake and what strategies can be leveraged to drive their use. It is of great importance to understand the user's perceived barriers and enablers to use mHealth technology to maximize its success [12].

Our team developed the Electronic Asthma Management System (eAMS), a point-of-care CDSS, to bridge key care gaps in the management of asthma in primary care settings [5]. The eAMS consists of a previsit electronic patient questionnaire that collects asthma-related parameters and a CDSS that receives and processes questionnaire data to produce guideline-based decision support (asthma control level, corresponding medication change recommendations, and a self-management tool called an asthma action plan (AAP) [4] integrated into the EHR in real time). In a 2-year interrupted time series study involving 890 patients and 237 clinicians, the eAMS significantly improved the frequency of asthma control assessment, the ratio of controller to reliever medication prescriptions, and delivery of AAPs to patients [5]. However, the previsit questionnaires were used by only 61.2% (551/890) of patients [5]. A detailed uptake analysis identified several quantitative predictors of questionnaire uptake [13].

### Objectives

In this study, we sought to identify the determinants (barriers and enablers) of uptake and completion of the eAMS mHealth previsit patient questionnaire, with a view to tailoring targeted strategies to overcome barriers and leverage enablers to drive patient usage. Given the growing popularity of mobile patient questionnaires, we identified themes that would be broadly relevant across such systems.

## Methods

### Recruitment

This was a qualitative study. Patients were recruited from the St. Michael's Hospital's Respiratory Patient Database via

telephone or email, and each provided informed written consent. We included adults (aged  $\geq 18$  years) with a clinical diagnosis of asthma, who were able to speak and write English, and who owned a smartphone. Each participant received a Can \$50 (US \$38) gift card for participating in the study. Ethics approval for the study was obtained from the Research Ethics Boards at Ryerson University and St. Michael's Hospital.

### Sample Size

For theory-based interview studies, Francis et al [14] recommended a minimum of 10 interviews, followed by an initial analysis, and up to an additional 3 interviews or fewer if data saturation (ie, no new themes generated) is achieved before that. Accordingly, we sought to interview at least 10 participants and up to 3 more pending interim analyses for data saturation.

### Interview Procedure

All interviews were conducted by 3 members of the research team trained in conducting qualitative interviews. Interviews were audio recorded and transcribed verbatim. After consent was obtained, participants were asked to complete a data collection form for background information (demographics, asthma-related information, and smartphone usage). Given that group sessions featured direct use of the eAMS questionnaire, each participant was asked to bring his or her own smartphone and was emailed a link to the questionnaire at the start of the session. We then explained and demonstrated the functions, role, and workflow of the questionnaire (including information transmission to the clinician's CDSS) and asked participants to independently complete the questionnaire on their device.

The questionnaire takes 5 to 10 min to complete and poses a series of questions that determine guideline-based asthma control levels, identify the individual's asthma symptoms and triggers (eg, sports, activities, location), and elicit current medication use. We created a web-responsive version of the tablet-based waiting room questionnaire that was used in the original eAMS study [5]. The original questionnaire's design and testing process are described in detail elsewhere [15,16]. Briefly, the questionnaire was designed based on the best principles for touch questionnaire usability [16], and both its content and usability were optimized through patient feedback in a rapid cycle design process (with summative qualitative analysis) involving 20 patients with asthma sampled purposively to represent a broad range of ages and electronic technology and touch device experience. The questionnaire achieved an *excellent* System Usability Scale score of 84.2 (SD 14.7). The System Usability Scale is a 10-item Likert scale questionnaire that has been used widely to measure the perceived effectiveness of, satisfaction with, and system efficacy of technological apps [15-17].

Immediately after completing the questionnaire, participants were asked specific questions related to the barriers and enablers to complete the questionnaire in a real-world setting. The described real-world workflow included downloading an app to their smartphone or tablet and accessing the questionnaire through this app for up to 1 week before their doctor's appointment (up to and including in the waiting room immediately before the appointment). The interview guide was



based on the Theoretical Domains Framework (TDF), an integrative framework comprised of 14 theoretical domains derived from 33 validated health and social psychology theories and 128 constructs that may drive and explain health-related behavior change from a psychological perspective [18,19]. The TDF is a comprehensive approach to understanding the determinants of behaviors in health care professionals and patients; it has been applied successfully for this purpose across diverse diseases and health care settings; and it has emerged as a standard for barrier and enabler measurement in implementation research [20]. The 14 theoretical domains include (1) knowledge; (2) skills; (3) social/professional role

and identity; (4) beliefs about capabilities; (5) optimism; (6) beliefs about consequences; (7) reinforcement; (8) intentions; (9) goals; (10) memory, attention and decision processes; (11) environmental context and resources; (12) social influences; (13) emotions; and (14) behavioral regulation [20]. The interview guide was designed to explore which domains in the TDF were relevant for the targeted behavior (ie, completion of the eAMS questionnaire before the clinical visit) and how each of those domains influenced this behavior [20]. [Table 1](#) describes the 14 TDF domains [19] and associated interview questions.

**Table 1.** Theoretical Domains Framework: theoretical domains and associated questions.

Theoretical Domains Framework domain	Definitions	Example interview questions
Knowledge	An awareness of the existence of something	<ul style="list-style-type: none"> <li>Having seen the questionnaire, do you need more information? What would need to be in place for you to do this questionnaire before seeing your doctor?</li> </ul>
Skills	An ability or proficiency acquired through practice	<ul style="list-style-type: none"> <li>What types of training, if any, would have been required to help you to complete the questionnaire?</li> </ul>
Social/professional role and identity	A coherent set of behaviors and displayed personal qualities of an individual in a social or work setting	<ul style="list-style-type: none"> <li>Do you feel it is part of your responsibility to complete this questionnaire? Do you think this is something you should do?</li> </ul>
Beliefs about capabilities	Acceptance of the truth, reality, or validity about an ability, a talent, or a facility that a person can put to constructive use	<ul style="list-style-type: none"> <li>How easy or difficult was it to complete the asthma questionnaire? What parts of the questionnaire were easy or difficult to complete? How confident were you in your ability to complete the questionnaire?</li> </ul>
Optimism	The confidence that things will happen for the best or that desired goals will be obtained	<ul style="list-style-type: none"> <li>How likely do you think it is that completing the asthma questionnaire would lead to better management of your asthma by your doctor?</li> </ul>
Beliefs about consequences	Acceptance of the truth, reality, or validity about outcomes of a behavior in a given situation	<ul style="list-style-type: none"> <li>What do you think are the benefits of completing the asthma questionnaire? What are the good things that can happen as a result of completing the asthma questionnaire?</li> </ul>
Reinforcement	Increasing the probability of a response by arranging a dependent relationship, or contingency, between the response and a given stimulus	<ul style="list-style-type: none"> <li>Are there any good experiences that you have had with your asthma management that would increase your likelihood of completing the asthma questionnaire?</li> </ul>
Intentions	A conscious decision to perform a behavior or a resolve to act in a certain way	<ul style="list-style-type: none"> <li>Is completing the questionnaire something you plan to do if you are asked to complete the questionnaire in the future?</li> </ul>
Goals	Mental representations of outcomes or end states that an individual wants to achieve	<ul style="list-style-type: none"> <li>How much of a priority would it be to complete the asthma questionnaire before you see your doctor, compared to other priorities?</li> </ul>
Memory, attention and decision processes	The ability to retain information, focus selectively on aspects of the environment, and choose between 2 or more alternatives	<ul style="list-style-type: none"> <li>Is completing a questionnaire before seeing a doctor something you usually do? Would you remember to complete the questionnaire?</li> <li>When could you see yourself forgetting to complete the questionnaire? Why? What could help to prevent this?</li> <li>Would a text or email message 1 week before your appointment be a sufficient reminder?</li> </ul>
Environmental context and resources	Any circumstance of a person's situation or environment that discourages or encourages the development of skills and abilities, independence, social competence, and adaptive behavior	<ul style="list-style-type: none"> <li>Is there anything that might influence whether or not you would complete the questionnaire? (eg, Wi-Fi availability, equipment to complete, links broken, time, competing tasks, etc)</li> </ul>
Social influences	Those interpersonal processes that can cause individuals to change their thoughts, feelings, or behaviors	<ul style="list-style-type: none"> <li>Who would influence you to complete the questionnaire? (eg, family member, doctor, etc)</li> </ul>
Emotions	A complex reaction pattern, involving experiential, behavioral, and physiological elements, by which the individual attempts to deal with a personally significant matter or event	<ul style="list-style-type: none"> <li>Tell me how you feel about completing the asthma questionnaire before you see your doctor. Are you worried, or do you have any concerns about completing the questionnaire?</li> </ul>
Behavioral regulation	Anything aimed at managing or changing objectively observed or measured actions	<ul style="list-style-type: none"> <li>Can you think of a plan for how you would ensure that you complete the asthma questionnaire before your appointment?</li> </ul>

## Data Analysis

Participant characteristics were summarized in detail. In total, 2 trained qualitative analysts (JY and SS) independently coded transcripts into the 14 TDF domains (using directed content analysis) [21]. They first coded 2 transcripts independently and together defined a consensus coding scheme (ie, codes, definitions of codes, examples of quotes under each code). This coding scheme was used to independently code the remaining transcripts (using NVivo 12 software, QSR International). The coders met on a regular basis to discuss discrepant codes, which were resolved through consensus. For each coded quote, the analysts generated a statement that reflected the core belief expressed by that response (a *belief statement*) [20]. These statements were grouped into themes that suggested significant influence on the uptake of the questionnaire. In instances where a single domain allocation agreement could not be reached, the statement was placed in both pertinent domains [20]. Emerging themes and belief statements were verified by each analyst and adjusted for consensus. To ensure the trustworthiness of the data, the analysts used memos during the coding process and developed an audit trail to track progress and enable easy reference to the primary data at a later time, if necessary. TDF domains that were considered of high importance were those with (1) belief statements in the domain that had relatively high

frequencies (ie, more than one-third of respondents identified the belief), (2) conflicting beliefs in the domain (ie, where participants identified opposing beliefs), and (3) evidence of strong beliefs in a domain that is believed to directly impact uptake (as determined by the consensus of research team members) [20,22]. Verbatim quotes along with their categories and themes were used to support the belief statements.

## Results

### Overview

Data were collected from June 2018 to April 2019. Of the 21 patients who were approached to participate, 12 were recruited and consented to be interviewed. We performed an interim analysis after the first 10 interviews and coded each transcript thereafter until saturation. Saturation was achieved after another 2 interviews, for a total of 12 interviews.

### Participant Characteristics

Participant characteristics are summarized in Table 2. All participants were aged  $\geq 31$  years, 58% (7/12) of the participants were female, and a majority had a college education or higher. All participants used their smartphones frequently during the day for a variety of purposes. Each face-to-face interview lasted approximately 45 min.

**Table 2.** Participant characteristics (n=12).

Variable	Values
<b>Age (years), n (%)</b>	
25-30	0 (0)
31-40	3 (25)
41-50	3 (25)
51-60	2 (17)
≥60	4 (33)
<b>Sex, n (%)</b>	
Female	7 (58)
<b>Highest level of education completed, n (%)</b>	
University	8 (66)
College or trade school or other	2 (17)
High school	2 (17)
Elementary school	0 (0)
Number of years since asthma diagnosis, median (IQR)	8 (5.25-36)
Inhaler medication for asthma, n (%)	12 (100)
<b>Use of rescue inhaler, n (%)</b>	
Yes	8 (66)
No	2 (17)
Not reported	2 (17)
<b>Use of controller inhaler, n (%)</b>	
Yes	11 (92)
No	1 (8)
<b>Type of smartphone, n (%)</b>	
Blackberry OS <sup>a</sup>	1 (8)
iPhone	6 (50)
Android OS <sup>b</sup>	5 (42)
<b>Smartphone activity (check all that apply), n (%)</b>	
Email	10 (83)
Recreation (eg, text messaging)	10 (83)
Information seeking (eg, finding addresses, directions)	9 (75)
Information storage (eg, contacts)	9 (75)
Scheduling (eg, to-do lists, appointments)	8 (66)
Other (eg, Facebook, GPS)	6 (50)

<sup>a</sup>OS: operating system.

<sup>b</sup>Android OS: included phones manufactured by Samsung, Huawei, LG, and OnePlus.

## Relevant Theoretical Domains

Participants spent an average of 9.6 (SD 5.2) min to complete and review the questionnaire before the interview. All 14 TDF domains were identified as relevant. Most of the belief statements focused on enablers to use the questionnaire. A summary of the 14 TDF domains, related belief statements, and participant quotes is provided in [Multimedia Appendix 1](#).

## Enablers

Overall, the majority of participants indicated that the eAMS questionnaire was easy and straightforward to complete, with no training required (knowledge, beliefs about capabilities, and skills). Most participants stated that they would complete the questionnaire on their phone if they were given that option (intention). Completion of the questionnaire was viewed by participants as important and a priority (goals). Having the

option to complete the questionnaire on their own time was considered a benefit to participants (environmental context and resources):

*I think it's more relaxed which makes the patient feel easier and there's no rush. Like okay, you're doing the questionnaire and somebody comes to the door or you can just go answer the door, and then go back to the questionnaire, or be doing the questionnaire at 3 o'clock in the morning if you want... [P1003]*

Several participants reported that their prior experiences with poor disease control motivated them to use the tool (reinforcement), and the majority of participants reported that it was their responsibility to complete the questionnaire to enhance their level of care (social/professional role and identity):

*I feel it is [my responsibility] because the way I look at it, the more information my doctor has about what's going on with me, the better he or she is able to help me manage my symptoms and cope. [P1004]*

Most participants described benefits to completing a patient questionnaire before their medical appointment, including that their primary care physician would have a more thorough understanding of their condition and that this would translate into better health care and disease control (eg, this would inform their self-management plans; beliefs about consequences). Participants also indicated that completing the questionnaire would enhance their understanding of their own disease control (beliefs about consequences):

*I think from the patient's perspective, it would give them more insight as to what's happening because they actually have to think about what their symptoms are and what they're doing to help the asthma, so therefore, they can bring that information to the doctor. [P1008]*

More than half of the participants felt that receiving a reminder from the doctor's office via email, text, or phone before their appointment would ensure that the questionnaire would be completed (behavioral regulation, memory, attention and decision processes; environmental context and resources).

## Barriers

Several participants mentioned that competing priorities might influence whether they would complete the questionnaire in advance of their medical appointment (environmental context and resources):

*I have my two personal emails. I have 4 different work emails that I have to go through every day. Like it's swamped with these things. [P1003]*

Some would have preferred to complete the questionnaire on a larger device (eg, computer, iPad; environmental context and resources) or on paper. Barriers to using the questionnaire also included conflicting beliefs. A small number of participants felt that completing the questionnaire would not help them to understand their asthma nor lead to better asthma control (beliefs about consequences). Several participants expressed concerns about web-based security of questionnaire data and whether they could complete the questionnaire correctly (emotions).

## Discussion

### Principal Findings

To our knowledge, this is the first study to formally assess barriers and enablers to patient use of an mHealth questionnaire and the first to apply the validated TDF for this purpose. In identifying the determinants of uptake of the eAMS patient questionnaire, we identified constructs that are likely applicable to mHealth questionnaires across diseases.

First, we noted that factors relating to the content, format, and usability of the questionnaire itself were mostly perceived favorably and characterized as enablers to its use, as opposed to barriers. Specifically, most users did not believe that they needed extra training in the form of external knowledge or skills to complete the questionnaire (knowledge and skills) and clearly indicated that they were capable of completing it as required (beliefs about capabilities). This is likely a direct consequence of the user-centered approach taken in questionnaire development, whereby both content and format were serially improved in response to user feedback in focus group settings [15,16]. Indeed, Sun et al [23] previously reported that perceived ease of use is positively associated with behavioral intention to use mHealth systems. Accordingly, our findings reinforce the value of the integrated knowledge translation approach [24] in general and of rapid cycle design in particular, whereby end user engagement at an early stage of development is critical to ensure that the intrinsic features of a tool are conducive to later uptake [25]. Yet, it should be noted that 2 participants indicated that some users might require telephone or in-person support during questionnaire use (environmental context and resources), and it is likely that universal and sustainable adoption of any such mobile questionnaire will require investment in some form of technical support for a small number of users [26], even if they own a smartphone. As noted, the questionnaire completion rate was only approximately 60% in our previous study, which required patients to complete the questionnaire on a waiting room-based tablet [5]. In this study, the use case involved completion on their own device, up to 1 week in advance of the appointment, and patients indicated that both the ability to complete it on their own time schedule and in their own physical space were facilitators (environmental context and resources). This preference for e-questionnaire completion at home as opposed to within a health care facility was also noted in a previous study [27]. This suggests that questionnaire modalities that are limited to the waiting room environment, such as tablets or kiosks, may be less favorable. However, the advance completion approach would not be able to leverage typical prompts and reminders used in the waiting room environment (both human [eg, receptionist] and visual [eg, waiting room posters]). Correspondingly, some patients believed that they might fail to complete the questionnaire because of barriers presented by a lack of time or competing day-to-day priorities (environmental context and resources), and others indicated that they might simply forget (memory, attention and decision processes). Although some patients noted that they could set a calendar or personal email or device reminder (behavioral regulation), the vast majority indicated that a *friendly* (external) reminder would be a key facilitator (memory, attention and



decision processes), with preferred modalities being email or SMS text message (environmental context and resources). In summary, the balance of our findings suggests that completion at home on a personal device is preferable but that electronic reminders should also be incorporated into any e-questionnaire implementation plan. Given that health practitioners were the most commonly reported influencers who could drive usage (social influences), invitations and reminders should ideally be addressed from the patient's own health care provider.

As for the mode of delivery of the questionnaire, most users preferred the electronic system to a conventional paper-based questionnaire, although it is of note that 2 participants would have preferred paper (environmental context and resources). Several previous studies have similarly reported that most patients, across conditions, prefer an e-questionnaire to a paper-based one. [28-30]. Barentz et al [29] found that a preference for e-questionnaires was greatest in younger and more educated patients. Although our study included participants across a wide age spectrum, most were highly educated (10/12, 83% had a college education or higher), which may have contributed to this finding. It is also of note that some participants would prefer using a larger device, such as a personal computer or laptop (environmental context and resources). This supports previous observations that mobile apps should not exist in isolation [26]; any such mHealth questionnaire should be complemented by both a conventional website and tailored or web-responsive content accessible across platforms [26].

The vast majority of participants in our study felt that it was part of their responsibility to complete the questionnaire before their appointment (social/professional role and identity). Indeed, given a gradual movement away from the conventional unidirectional model of care (provider to patient) toward a bidirectional model with direct patient engagement and shared decision making over the last decade [31], it is not surprising that many patients perceived completion of a previsit questionnaire as a natural part of the health care interaction. Not only do studies suggest that a majority of patients would prefer to have an active role in their health care [32] but that such participation is associated with improved health outcomes [33]. This perceived responsibility was complemented by the perceived benefits of completing the questionnaire (beliefs about consequences). Some participants believed that completing the questionnaire before their medical appointment would benefit their physician by facilitating history taking and saving time. This finding is similar to that of Howell et al [27], who reported that a desire to reduce consultation time and enhance clinic efficiency motivated patients to complete a preoperative tablet-based questionnaire. Other direct benefits of questionnaire use mentioned by our participants included improving their own understanding of their disease and enabling both more personalized care (eg, a personalized AAP) and a higher quality of care. This perception of higher quality care was also the driving factor behind goal setting and prioritizing questionnaire completion (goals). Given that these factors appear to be strong enablers, patient-facing messaging at the time of an

e-questionnaire rollout could reinforce a sense of patient responsibility and these myriad benefits of completing such a questionnaire in an effort to drive usage.

Some participants also noted that their own previous experience with poor health outcomes acted as facilitators by enabling them to recognize the importance of completing the questionnaire (reinforcement). Correspondingly, previous work has shown that perceived health threats predict patient intentions to use smartphone-based health technology [34]. This facilitator may be particularly relevant in diseases that are characterized by episodic flares, such as asthma [35], chronic obstructive pulmonary disease [36], and congestive heart failure [37] and can also be leveraged in patient-facing messaging to drive usage.

Although the vast majority of participants indicated that they would ultimately complete such a questionnaire (intentions), certain barriers and strategies to address them are worth noting. Some participants were not convinced that their questionnaire responses would ultimately prove useful (optimism). This barrier to uptake has been echoed in previous eHealth literature, particularly among older adults concerned about *getting it wrong* [26]. Similarly, some participants expressed concerns about the stress of completing the questionnaire on time and accurately (emotions). These findings suggest that it would be important for any e-questionnaire to include an option to skip questions, and to include response options such as "I don't know" or "I'm not sure." Additional messaging could explain that missing answers and data entry errors are acceptable, that patients are simply being asked to "do their best," and that clinicians will verify and complete data where required. Another concern raised by one participant was the risk of a data security breach (emotions). Indeed, patient concerns about data security are a well-described barrier to the use of social media [38], mHealth apps [39], and eHealth questionnaires in particular [26,27]. These concerns can be addressed upfront through transparent terms of use, privacy, and data policies as well as through opt-out clauses for users. Ideally, a formal privacy impact assessment and threat risk assessment should be performed. Although the cost of formal analyses may be prohibitive for small organizations, some jurisdictions are facilitating this process for developers, such as the National Health Service (NHS) in the United Kingdom, through the NHS Health Apps Library [40].

## Practice Implications

Our findings consisted mostly of enablers to using the e-questionnaire and were identified across participants, and many were supported by the literature in parallel areas. On the basis of these findings, we were able to suggest simple solutions that might drive questionnaire usage. A summary of these identified barriers and enablers and corresponding optimization strategies to address and leverage them are provided in Table 3. These strategies are rooted in concepts related to behavior change techniques (BCTs); however, as a next step, we will employ a formal process (ie, a behavior change matrix) to identify BCTs that will be mapped onto these relevant TDF domains (ie, modifiable barriers and enablers) [41,42].

**Table 3.** Practice implications for uptake and completion of the Electronic Asthma Management System questionnaire based on identified enablers and barriers.

Identified barrier and enabler	Corresponding system optimization strategy
<ul style="list-style-type: none"> <li>Competing priorities may impede questionnaire completion before appointment (barrier: environmental context and resources)</li> </ul>	<ul style="list-style-type: none"> <li>Reinforce that questionnaire takes only 5 to 10 min to complete and can be done at any time before appointment</li> <li>Provide a reminder to administrative staff at the time of appointment to identify patients who have not completed the questionnaire and prompt them to complete it in waiting room</li> </ul>
<ul style="list-style-type: none"> <li>Preference to complete questionnaire on a larger device (barrier: environmental context and resources)</li> </ul>	<ul style="list-style-type: none"> <li>Develop a web-responsive version of the questionnaire, optimized for completion on large tablets, laptops, and desktop computers</li> </ul>
<ul style="list-style-type: none"> <li>Questionnaire completion will enhance one's understanding of one's own disease control (enabler: beliefs about consequences)</li> <li>Questionnaire completion would not help one to better understand their asthma (barrier: beliefs about consequences)</li> </ul>	<ul style="list-style-type: none"> <li>Position the questionnaire app as part of a general educational tool for asthma by developing dedicated web-based educational content that users can access simultaneously through the same app</li> </ul>
<ul style="list-style-type: none"> <li>Prior experiences of poor disease control motivated questionnaire completion (enabler: reinforcement)</li> <li>Questionnaire completion would not lead to better asthma control (barrier: beliefs about consequences)</li> </ul>	<ul style="list-style-type: none"> <li>Include patient-facing messages describing the personal impact of poor disease control, how doing the questionnaire will lead to an AAP<sup>a</sup>, and how an AAP has been shown to reduce symptoms of poor control</li> <li>Enhance the questionnaire app by providing access to the AAP itself through the app</li> </ul>
<ul style="list-style-type: none"> <li>It is part of one's responsibility to complete the questionnaire to enhance their care (enabler: social/professional role and identity)</li> <li>Questionnaire completion leads to the physician having a more thorough understanding of one's condition, translating into better health care and disease control (enabler: beliefs about consequences)</li> </ul>	<ul style="list-style-type: none"> <li>Ask each user's <i>own</i> health care provider (physician or nurse practitioner) to send a message to patients emphasizing the following: <ul style="list-style-type: none"> <li>This is a role that they would like the patient to play in their own care</li> <li>Completing this will enable them to better understand the patient's disease and to offer better care</li> </ul> </li> </ul>
<ul style="list-style-type: none"> <li>Concerns about web-based security of questionnaire data (barrier: emotions)</li> </ul>	<ul style="list-style-type: none"> <li>Conduct a formal privacy impact assessment and threat risk assessment and clearly communicate this robust security approach to users</li> </ul>
<ul style="list-style-type: none"> <li>Concerns about whether one can complete the questionnaire correctly (barrier: emotions)</li> </ul>	<ul style="list-style-type: none"> <li>Provide patients with options to skip questions and responses and ensure availability of telephone and email technical support</li> </ul>
<ul style="list-style-type: none"> <li>A reminder would ensure that the questionnaire would be completed (enabler: behavioral regulation, memory, attention and decision processes; environmental context and resources)</li> </ul>	<ul style="list-style-type: none"> <li>Include <i>friendly</i> reminders via automated phone call, email, or text message to complete the questionnaire (and allow personalization of reminder preferences at the time of patient registration)</li> </ul>

<sup>a</sup>AAP: asthma action plan.

## Limitations

Our study applied a validated, theory-based methodology to identify and classify barriers and enablers and sampled a diverse population (varied ages and smartphone operating system types). Smartphone ownership is approaching universality in industrialized nations, and we expect that most patients would complete such a questionnaire on a smartphone; however, our results are not pertinent to users without a smartphone who would have to access the system through a web-based, web-responsive questionnaire at home or on a tablet device made available for this purpose in the clinic setting. Recruitment of participants from a quaternary care center may also limit generalizability to other settings. Given that patients with asthma prefer a collaborative role in managing their illness, across a range of asthma severity levels [43], we do not believe that the

severity of the disease itself would influence the barriers and enablers that we identified, with the exception of previous poor health outcome experiences acting as a facilitator.

## Conclusions

In this study, we formally identified barriers and enablers in the uptake of a patient-facing mHealth questionnaire. We believe that our findings are broadly relevant to the rapidly growing use of e-questionnaires across health disciplines. Where possible, we suggested strategies that can be used to address barriers and leverage enablers. Future studies are recommended to employ a formal process (ie, a behavior change matrix) to identify behavior change strategies and techniques that are mapped onto the relevant TDF domains (ie, modifiable barriers and enablers) [41,42] and test their impact on actual questionnaire uptake.

## Acknowledgments

The authors would like to thank the participants for providing extensive feedback on the e-questionnaire. This study was funded by the Lung Association Provincial Grant-in-Aid/National Grant Review competition. AK is supported by the Thorpe Family Fellowship in Respiratory Medicine at St. Michael's Hospital. SG is supported by the Michael Locke Term Chair in Knowledge Translation and Rare Lung Disease Research. Funders had no role in the design of the study; in the collection, analysis, or interpretation of data; or in manuscript preparation.

## Conflicts of Interest

SG owns the intellectual property associated with eAMS and the eAMS questionnaire and would have an ownership interest in any commercial enterprise derived from this system. Other authors report no potential conflicts of interest.

## Multimedia Appendix 1

Relevant theoretical domains framework domains, belief statements, and sample quotes.

[DOC File, 64 KB - [jmir\\_v22i10e19474\\_app1.doc](#)]

## References

1. National Health Interview Survey. Centre for Disease Control and Prevention. 2017. URL: <https://www.cdc.gov/nchs/nhis/shs/tables.htm> [accessed 2017-09-19] [WebCite Cache ID <https://www.cdc.gov/nchs/nhis/>]
2. Barnett SB, Nurmagambetov TA. Costs of asthma in the United States: 2002-2007. *J Allergy Clin Immunol* 2011 Jan;127(1):145-152. [doi: [10.1016/j.jaci.2010.10.020](#)] [Medline: [21211649](#)]
3. Chapman KR, Boulet LP, Rea RM, Franssen E. Suboptimal asthma control: prevalence, detection and consequences in general practice. *Eur Respir J* 2008 Feb;31(2):320-325. [doi: [10.1183/09031936.00039707](#)] [Medline: [17959642](#)]
4. Price C, Agarwal G, Chan D, Goel S, Kaplan AG, Boulet L, et al. Large care gaps in primary care management of asthma: a longitudinal practice audit. *BMJ Open* 2019 Jan 29;9(1):e022506. [doi: [10.1136/bmjopen-2018-022506](#)] [Medline: [30696669](#)]
5. Gupta S, Price C, Agarwal G, Chan D, Goel S, Boulet L, et al. The Electronic Asthma Management system (eAMS) improves primary care asthma management. *Eur Respir J* 2019 Apr;53(4). [doi: [10.1183/13993003.02241-2018](#)] [Medline: [30765503](#)]
6. Mobile Fact Sheet. Pew Research Center. 2019. URL: <https://www.pewresearch.org/internet/fact-sheet/mobile/> [accessed 2020-09-19] [WebCite Cache ID <https://www.pewresearch.org/internet/fact-sheet/mobile/>]
7. Vangeepuram N, Mayer V, Fei K, Hanlen-Rosado E, Andrade C, Wright S, et al. Smartphone ownership and perspectives on health apps among a vulnerable population in East Harlem, New York. *Mhealth* 2018;4:31 [FREE Full text] [doi: [10.21037/mhealth.2018.07.02](#)] [Medline: [30221166](#)]
8. Koh S, Choi Y, Cheon J, Kim CR, Kim JH, Kim J. Using a wireless touch screen tablet personal computer is feasible to assess the quality of breast cancer survivorship. *J Cancer Res Pract* 2018 Dec;5(4):136-140. [doi: [10.1016/j.jcrpr.2018.08.002](#)]
9. Kim H, Park HC, Yoon SM, Kim TH, Kim J, Kang MK, et al. Evaluation of quality of life using a tablet PC-based survey in cancer patients treated with radiotherapy: a multi-institutional prospective randomized crossover comparison of paper and tablet PC-based questionnaires (KROG 12-01). *Support Care Cancer* 2016 Oct;24(10):4399-4406. [doi: [10.1007/s00520-016-3280-5](#)] [Medline: [27220523](#)]
10. Moura LM, Schwamm E, Moura Junior V, Seitz MP, Hsu J, Cole AJ, et al. Feasibility of the collection of patient-reported outcomes in an ambulatory neurology clinic. *Neurology* 2016 Dec 6;87(23):2435-2442 [FREE Full text] [doi: [10.1212/WNL.0000000000003409](#)] [Medline: [27815405](#)]
11. Suzuki E, Mackenzie L, Sanson-Fisher R, Carey M, D'Este C, Asada H, et al. Acceptability of a touch screen tablet psychosocial survey administered to radiation therapy patients in Japan. *Int J Behav Med* 2016 Aug;23(4):485-491. [doi: [10.1007/s12529-015-9502-2](#)] [Medline: [26249725](#)]
12. Matricardi PM, Dramburg S, Alvarez-Perea A, Antolín-Amérigo D, Apfelbacher C, Atanaskovic-Markovic M, et al. The role of mobile health technologies in allergy care: an EAACI position paper. *Allergy* 2020 Feb;75(2):259-272. [doi: [10.1111/all.13953](#)] [Medline: [31230373](#)]
13. Kouri A, Yamada J, Sale J, Straus SE, Gupta S. Primary care pre-visit electronic patient questionnaire for asthma: uptake analysis and predictor modeling. *J Med Internet Res* 2020 Sep 18;22(9):e19358 [FREE Full text] [doi: [10.2196/19358](#)] [Medline: [32945779](#)]
14. Francis JJ, Johnston M, Robertson C, Glidewell L, Entwistle V, Eccles MP, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychol Health* 2010 Dec;25(10):1229-1245. [doi: [10.1080/08870440903194015](#)] [Medline: [20204937](#)]
15. Gupta S, Lam Shin Cheung V, Kastner M, Straus S, Kaplan A, Boulet L, et al. Patient preferences for a touch screen tablet-based asthma questionnaire. *J Asthma* 2019 Jul;56(7):771-781. [doi: [10.1080/02770903.2018.1490750](#)] [Medline: [29972087](#)]

16. Lam Shin Cheung V, Kastner M, Sale JE, Straus S, Kaplan A, Boulet L, et al. Development process and patient usability preferences for a touch screen tablet-based questionnaire. *Health Informatics J* 2020 Mar;26(1):233-247. [doi: [10.1177/1460458218824749](https://doi.org/10.1177/1460458218824749)] [Medline: [30672358](https://pubmed.ncbi.nlm.nih.gov/30672358/)]
17. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud* 2009 May;4(3):114-123.
18. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A, Psychological Theory Group. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care* 2005 Feb;14(1):26-33. [doi: [10.1136/qshc.2004.011155](https://doi.org/10.1136/qshc.2004.011155)] [Medline: [15692000](https://pubmed.ncbi.nlm.nih.gov/15692000/)]
19. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implement Sci* 2012 Apr 24;7:37 [FREE Full text] [doi: [10.1186/1748-5908-7-37](https://doi.org/10.1186/1748-5908-7-37)] [Medline: [22530986](https://pubmed.ncbi.nlm.nih.gov/22530986/)]
20. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implement Sci* 2017 Jun 21;12(1):77 [FREE Full text] [doi: [10.1186/s13012-017-0605-9](https://doi.org/10.1186/s13012-017-0605-9)] [Medline: [28637486](https://pubmed.ncbi.nlm.nih.gov/28637486/)]
21. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005 Nov;15(9):1277-1288. [doi: [10.1177/1049732305276687](https://doi.org/10.1177/1049732305276687)] [Medline: [16204405](https://pubmed.ncbi.nlm.nih.gov/16204405/)]
22. Yamada J, Potestio ML, Cave AJ, Sharpe H, Johnson DW, Patey AM, et al. Using the theoretical domains framework to identify barriers and enablers to pediatric asthma management in primary care settings. *J Asthma* 2018 Nov;55(11):1223-1236. [doi: [10.1080/02770903.2017.1408820](https://doi.org/10.1080/02770903.2017.1408820)] [Medline: [29261346](https://pubmed.ncbi.nlm.nih.gov/29261346/)]
23. Sun Y, Wang N, Guo X, Peng Z. Understanding the acceptance of mobile health services: a comparison and integration of alternative models. *J Electron Commer Res* 2013;14(2):183-200.
24. Kothari A, McCutcheon C, Graham ID. Defining integrated knowledge translation and moving forward: a response to recent commentaries. *Int J Health Policy Manag* 2017 May 1;6(5):299-300 [FREE Full text] [doi: [10.15171/ijhpm.2017.15](https://doi.org/10.15171/ijhpm.2017.15)] [Medline: [28812820](https://pubmed.ncbi.nlm.nih.gov/28812820/)]
25. Johnson K, Gustafson D, Ewigman B, Provost L, Roper R. Using Rapid-Cycle Research to Reach Goals: Awareness, Assessment, Adaptation, Acceleration. Agency for Healthcare Research and Quality. 2015. URL: <https://pbrn.ahrq.gov/sites/default/files/docs/page/AHRQPBRNFinalRapidCycleResearchGuidanceDocument.pdf> [accessed 2020-09-19]
26. Laxman K, Krishnan SB, Dhillon JS. Barriers to adoption of consumer health informatics applications for health self management. *Health Sci J* 2015;9(5):1-7.
27. Howell M, Hood AJ, Jayne DG. Use of a patient completed iPad questionnaire to improve pre-operative assessment. *J Clin Monit Comput* 2017 Feb;31(1):221-225. [doi: [10.1007/s10877-015-9818-0](https://doi.org/10.1007/s10877-015-9818-0)] [Medline: [26715416](https://pubmed.ncbi.nlm.nih.gov/26715416/)]
28. Fritz F, Balhorn S, Riek M, Breil B, Dugas M. Qualitative and quantitative evaluation of EHR-integrated mobile patient questionnaires regarding usability and cost-efficiency. *Int J Med Inform* 2012 May;81(5):303-313. [doi: [10.1016/j.jimedinf.2011.12.008](https://doi.org/10.1016/j.jimedinf.2011.12.008)] [Medline: [22236957](https://pubmed.ncbi.nlm.nih.gov/22236957/)]
29. Barentsz MW, Wessels H, van Diest PJ, Pijnappel RM, Haaring C, van der Pol CC, et al. Tablet, web-based, or paper questionnaires for measuring anxiety in patients suspected of breast cancer: patients' preferences and quality of collected data. *J Med Internet Res* 2014 Oct 31;16(10):e239 [FREE Full text] [doi: [10.2196/jmir.3578](https://doi.org/10.2196/jmir.3578)] [Medline: [25364951](https://pubmed.ncbi.nlm.nih.gov/25364951/)]
30. Martin P, Brown MC, Espin-Garcia O, Cuffe S, Pringle D, Mahler M, et al. Patient preference: a comparison of electronic patient-completed questionnaires with paper among cancer patients. *Eur J Cancer Care (Engl)* 2016 Mar;25(2):334-341. [doi: [10.1111/ecc.12318](https://doi.org/10.1111/ecc.12318)] [Medline: [25899560](https://pubmed.ncbi.nlm.nih.gov/25899560/)]
31. Pieterse AH, Stiggelbout AM, Montori VM. Shared decision making and the importance of time. *J Am Med Assoc* 2019 Jul 2;322(1):25-26. [doi: [10.1001/jama.2019.3785](https://doi.org/10.1001/jama.2019.3785)] [Medline: [31002337](https://pubmed.ncbi.nlm.nih.gov/31002337/)]
32. Kiesler DJ, Auerbach SM. Optimal matches of patient preferences for information, decision-making and interpersonal behavior: evidence, models and interventions. *Patient Educ Couns* 2006 Jun;61(3):319-341. [doi: [10.1016/j.pec.2005.08.002](https://doi.org/10.1016/j.pec.2005.08.002)] [Medline: [16368220](https://pubmed.ncbi.nlm.nih.gov/16368220/)]
33. Légaré F, Adekpedjou R, Stacey D, Turcotte S, Kryworuchko J, Graham ID, et al. Interventions for increasing the use of shared decision making by healthcare professionals. *Cochrane Database Syst Rev* 2018 Jul 19;7:CD006732 [FREE Full text] [doi: [10.1002/14651858.CD006732.pub4](https://doi.org/10.1002/14651858.CD006732.pub4)] [Medline: [30025154](https://pubmed.ncbi.nlm.nih.gov/30025154/)]
34. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. *JMIR Mhealth Uhealth* 2017 Dec 6;5(12):e177 [FREE Full text] [doi: [10.2196/mhealth.7886](https://doi.org/10.2196/mhealth.7886)] [Medline: [29212629](https://pubmed.ncbi.nlm.nih.gov/29212629/)]
35. Loughheed MD, Leniere C, Ducharme FM, Licskai C, Dell SD, Rowe BH, Canadian Thoracic Society Asthma Clinical Assembly. Canadian Thoracic Society 2012 guideline update: diagnosis and management of asthma in preschoolers, children and adults: executive summary. *Can Respir J* 2012;19(6):e81-e88 [FREE Full text] [doi: [10.1155/2012/214129](https://doi.org/10.1155/2012/214129)] [Medline: [23248807](https://pubmed.ncbi.nlm.nih.gov/23248807/)]
36. Bourbeau J, Bhutani M, Hernandez P, Aaron SD, Balter M, Beauchesne M, et al. Canadian Thoracic Society Clinical Practice Guideline on pharmacotherapy in patients with COPD – 2019 update of evidence. *Can J Resp Crit Care Sleep Med* 2019 Oct 18;3(4):210-232. [doi: [10.1080/24745332.2019.1668652](https://doi.org/10.1080/24745332.2019.1668652)]



37. McKelvie RS, Moe GW, Ezekowitz JA, Heckman GA, Costigan J, Ducharme A, et al. The 2012 Canadian Cardiovascular Society heart failure management guidelines update: focus on acute and chronic heart failure. *Can J Cardio* 2013 Feb;29(2):168-181. [doi: [10.1016/j.cjca.2012.10.007](https://doi.org/10.1016/j.cjca.2012.10.007)]
38. Fisher J, Clayton M. Who gives a tweet: assessing patients' interest in the use of social media for health care. *Worldviews Evid Based Nurs* 2012 Apr;9(2):100-108. [doi: [10.1111/j.1741-6787.2012.00243.x](https://doi.org/10.1111/j.1741-6787.2012.00243.x)] [Medline: [22432730](https://pubmed.ncbi.nlm.nih.gov/22432730/)]
39. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. *JMIR Mhealth Uhealth* 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: [10.2196/11223](https://doi.org/10.2196/11223)] [Medline: [30990458](https://pubmed.ncbi.nlm.nih.gov/30990458/)]
40. Guidance for health app developers, commissioners and assessors. NHS Digital. 2019. URL: <https://digital.nhs.uk/services/nhs-apps-library/guidance-for-health-app-developers-commissioners-and-assessors> [accessed 2020-09-19]
41. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. *Appl Psychol* 2008;57(4):660-680. [doi: [10.1111/j.1464-0597.2008.00341.x](https://doi.org/10.1111/j.1464-0597.2008.00341.x)]
42. Cane J, Richardson M, Johnston M, Ladha R, Michie S. From lists of behaviour change techniques (BCTs) to structured hierarchies: comparison of two methods of developing a hierarchy of BCTs. *Br J Health Psychol* 2015 Feb;20(1):130-150. [doi: [10.1111/bjhp.12102](https://doi.org/10.1111/bjhp.12102)] [Medline: [24815766](https://pubmed.ncbi.nlm.nih.gov/24815766/)]
43. Kew KM, Malik P, Aniruddhan K, Normansell R. Shared decision-making for people with asthma. *Cochrane Database Syst Rev* 2017 Oct 3;10:CD012330 [FREE Full text] [doi: [10.1002/14651858.CD012330.pub2](https://doi.org/10.1002/14651858.CD012330.pub2)] [Medline: [28972652](https://pubmed.ncbi.nlm.nih.gov/28972652/)]

## Abbreviations

**AAP:** asthma action plan  
**BCT:** behavior change technique  
**CDSS:** computerized decision support system  
**eAMS:** Electronic Asthma Management System  
**EHR:** electronic health record  
**mHealth:** mobile health  
**NHS:** National Health Service  
**TDF:** Theoretical Domains Framework

*Edited by G Eysenbach; submitted 21.04.20; peer-reviewed by D Ryan, E Arden-Close; comments to author 29.07.20; revised version received 24.08.20; accepted 01.09.20; published 08.10.20.*

### *Please cite as:*

Yamada J, Kouri A, Simard SN, Segovia SA, Gupta S

Barriers and Enablers to Using a Patient-Facing Electronic Questionnaire: A Qualitative Theoretical Domains Framework Analysis  
*J Med Internet Res* 2020;22(10):e19474

URL: <http://www.jmir.org/2020/10/e19474/>

doi: [10.2196/19474](https://doi.org/10.2196/19474)

PMID: [33030437](https://pubmed.ncbi.nlm.nih.gov/33030437/)

©Janet Yamada, Andrew Kouri, Sarah-Nicole Simard, Stephanie A Segovia, Samir Gupta. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org/>), 08.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Automating the Generation of Antimicrobial Resistance Surveillance Reports: Proof-of-Concept Study Involving Seven Hospitals in Seven Countries

Cherry Lim<sup>1,2</sup>, MSc; Thyl Miliya<sup>3</sup>, MD; Vilada Chansamouth<sup>4</sup>, MD; Myint Thazin Aung<sup>5</sup>, PhD; Abhilasha Karkey<sup>2,6,7</sup>, PhD; Prapit Teparrukkul<sup>8</sup>, MD; Batra Rahul<sup>9</sup>, MD; Nguyen Phu Huong Lan<sup>10</sup>, PhD; John Stelling<sup>11</sup>, MPH, MD; Paul Turner<sup>2,3</sup>, FRCP; Elizabeth Ashley<sup>2,4,12</sup>, MBBS; H Rogier van Doorn<sup>2,13</sup>, MD; Htet Naing Lin<sup>12</sup>, MD; Clare Ling<sup>14</sup>, PhD; Soawapak Hinjoy<sup>15,16</sup>, DrPH; Sopon Iamsirithaworn<sup>17</sup>, PhD; Susanna Dunachie<sup>1,2</sup>, PhD; Tri Wangrangsamakul<sup>1,2</sup>, MRCPE, FRCP; Viriya Hantrakun<sup>1</sup>, PhD; William Schilling<sup>1,2</sup>, MD; Lam Minh Yen<sup>13</sup>, MD; Le Van Tan<sup>13</sup>, PhD; Htay Htay Hlaing<sup>5</sup>, MD; Mayfong Mayxay<sup>2,4,18</sup>, MD; Manivanh Vongsouvath<sup>4</sup>, MD; Buddha Basnyat<sup>2,6,7</sup>, FRCPE; Jonathan Edgeworth<sup>9</sup>, FRCP; Sharon J Peacock<sup>19</sup>, FRCP; Guy Thwaites<sup>2,13</sup>, FRCP; Nicholas PJ Day<sup>1,2</sup>, FRCP; Ben S Cooper<sup>1,2</sup>, PhD; Direk Limmathurotsakul<sup>1,2</sup>, PhD

<sup>1</sup>Mahidol Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

<sup>2</sup>Nuffield Department of Medicine, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, United Kingdom

<sup>3</sup>Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia

<sup>4</sup>Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit, Mahosot Hospital, Vientiane, Lao People's Democratic Republic

<sup>5</sup>North Okkalapa General and Teaching Hospital, Yangon, Myanmar

<sup>6</sup>Patan Hospital, Kathmandu, Nepal

<sup>7</sup>Oxford University Clinical Research Unit, Patan Hospital, Kathmandu, Nepal

<sup>8</sup>Sunpasitthiprasong Hospital, Ubon Ratchathani, Thailand

<sup>9</sup>Department of Infectious Diseases, Centre for Clinical Infection and Diagnostic Research, King's College London & Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

<sup>10</sup>Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam

<sup>11</sup>Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

<sup>12</sup>Myanmar Oxford Clinical Research Unit, Yangon, Myanmar

<sup>13</sup>Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

<sup>14</sup>Shoklo Malaria Research Unit and Mahidol Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, Thailand

<sup>15</sup>Department of Disease Control, Bureau of Epidemiology, Ministry of Public Health, Nonthaburi, Thailand

<sup>16</sup>Department of Disease Control, Office of International Cooperation, Ministry of Public Health, Nonthaburi, Thailand

<sup>17</sup>Division of Communicable Diseases, Department of Disease Control, Ministry of Public Health, Nonthaburi, Thailand

<sup>18</sup>Institute of Research and Education Development, University of Health Sciences, Vientiane, Lao People's Democratic Republic

<sup>19</sup>Department of Medicine, University of Cambridge, Cambridge, United Kingdom

## Corresponding Author:

Cherry Lim, MSc

Mahidol Oxford Tropical Medicine Research Unit

Faculty of Tropical Medicine

Mahidol University

3rd Floor, 60th Anniversary Chalermprakiat Building

420/6 Ratchawithi Rd, Ratchathewi District

Bangkok, 10400

Thailand

Phone: 66 2 203 6333 ext 7352

Email: [cherry@tropmedres.ac](mailto:cherry@tropmedres.ac)

## Abstract

**Background:** Reporting cumulative antimicrobial susceptibility testing data on a regular basis is crucial to inform antimicrobial resistance (AMR) action plans at local, national, and global levels. However, analyzing data and generating a report are time consuming and often require trained personnel.

**Objective:** This study aimed to develop and test an application that can support a local hospital to analyze routinely collected electronic data independently and generate AMR surveillance reports rapidly.

**Methods:** An offline application to generate standardized AMR surveillance reports from routinely available microbiology and hospital data files was written in the R programming language (R Project for Statistical Computing). The application can be run by double clicking on the application file without any further user input. The data analysis procedure and report content were developed based on the recommendations of the World Health Organization Global Antimicrobial Resistance Surveillance System (WHO GLASS). The application was tested on Microsoft Windows 10 and 7 using open access example data sets. We then independently tested the application in seven hospitals in Cambodia, Lao People's Democratic Republic, Myanmar, Nepal, Thailand, the United Kingdom, and Vietnam.

**Results:** We developed the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS), which can support clinical microbiology laboratories to analyze their microbiology and hospital data files (in CSV or Excel format) onsite and promptly generate AMR surveillance reports (in PDF and CSV formats). The data files could be those exported from WHONET or other laboratory information systems. The automatically generated reports contain only summary data without patient identifiers. The AMASS application is downloadable from <https://www.amass.website/>. The participating hospitals tested the application and deposited their AMR surveillance reports in an open access data repository.

**Conclusions:** The AMASS is a useful tool to support the generation and sharing of AMR surveillance reports.

(*J Med Internet Res* 2020;22(10):e19762) doi:[10.2196/19762](https://doi.org/10.2196/19762)

## KEYWORDS

antimicrobial resistance; surveillance; report; data analysis; application

## Introduction

Generating and sharing antimicrobial resistance (AMR) surveillance reports are fundamental elements of actions against AMR infections at local, national, and international levels. Information on patterns of antimicrobial susceptibility is important to guide empiric choice of therapy, monitor resistance trends, and detect outbreaks of AMR infections at the local level [1-3]. Combining data and reports at the national level provides evidence to inform the implementation of national action plans, decide on resource allocation for interventions, and monitor the impact of those interventions [3-5]. The Review on AMR chaired by Jim O'Neill estimated that 700,000 global deaths are attributable to AMR infections each year (including bacterial infections and tuberculosis) [1,6], and they have an enormous global impact [7,8]. While this represents a very rough estimate subject to well-documented limitations [1,9,10], the report importantly highlighted the need for improved AMR surveillance.

Methods to analyze data and generate AMR surveillance reports are gradually being standardized worldwide [3,11-15]. Recently, the World Health Organization (WHO) launched the Global Antimicrobial Resistance Surveillance System (GLASS) with a defined protocol for AMR surveillance data collection for certain high-priority pathogens and resistance phenotypes [2,11,16]. In general, it is recommended that (1) repeat isolates of a given bacterial species from individual patients should be removed from the calculations and (2) data should be stratified by origin of infection (community or hospital) whenever possible [11,15]. A simple deduplication process is to include only the first isolate of a species per patient per specimen type per survey period in the report [3,11]. The origin of infection

is defined by using specimen collection date, location type (inpatient or outpatient setting), and hospital admission date for inpatient isolates as a proxy to define where the infection was likely contracted (community or hospital) [3,11].

Even in areas where AMR surveillance data are available, there are many barriers to utilizing such data [17]. Many hospitals in low and middle-income countries (LMICs) lack the time and resources needed to analyze the data (particularly to deduplicate and validate the accuracy of the summary data), write reports, and release the data or reports [17]. There are open-access laboratory information systems (LISs) [18] and microbiology laboratory database software, including WHONET [19], that are useful for recording and analyzing the data, and generating figures to support the generation of AMR surveillance reports for hospitals in LMICs. Most of these systems have limited ability to generate summary reports for immediate use and be operated by nontechnical staff. In addition, to generate AMR surveillance reports stratified by the origin of infection, additional data on hospital admission are frequently needed [20-23]. This is because hospital admission dates are not generally collected in microbiology laboratory data files. In many hospital settings, microbiology and hospital admission data are held in separate computers or systems with restricted access. Even in high-income countries, many hospitals lack well-trained clinical microbiologists, epidemiologists, or data experts with adequate skills in statistical software (such as R, SAS, SPSS, and STATA) to merge and deduplicate data in the separated databases and generate reports stratified by the origin of infection.

Here, we developed an application termed the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS), which can support a local hospital to independently analyze

routinely collected electronic data and rapidly generate AMR surveillance reports. We tested the AMASS in seven hospitals in Cambodia, Lao People's Democratic Republic, Myanmar, Nepal, Thailand, the United Kingdom, and Vietnam, and deposited the report from each hospital in an open-access platform.

## Methods

### Design of the Application

The tool operates by reading and processing the raw data files to automatically produce AMR surveillance reports. The application was designed to be open access, user friendly, and highly compatible with readily available data sets at local hospitals, and have high data security. To ensure the tool is fully open access and can run in a standalone inexpensive computer even without internet access, we built the application in R (version 3.6.2; R Project for Statistical Computing), which is a

free software environment. We then gave the application a user-friendly interface, which only requires double clicking on the application file to run the automation without the need to understand the R program. We decided to include both R portable (version 3.4.3; R Project for Statistical Computing) and RStudio (version 1.1.423; RStudio, Inc) within the downloadable package so that the application can run without the need to install R or any program prior to running the application. We designed the application so that it reads raw data files in either CSV or Excel format as many hospitals in LMICs can export their data in one of these formats [17]. Hospitals that store microbiology data and hospital data in software, such as WHONET [19], can export data in CSV or Excel format. During the development process, we revised the R codes repeatedly to ensure that the application would produce easy-to-use and easy-to-share outputs. Table 1 documents the features that we focused on when designing the application [24,25].

**Table 1.** Features of the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS).

Feature	Description
Open access	<p>The AutoMated tool for Antimicrobial resistance Surveillance System (AMASS) is open access and can be downloaded [24,25].</p> <p>The AMASS was developed using R, which is a free software. In the download package for the AMASS (AMASS.zip), there is a folder that contains R-portable and RStudio, which support data processing and analysis to generate the antimicrobial resistance (AMR) surveillance report automatically.</p> <p>The AMASS is under CC-BY 4.0 license. Users can share (copy and redistribute the material in any medium or format) and modify the R codes of the AMASS under the terms and conditions of the Creative Commons license.</p>
User friendly	<p>The AMASS can be run by double clicking on the application icon. Data analysis and AMR surveillance report generation are automated by the AMASS application.</p> <p>Data cleaning, deduplication, and analysis are performed rapidly (it takes about 1-3 minutes to automatically produce an AMR surveillance report using example data sets provided in the AMASS package).</p> <p>No additional program or software is needed. All the essential software is stored in the AMASS package and will operate automatically after double clicking the application file (AMASS.bat).</p> <p>Users do not need to understand R program or write any codes to run the AMASS application.</p>
Highly compatible	<p>The AMASS works with raw data files in either CSV or Excel format, which can be commonly exported from WHONET and other software, programs, or data management systems used for microbiology data and hospital admission data.</p> <p>The AMASS uses data dictionary files (in Excel format) to accommodate data exported from different software, programs, or systems that may have different ways to name data variables and data values.</p> <p>The AMASS dictionary files can be reused by users in the future (eg, monthly, quarterly, and yearly) if the structures of the new raw microbiology data file and hospital admission data file remain unchanged.</p> <p>The AMASS uses a tier-based approach based on availability of raw data files to generate reports. Users with limited data availability (eg, microbiology data with only culture positive results) can still utilize the deduplication and report generation functions of the AMASS. Users with additional data (eg, microbiology data with culture negative results and hospital admission date data) will receive additional reports (eg, sample-based surveillance reports with stratification by infection origin).</p>
High data security	<p>The AMASS does not require the internet for operation. Users do not have to transfer raw individual data (which may contain identifiable information) to any institution outside of the hospital to analyze the data and generate the reports. The AMASS can be run on a standalone computer within the local hospital under local data security. Hence, the AMASS does not increase any risks of breaching individual patient data confidentiality.</p>
Easy-to-use outputs	<p>The automatically generated AMR surveillance report is in PDF format, which is easy to print, read, and share within and outside the hospital.</p>
Easy-to-share outputs	<p>The report (in PDF format) and aggregated summary data files (in CSV format) contain no individual-level patient data and can be readily shared with national and international organizations.</p>

A schematic overview of input data, data processing, statistical analysis, and output is shown in [Multimedia Appendix 1](#). In brief, input data are microbiology data files (in CSV or Excel format) with or without hospital data files, which can be exported from WHONET or a LIS with data export capacity. The data processing and analysis algorithms, including data deduplication, were developed based on the WHO GLASS recommendations [2,11]. The reports were designed to have a format similar to that of the WHO GLASS 2018 report [14].

The application deduplicates the data by including only the first isolate per sample type per pathogen per survey period for each patient [11]. The application currently includes only blood specimens and the following eight pathogens: *Acinetobacter* spp., *Escherichia coli*, *Enterococcus* spp., *Klebsiella pneumoniae*, *Salmonella* spp., *Staphylococcus aureus*, *Streptococcus pneumoniae*, and *Pseudomonas aeruginosa*. Both *Enterococcus* spp. and *P. aeruginosa* were added on top of the six priority pathogens described by the WHO GLASS for bacteremia [2,26] because both pathogens are common causes of bacteremia [15,20], and were included in the 2015 global priority list of AMR bacteria [26]. The list of pathogen-antibiotic combinations was modified from the WHO GLASS ([Multimedia Appendix 2](#)). Infections are stratified into community-origin or hospital-origin infections using hospital admission dates and specimen collection dates, when available [2,11]. Patients with the first specimen culture positive for the pathogen taken in the outpatient setting or on the first or second day of hospitalization are classified as having community-origin infection [2,11]. Patients with the first specimen culture positive for the pathogen taken on hospital day three or later are classified as having hospital-origin infection [2,11]. Alternatively, in cases where users have data on the origin of infection assigned by the attending physician or infection control team of the hospital, the AMASS can instead use those categorizations to stratify the infection into community-origin or hospital-origin infection. Prevalence and incidence rates are estimated based on the recommendations of the WHO GLASS ([Multimedia Appendix 3](#)) [2,11]. An additional report on mortality involving AMR and non-AMR infections is generated when mortality data are available in the hospital admission data file. The term “mortality involving AMR and antimicrobial-susceptible infections” is used because the mortality reported in the hospital admission data represents mortality from any cause, not necessarily attributable to AMR [1]. This measure of mortality includes deaths caused or contributed by other underlying and intermediate causes. Therefore, the term “involving AMR infections” is used, in accordance with the term used by the UK Office for National Statistics [27,28]. The AMASS used the Wilson method to estimate the confidence intervals for proportions.

### Example Data Sets

Two example data sets are provided in the downloadable package of the AMASS application. The first example data set is the open-access demonstration file from WHONET [19]. The second example data set is a synthetic data set generated for the AMASS. The second example data set was created to represent a large data set from a 1000-bed hospital, containing both microbiology and hospital admission data. A detailed description

on how the second example data set was generated is provided in [Multimedia Appendix 4](#). In brief, two synthetic data files (microbiology\_data.xls and hospital\_admission\_data.xls files) were generated based on the summary data in the AMR surveillance report from 2015 of Sunpasitthiprasong Hospital, Ubon Ratchathani, Thailand. Variables in the microbiology data file include hospital number, specimen type, specimen collection date, culture result, and antibiotic susceptibility testing result, and each row contains information for each specimen. Variables in the hospital admission data file include hospital number, age, sex, admission date, discharge date, and in-hospital discharge outcome. All data were randomly generated using STATA 15.1 (StataCorp).

### Testing the AMASS Using Hospital Data

The following seven hospitals participated in the study: Angkor Hospital for Children in Cambodia, Mahosot Hospital in Lao People's Democratic Republic, North Okkalapa General and Teaching Hospital in Myanmar, Patan Hospital in Nepal, Sunpasitthiprasong Hospital in Thailand, St Thomas' Hospital in the United Kingdom, and Hospital for Tropical Diseases in Vietnam. The hospitals were selected because microbiology data are collected routinely, and they have prior experience in data quality controls. Moreover, the hospitals varied in the LIS used for data storage and are good examples to demonstrate the practicality and usability of the AMASS in different settings, to a wider audience. The study was approved by the Oxford Tropical Research Ethics Committee, University of Oxford, and local Ethics Committees. CL corresponded with the participating hospitals and demonstrated how the application operates using the example data files. The local hospital staff operated the application using their local data by themselves. All microbiology and hospital data were stored independently within their hospital computers under their local data protection standards. Automatically generated reports and anonymous summary data from each hospital are publicly available and deposited in data repositories with permission from each hospital.

## Results

### Overview of the AMASS

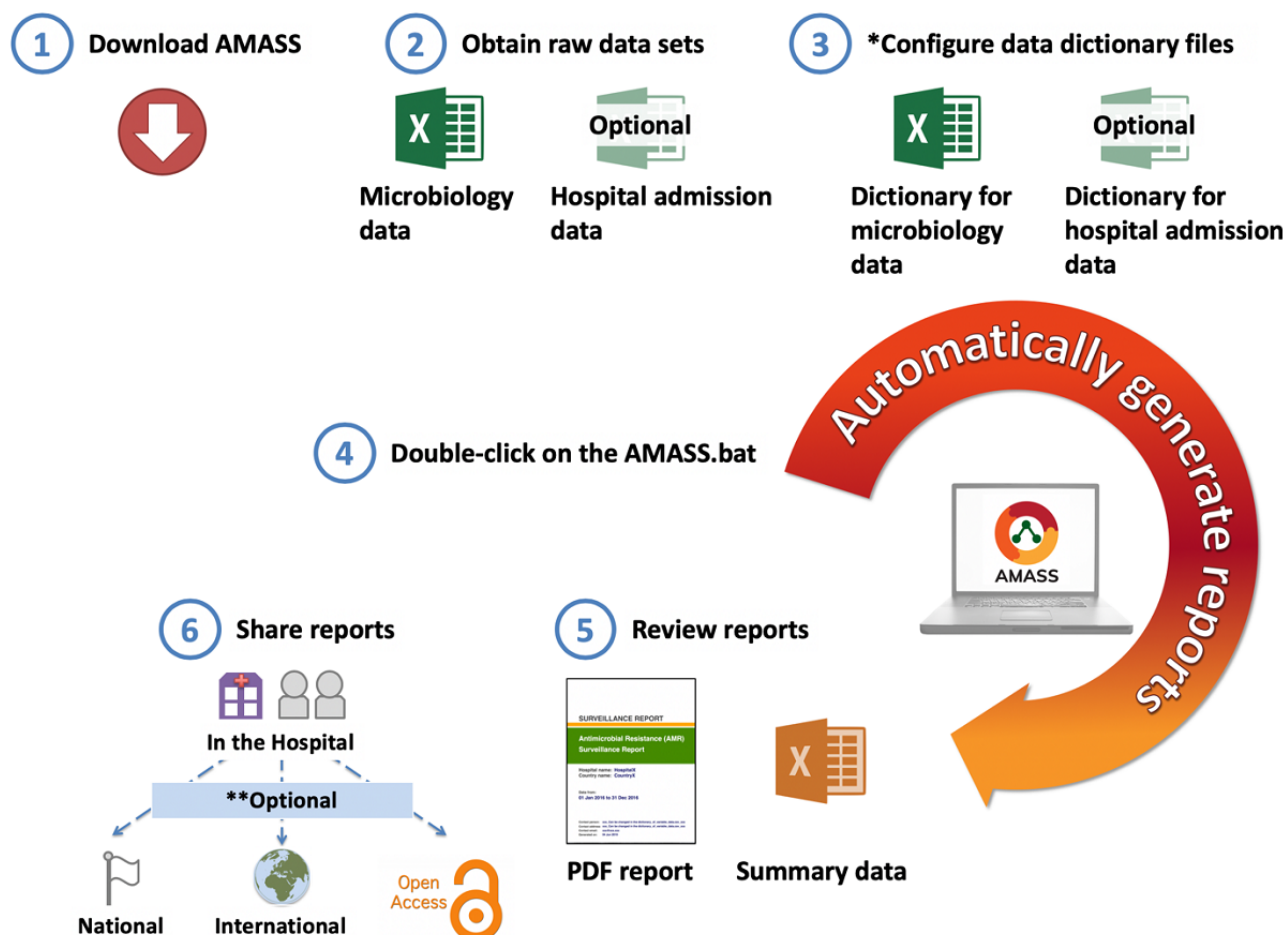
The AMASS application has been developed to support clinical microbiology laboratories to automatically analyze hospital local data files (in CSV or Excel format) and generate AMR surveillance reports (in PDF and CSV formats) promptly. Six steps are followed to generate the AMR surveillance reports ([Figure 1](#) and [Multimedia Appendix 5](#)). First, download the AMASS package from the website [24]. Second, obtain the routinely collected raw microbiology data file and, if available, hospital admission data file, and then, save the data files in the folder of the AMASS application. Third, configure data dictionaries. Two data dictionaries are provided to accommodate different ways of naming variables (eg, sex and gender) and data values (eg, M and F, or male and female). The functioning of the data dictionaries is described in more detail in the “Illustration on How to Use the AMASS” section. Fourth, double click on the AMASS.bat file to run the application. Fifth, review and validate the AMR surveillance reports (generated



in PDF format) and the anonymous summary data (generated in CSV format). Sixth, share the reports within the hospital, especially with the local infection control team. The reports and anonymous summary data contain no patient identifiers.

Therefore, users may share the reports and anonymous summary data with national and international organizations or make the reports and anonymous summary data open access. The key features of the AMASS are listed in [Table 1](#).

**Figure 1.** Conceptual flow of the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS). Step 1 (download the AMASS) and step 3 (configure data dictionary files) are one-time steps. Step 2 (obtain data), step 4 (run the AMASS), step 5 (review report), and step 6 (share report) are ongoing steps that users could repeat regularly (ie, monthly or quarterly). \*Two data dictionary files (in Excel format) are provided to allow the application to understand how variables and values of each variable are named in the raw data files in different settings. Those data dictionary files can be reused in the subsequent runs of the AMASS, as long as how variables and values of each variable are named in the raw data files remain the same. Details on how to configure the data dictionary can be found in [Figure 2](#) and [Multimedia Appendix 6](#). \*\*The antimicrobial resistance (AMR) surveillance report and summary data generated contain no patient identifiable information. The decision to share the report and summary data to national or international AMR organizations is solely up to the jurisdiction of the hospital.



## Illustration on How to Use the AMASS

### Input Requirements

To align with formats of commonly exported microbiology data and hospital admission data, the AMASS reads data files in both CSV and Excel formats. The microbiology data need to be in a wide format, meaning that each row should contain data from a single clinical isolate. The key variables required for the microbiology data file include patient identifier, specimen collection date, specimen type, culture result, and antimicrobial susceptibility test interpretation per antibiotic. The hospital admission data file is optional but, if available, also needs to be in a wide format. The key variables needed for the hospital admission data file include patient identifier, admission date, gender, age, and, if possible, in-hospital discharge outcome. The current version of the AMASS uses the Gregorian calendar

and requires the date to be in the order of day, month, and year in any format (ie, either text [English] or numeric).

There are two data dictionary files provided for the users to accommodate different ways of naming data variables and data values ([Figure 2](#)). The first data dictionary file (`dictionary_for_microbiology_data.xls`) is for the microbiology data file. For example, the AMASS uses the variable name “hospital\_number” as a patient identifier (Row 3, Column A of the data dictionary file). In cases where the raw microbiology data file uses a different name for the patient identifier (eg, hn), users would need to fill “hn” in the data dictionary file (Row 3, Column B of the data dictionary file). This allows the AMASS application to know that the variable “hn” of the raw microbiology data file is the patient identifier (ie, “hospital\_number”). The second data dictionary file (`dictionary_for_hospital_admission_data.xls`) is for the hospital admission data file, which is to be used likewise. [Multimedia](#)



Appendix 6 presents a step-by-step tutorial on how to use and configure the data dictionaries.

**Figure 2.** An example of how to complete a data dictionary file. For a first-time user, the user may need to complete a data dictionary file by filling in variable names used in the raw data files into the data dictionary files (eg, arrow A). This is to allow the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS) to understand that the variable “hospital\_number” used by the AMASS is named as “hn” in the user’s raw microbiology data file. Thereafter, users need to enter how data values are named in their raw data files (e.g. arrow B). This is to allow the AMASS to understand that the data value named “blood\_specimen” is named as “blood” in user’s raw microbiology data file. Please note that the contents in the first column of the data dictionary file must remain unchanged. Users can add new rows but the content in the cell in the first column must not be changed. For example, users can define that both “*E. coli* (ESBL-producing strain)” and “*Escherichia coli*” in their raw microbiology data file mean “organism\_escherichia\_coli” by the AMASS. The example data dictionary files shown in the figure are available in the Example\_Dataset\_2 folder (within the AMASS download package).

**An example dictionary for microbiology data file**

A	B	C	D
"Don't change content in this column, but you can add rows with the same content if needed"			
"Change content in this column to represent how variable names, antibiotic names, or variable values are written in your raw microbiology data file"		Requirements	Explanation
1	Variable names used in AMASS	Variable names used in your microbiology data file	
2	hospital_number	hn	Required
3	specimen_collection_date	spcdate	Required
4	specimen_type	spctype	Required
5	organism	organism	Required
6	infection_origin		Optional
7	Data values described in AMASS	Data values used in your microbiology data file	
8	blood_specimen	blood	Required
9	organism_no_growth	no growth	Required

**An example microbiology data file**

A	B	C	D	E	
1	hn	spcdate	spctype	organism	FOX
2	608	17dec2016	blood	no growth	
3	609	28may201	blood	no growth	
4	610	10apr2016	blood	no growth	
5	611	04oct2016	blood	other	
6	612	14apr2016	blood	no growth	
7	613	02jan2017	blood	no growth	
3705	196	12dec2016	blood	Escherichia coli	
3706	197	28nov2016	blood	Escherichia coli	
3707	198	17oct2016	blood	Escherichia coli	

**Callout boxes:**

- The section under this heading is for **variable names**
- The section under this heading is for **data values**

**Arrows:**

- A: Points from the dictionary to the data file (e.g., from 'hn' in the dictionary to 'hn' in the data file).
- B: Points from the dictionary to the data file (e.g., from 'blood' in the dictionary to 'blood' in the data file).

### Outputs Generated by the AMASS

The AMR surveillance reports generated from the AMASS are illustrated using the two open-access example data sets provided within the download package. Multimedia Appendices 7 and 8 contain AMR surveillance reports generated from the first and second example data sets, respectively. Multimedia Appendix 5 illustrates how to test the AMASS using example data sets in a step-by-step fashion.

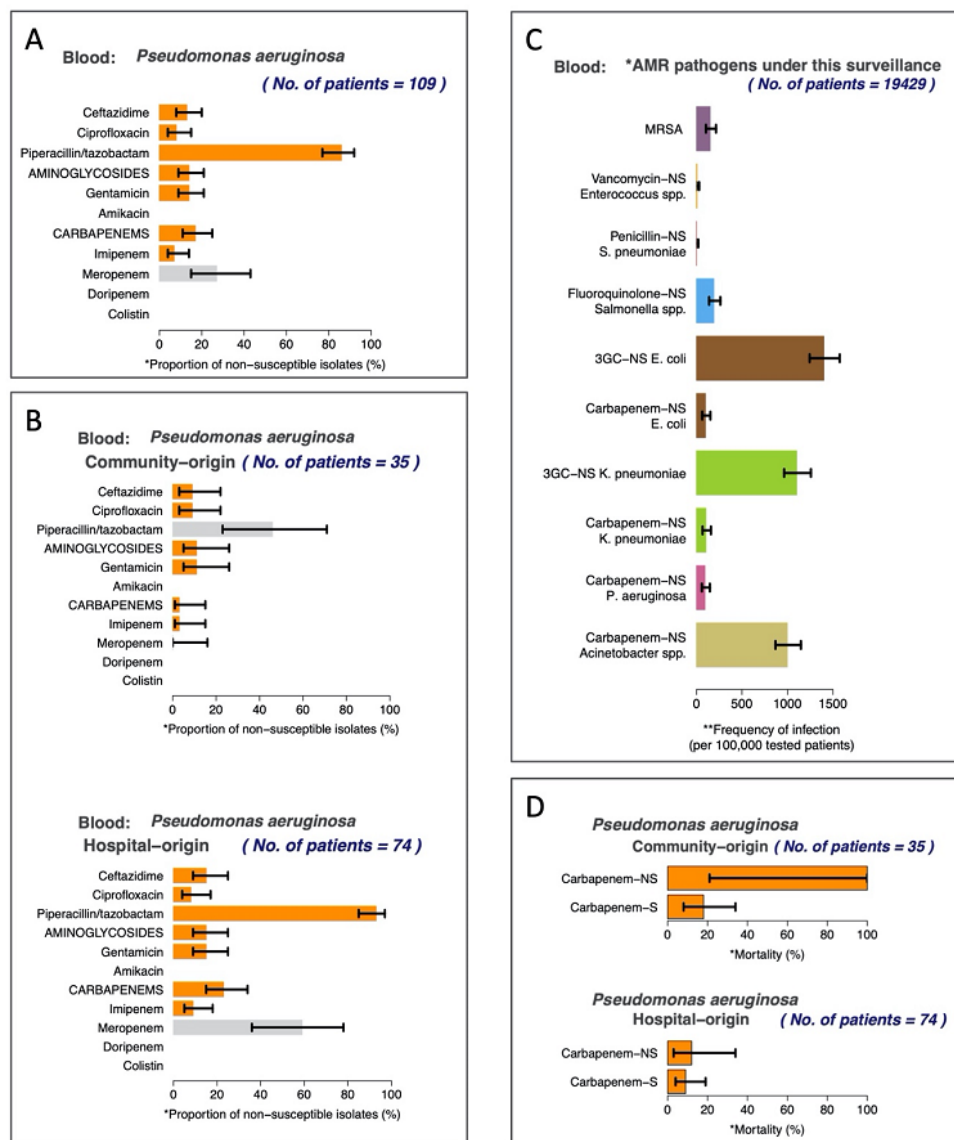
In short, the automatically generated report on AMR surveillance contains the following six sections (Figure 3): (1) data overview; (2) an isolate-based report; (3) an isolate-based report with stratification by origin of infection; (4) a sample-based report; (5) a sample-based report without stratification by infection; and (6) mortality involving AMR and antimicrobial-susceptible infections.

The AMASS uses a tier-based approach. When only the microbiology data file with the results of culture positive samples is available, only section one and two are automatically generated for users (as shown in Multimedia Appendix 7,

generated from the first example data set). Section three is generated only when data on admission dates are available. This is because these data are required for stratification by the origin of infection. Section four is generated only when data of culture negative specimens (no microbial growth) are available in the microbiology data file. This is because these data are required for the sample-based approach. Section five is section four stratified by the origin of infection and is generated if admission date data are also available. Section six is generated only when mortality data are available (as shown in Multimedia Appendix 8, generated from the second example data set).

The AMASS also generates two log files. The first log file (generated in PDF format) is for users to validate the input data used by the AMASS to generate the AMR surveillance report. It contains information such as the total number of records analyzed, age distribution, number of missing values, and total number of isolates per organism in the raw microbiology data file. The second log file (generated in plain text format) could be used for consultation with R users, statisticians, or the AMASS development team in case of any technical issue when running the AMASS.

**Figure 3.** Examples of figures automatically generated by the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS). All figures are from the report (Multimedia Appendix 8 Multimedia Appendix 8) automatically generated by the AMASS application using an example data set provided in the download package. Figure 3A represents the overall proportion of nonsusceptible (intermediate and resistant) isolates in an isolate-based report (section two in the report). Figure 3B represents the proportion of nonsusceptible isolates stratified by the origin of infection (section three in the report). Figure 3C represents the frequency of bloodstream infections per 100,000 tested patients (section four in the report). Figure 3D represents mortality involving antimicrobial-resistant and antimicrobial-susceptible bloodstream infections (section six in the report).

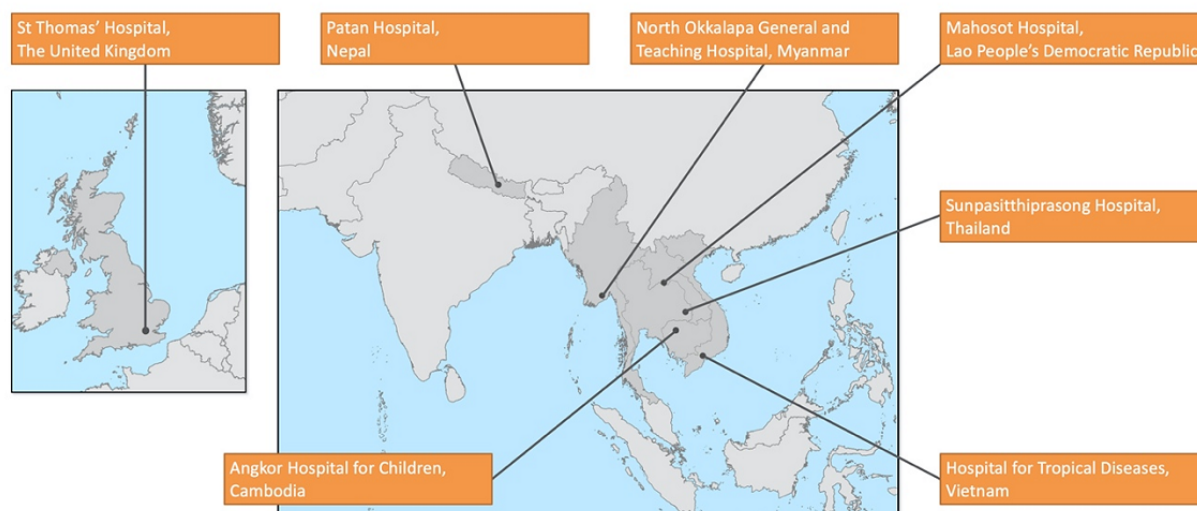


### Testing the AMASS Using Hospital Data

The AMASS was tested in seven hospitals in seven countries (Figure 4). The hospitals varied in data availability, data structure, naming of the variables, and definitions for data values (Multimedia Appendix 9). Overall, the proportions of patients

having *Escherichia coli* bacteremia caused by third-generation cephalosporin-resistant isolates ranged from 19% to 85%. The incidence rates of *Escherichia coli* bacteremia caused by third-generation cephalosporin-resistant isolates ranged from 283 to 2737 per 100,000 tested patients among participating hospitals with available data on negative cultures.

**Figure 4.** A map of participating hospitals and examples of summary data from the automatically generated antimicrobial resistance surveillance reports. The reports and summary data from St Thomas' Hospital, Patan Hospital, North Okkalapa General and Teaching Hospital, Mahosot Hospital, Sunpasitthiprasong Hospital, Hospital for Tropical Diseases, and Angkor Hospital for Children are open access [28-34].



All participating hospitals deposited their output files and their data dictionary files at figshare [29-35]. Angkor Hospital for Children, North Okkalapa General and Teaching Hospital, and Sunpasitthiprasong Hospital had all six sections of the AMR surveillance report, as microbiology and hospital admission data, data on negative cultures, and in-hospital discharge outcome data were available. Mahosot Hospital and Patan Hospital had sections one and two of the AMR surveillance report, as only microbiology data on positive cultures were available to test in the AMASS. St Thomas' Hospital had sections one, two, three, and six of the AMR surveillance report because negative culture data were not available to test in the AMASS. Hospital for Tropical Diseases had sections one, two, and four available, as hospital admission data were not available to test in the AMASS.

Data storage systems varied across the hospitals. Angkor Hospital for Children used the ACORN LIS that is based on the Microsoft Access program [36]. Mahosot Hospital used a local LIS that is also based on the Microsoft Access program [37]. Sunpasitthiprasong Hospital used the MLAB program [38]. St Thomas' Hospital used ICNET [39], which is a commercial clinical surveillance software package. North Okkalapa General and Teaching Hospital used WHONET 2019 (modernized version of WHONET 5.6) [19]. Patan Hospital and Hospital for Tropical Diseases used an in-house LIS based on the MySQL system. Inputting data dictionary files for the first time took about 1 to 3 hours. However, the data dictionary files could be stored and reused when the raw microbiology data file or hospital admission data file was revised or updated. North Okkalapa General and Teaching Hospital used the data dictionary file from the example data set (in the Example\_Dataset\_1\_WHONET folder), which was generated to comply with WHONET exported data (in .XLSX format; [Multimedia Appendix 10](#) illustrates how microbiology data were exported from WHONET 5.6). This saved the time and effort needed to complete the data dictionaries for the hospital.

We found that the AMASS took about 1 to 3 minutes to run and automatically generate an AMR surveillance report using

local data and local hospital computers. The AMASS works on Microsoft Windows 10 and with data containing a non-English (Latin and non-Latin characters) language. For example, “blood\_specimen” was recorded as “☒” in the raw microbiology excel file at Hospital for Tropical Diseases, Vietnam. As another example, “male” was recorded as “☒” and “female” was recorded as “☒” in the raw hospital admission excel file at Sunpasitthiprasong Hospital, Thailand. By completing data dictionary files, users could use the AMASS to analyze the data and generate the AMR surveillance report even though the raw data were recorded in a local language.

All participating hospitals had previous experience in data verification and in AMR surveillance report validation. All participating hospitals were familiar with the recommendations of the Clinical and Laboratory Standard Institute (CLSI) and European Committee on Antimicrobial Susceptibility Testing (EUCAST) on antimicrobial susceptibility testing (AST) data verification. Additionally, all participating hospitals were familiar with validating summary data by comparing the summary data automatically generated by the AMASS with manual calculations.

Users' suggestions for improvement were received when the investigation team visited the participating hospitals. The users expressed a need for including other clinical specimen types (eg, urine samples) and a function to support data verification or provide a list of isolates with unusual AST profiles. This is because a data verification function is not currently available in many in-house LISs. Users also expressed the need for a way to import multiple data files and different data formats, particularly the long format (ie, each row represents an AST result obtained from each antibiotic against each isolate). This is because the data exported from automated machines for AST testing are commonly in the long format and it takes time to integrate the data from the automated AST machines and from disk diffusion AST results. Users also noted that the requirement to manually export data from the existing LIS can be tedious



and time consuming to perform on a regular basis, and automation that can bypass the data extraction step would be useful.

## Discussion

We developed an open-access, offline, and easy-to-use application, AMASS, which allows hospitals, especially in LMICs, to automatically generate AMR surveillance reports from routinely collected electronic microbiology data. Seven hospitals used the AMASS, and shared their reports and summary data files by depositing them in an open-access data repository website. We propose that routine generation and sharing of AMR surveillance reports (ie, cumulative antimicrobial susceptibility reports or antibiograms) in open-access data repositories from hospitals with a microbiology laboratory should be supported worldwide [40].

The AMASS empowers data sharing by reducing the time and effort needed to prepare the summary data and reports without increasing the risks of breaching individual patient data confidentiality. This is because the AMASS can analyze the data and generate the reports promptly without needing to transfer raw data files to any party outside of the hospital. The autogenerated AMR surveillance reports are in PDF format, which can be reviewed and shared locally and immediately. This is important to improve the understanding of the local AMR burden and to inform local patient management. The PDF report and summary CSV files contain no individual patient-level information and can be shared with national and international organizations to support action plans on AMR. Any attempt to compare AMR surveillance reports among different hospitals should be done cautiously, as factors, including the type of hospital, blood culture utilization rates and practices, and patient characteristics, may influence the estimated prevalence and incidence rates of AMR infections [12-14]. An AMR surveillance report should clearly present the denominators used to calculate incidence rates. Selecting an appropriate denominator that represents the local setting is important for estimating the incidence rate of AMR, but remains challenging [41]. The AMASS calculates the total numbers of tested populations and uses them as denominators in section four and section five of the output report based on the WHO GLASS recommendations, and these denominators will not be available without data on negative cultures.

The AMASS uses currently recommended analytical approaches to generate the summary data and reports. The entire code is open access, verifiable, and modifiable. The AMASS can also be expanded and tailored based on local requirements. For instance, notifiable bacterial infections that are important for a local setting can be included by revising the R code provided in the AMASS package. Analytical methods can also be constantly updated and improved over time.

The AMASS is an add-on automatized report-generating tool that can be easily used, even by hospital users with limited information technology skills. It does not replace WHONET, LISs, quality assurance programs, or antimicrobial surveillance systems (including the WHO GLASS). The AMASS therefore differs from the “Macros and Excel Reports” function of

WHONET, which allows users to regularly generate reports on screen and export data as Excel files using macro language (ie, a series of written study parameters generated automatically through the WHONET user interface grouped together) [19]. In contrast, the AMASS is designed to automatically compile summary results of multiple organisms into a single report. The current version of the AMASS does not support quality improvement or alert individuals to unexpected antibiogram results, which WHONET is capable of doing [19]. Thus, while WHONET is an appropriate choice for the main software package for microbiology laboratories in LMICs to capture and store microbiology data and to export data for the WHO GLASS [19], the AMASS provides important additional functionality [19].

Verifying individual AST results and validating summary data in the reports (including those generated by the AMASS) remain major barriers to the generation of accurate surveillance data on AMR. Only verified AST results on each patient's isolate should be used in the data analysis [12-14]. WHONET and few LISs include functions that automatically check the results to ensure that they appear reasonable or alert users to confirm unusual results (eg, amikacin resistance coexisting with gentamicin and tobramycin susceptibility in *E. coli* is unusual, and data on such cases should be verified) [12-14,19]. In addition, it is important to validate the calculations of any analytical software used to generate the summary data and reports. We recommended that users of the AMASS perform manual validation (such as printing a line listing of all isolates of the species to cross-check with the reports), particularly when the program is used for the first time [12-14]. Users should also understand that the precision of the estimate is based on the sample size, which is illustrated by estimated confidence intervals. In addition, the representativeness of the summary cumulative AST data could be impacted by the sampling strategy used, particularly if blood culture is mainly performed for patients with treatment failure or prolonged hospitalization [11,12,14]. Nonetheless, generating AMR surveillance reports from existing microbiology databases could be the first step for hospitals in LMICs to understand and validate their own data, and support quality control programs at the local level.

Local hospitals must not be discredited because of the statistics in any AMR surveillance report, as possible criticism could be a barrier to data sharing [42]. Negative criticism of the data shared by local hospitals needs to be avoided. It is inevitable for errors to occur in any data set, especially in settings with limited resources and experience on quality assurance and data validation. Tools for data validation should be provided to local hospitals, and the statistics in an AMR report should act as a guide on the direction of improving data quality control, data validation, and infection prevention and control. Credit should be given to the hospitals that share their local AMR data and reports on an open-access data repository platform. Data repository platforms, such as figshare, allow flexibility on updating the data and revising the AMR surveillance reports with appropriate version controls by local hospitals. Publishing open-access AMR data on data repository platforms provides a digital object identifier (DOI) that should be used for data citation [43].

The AMASS has a number of limitations. First, the AMASS is not applicable for hospitals that only store data in paper forms. Second, the AMASS cannot work with raw microbiology data files that are not in a wide format or combine many files in multiple formats. For example, raw microbiology data file where each row contains data of each antibiotic susceptibility result for one single specimen. Third, the current version of the AMASS can only analyze microbiology data that include antimicrobial susceptibility test interpretive categories (susceptible, intermediate, and resistant) based on guidelines that the local hospital uses. Fourth, the AMASS cannot automatically validate the reliability of the data that are imported into the application and used for analysis. Data verification and quality checks will be included in future versions of the application. Fifth, the AMASS was tested with only few data sets that included non-English languages, and further testing on data with other languages may be needed. The current version

of the AMASS has been tested on Windows 10 and 7 and may not work on other operating systems. Sixth, the current version of the AMASS cannot provide options to export the AMR surveillance report in different formats (ie, in document or text format). However, users can reuse the summary statistics, which are saved in CSV format under the ResultData folder. Finally, feedback from participating hospitals was not formally evaluated. Studies evaluating the barriers, facilitators, and utilities of supporting applications, such as the AMASS, to generate AMR surveillance reports are needed.

In conclusion, the AMASS can be used to support hospitals with microbiology laboratories in analyzing routinely collected data and generating reports with minimal resources and expertise required. This may empower individual hospitals to contribute to the understanding of and actions on AMR in local settings and maximize the utility of local data.

---

## Acknowledgments

We thank all hospitals that participated in the AMASS project. We thank Dr Thet Wai Nwe from the Central Epidemiology Unit of Myanmar and Dr Htay Htay Tin from the National Health Laboratory of Myanmar for their comments on the AMASS. We also thank Dr Kyi Soe, who is the Senior Medical Superintendent of North Okkalapa General and Teaching Hospital, for his support for the AMASS and for hosting the AMASS training workshop at the hospital. We thank Professor Frank Smithuis and Dr Aung Pyae Phyo from Myanmar Oxford Clinical Research Unit (MOCRU) for their support on the AMASS project. We thank Professor Paul Newton and Professor David Dance from the Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit (LOMRU) for their support on the AMASS project. We thank Ronas Shakya from Patan Hospital, Nepal for technical assistance. We thank Joe Hessel from the Diagnostic Microbiology Development Program for comments on the AMASS. We thank Chotipong Chaisiripreeyakul, Techat Totemchokchaikarn, Sitthinun Pasirichusaree, and Papitchaya Achavakulthep for the tutorial videos on the AMASS. Mahidol Oxford Tropical Medicine Research Unit (MORU) is funded by the Wellcome Trust (grant: 106698/Z/14/Z). Oxford University Clinical Research Unit (OUCRU) is funded by the Wellcome Trust (grant: 106680/B/14/Z). The investigators are funded by the Wellcome Trust (CLim is funded by a Training Research Fellowship [grant: 206736] and DL is funded by an Intermediate Training Fellowship [grant: 101103]). BC is funded by the UK Medical Research Council and Department for International Development (grant: MR/K006924/1). The funders of the investigators and study had no role in the study, data collection, data analysis, data interpretation, or writing of the manuscript. The corresponding authors had the final responsibility for the decision to submit for publication.

---

## Authors' Contributions

CLim and DL conceptualized, designed, and developed the AMASS, and wrote the original draft. TM, VC, MTA, AK, PTe, RB, and LNPH tested the AMASS, curated data, and reviewed reports generated from the AMASS using local data. JS, PTu, EA, RvD, HNL, CLing, SH, SI, SD, TW, VH, WS, YLM, TLV, HHH, MM, MV, BB, JE, SP, GT, ND, and BC reviewed and interpreted the reports. All authors participated in drafting, editing, and revising the final manuscript.

---

## Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

An overview of the flow of data processing and analyses performed by the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS).

[DOCX File, 242 KB - [jmir\\_v22i10e19762\\_app1.docx](#)]

---

### Multimedia Appendix 2

The list of pathogen-antibiotic combinations modified from the World Health Organization Global Antimicrobial Resistance Surveillance System priority list and included in the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS).

[DOCX File, 31 KB - [jmir\\_v22i10e19762\\_app2.docx](#)]

---

### Multimedia Appendix 3



The methods to estimate prevalence and incidence rates were based on the recommendation of the World Health Organization Global Antimicrobial Resistance Surveillance System.

[DOCX File, 29 KB - [jmir\\_v22i10e19762\\_app3.docx](#)]

#### Multimedia Appendix 4

Simulating hypothetical data sets.

[DOCX File, 27 KB - [jmir\\_v22i10e19762\\_app4.docx](#)]

#### Multimedia Appendix 5

How to run the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS) using example data files.

[DOCX File, 26 KB - [jmir\\_v22i10e19762\\_app5.docx](#)]

#### Multimedia Appendix 6

How to configure the data dictionary files.

[DOCX File, 26 KB - [jmir\\_v22i10e19762\\_app6.docx](#)]

#### Multimedia Appendix 7

AMR surveillance report generated from the first example data set.

[PDF File (Adobe PDF File), 44 KB - [jmir\\_v22i10e19762\\_app7.pdf](#)]

#### Multimedia Appendix 8

AMR surveillance report generated from the second example data set.

[PDF File (Adobe PDF File), 124 KB - [jmir\\_v22i10e19762\\_app8.pdf](#)]

#### Multimedia Appendix 9

The list of participating hospitals.

[DOCX File, 30 KB - [jmir\\_v22i10e19762\\_app9.docx](#)]

#### Multimedia Appendix 10

How to export data from WHONET 5.6 (2019) in an Excel format that can be used by the AutoMated tool for Antimicrobial resistance Surveillance System (AMASS).

[DOCX File, 2846 KB - [jmir\\_v22i10e19762\\_app10.docx](#)]

## References

1. Limmathurotsakul D, Dunachie S, Fukuda K, Feasey NA, Okeke IN, Holmes AH, et al. Improving the estimation of the global burden of antimicrobial resistant infections. *The Lancet Infectious Diseases* 2019 Nov;19(11):e392-e398. [doi: [10.1016/s1473-3099\(19\)30276-2](#)]
2. Global Antimicrobial Resistance Surveillance System (GLASS) Report: Early implementation, 2017-2018. World Health Organization. 2018. URL: <https://apps.who.int/iris/bitstream/handle/10665/279656/9789241515061-eng.pdf?ua=1> [accessed 2019-12-01]
3. Annual epidemiological commentary: Gram-negative bacteraemia, MRSA bacteraemia, MSSA bacteraemia and C. difficile infections, up to and including financial year April 2018 to March 2019. Public Health England. 2018. URL: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/815449/Annual\\_epidemiological\\_commentary\\_April\\_2017\\_to\\_March\\_2019.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/815449/Annual_epidemiological_commentary_April_2017_to_March_2019.pdf) [accessed 2019-12-01]
4. Johnson AP, Davies J, Guy R, Abernethy J, Sheridan E, Pearson A, et al. Mandatory surveillance of methicillin-resistant Staphylococcus aureus (MRSA) bacteraemia in England: the first 10 years. *J Antimicrob Chemother* 2012 Apr;67(4):802-809. [doi: [10.1093/jac/dkr561](#)] [Medline: [2223229](#)]
5. Amaratunga K, Tarasuk J, Tsegaye L, Archibald C, 2015 Communicable Infectious Disease Steering Committee (CIDSC), Antimicrobial Resistance (AMR) Surveillance Task Group. Advancing surveillance of antimicrobial resistance: Summary of the 2015 CIDSC Report. *Can Commun Dis Rep* 2016 Nov 03;42(11):232-237 [FREE Full text] [doi: [10.14745/ccdr.v42i11a03](#)] [Medline: [29769992](#)]
6. Tackling Drug-Resistant Infections Globally: Final Report and Recommendation. The Review on Antimicrobial Resistance. 2016. URL: [https://amr-review.org/sites/default/files/160518\\_Final%20paper\\_with%20cover.pdf](https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf) [accessed 2019-12-01]
7. Interagency Coordination Group on Antimicrobial Resistance. No time to wait: securing the future from drug-resistant infections. World Health Organization. 2019. URL: [https://www.who.int/antimicrobial-resistance/interagency-coordination-group/IACG\\_final\\_report\\_EN.pdf?ua=1](https://www.who.int/antimicrobial-resistance/interagency-coordination-group/IACG_final_report_EN.pdf?ua=1) [accessed 2019-12-01]

8. The Global Risks Report 2018, 13th Edition. World Economic Forum. 2018. URL: [http://www3.weforum.org/docs/WEF\\_GRR18\\_Report.pdf](http://www3.weforum.org/docs/WEF_GRR18_Report.pdf) [accessed 2019-12-01]
9. Hall W, McDonnell A, O'Neill J. Superbugs: An Arms Race Against Bacteria. Cambridge, Massachusetts: Harvard University Press; 2018.
10. de Kraker ME, Stewardson AJ, Harbarth S. Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? PLoS Med 2016 Nov 29;13(11):e1002184 [FREE Full text] [doi: [10.1371/journal.pmed.1002184](https://doi.org/10.1371/journal.pmed.1002184)] [Medline: [27898664](https://pubmed.ncbi.nlm.nih.gov/27898664/)]
11. Global Antimicrobial Resistance Surveillance System: Manual for Early Implementation. World Health Organization. 2015. URL: [https://apps.who.int/iris/bitstream/handle/10665/188783/9789241549400\\_eng.pdf?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/188783/9789241549400_eng.pdf?sequence=1) [accessed 2019-12-01]
12. CLSI. Analysis and Presentation of Cumulative Antimicrobial Susceptibility Test Data, Fourth Edition. In: CLSI document M39-A4. Wayne, Pennsylvania: Clinical and Laboratory Standards Institute; 2014.
13. Public Health England. Biannual commentary on MRSA, MSSA and Gram-negative bacteraemia and Clostridium difficile infection from Independent Sector healthcare organisations in England (April 2017 to September 2017). London: Public Health England; Apr 2018.
14. Antimicrobial resistance (AMR) reporting protocol 2018. European Centre for Disease Prevention and Control. URL: <https://www.ecdc.europa.eu/sites/portal/files/documents/EARS-Net%20reporting%20protocol%202018.%20docx.pdf> [accessed 2019-12-01]
15. Antibiotic resistance threats in the United States 2019. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf> [accessed 2019-12-01]
16. Antimicrobial Resistance: Global Report on Surveillance. World Health Organization. 2014. URL: [https://apps.who.int/iris/bitstream/handle/10665/112642/9789241564748\\_eng.pdf;jsessionid=18CA836B4A59D8644F0976415FC00F67?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/112642/9789241564748_eng.pdf;jsessionid=18CA836B4A59D8644F0976415FC00F67?sequence=1) [accessed 2019-12-01]
17. Ashley E, Shetty N, Patel J, van Doorn R, Limmathurotsakul D, Feasey NA, et al. Harnessing alternative sources of antimicrobial resistance data to support surveillance in low-resource settings. J Antimicrob Chemother 2019 Mar 01;74(3):541-546 [FREE Full text] [doi: [10.1093/jac/dky487](https://doi.org/10.1093/jac/dky487)] [Medline: [30544186](https://pubmed.ncbi.nlm.nih.gov/30544186/)]
18. Kyobe S, Musinguzi H, Lwanga N, Kezimbira D, Kigozi E, Katabazi FA, H3Africa Biorepository PI Working G. Selecting a Laboratory Information Management System for Biorepositories in Low- and Middle-Income Countries: The H3Africa Experience and Lessons Learned. Biopreservation and Biobanking 2017 Apr;15(2):111-115. [doi: [10.1089/bio.2017.0006](https://doi.org/10.1089/bio.2017.0006)]
19. Stelling J, O'Brien TF. Surveillance of antimicrobial resistance: the WHONET program. Clin Infect Dis 1997 Jan;24 Suppl 1:S157-S168. [doi: [10.1093/clinids/24.supplement\\_1.s157](https://doi.org/10.1093/clinids/24.supplement_1.s157)] [Medline: [8994799](https://pubmed.ncbi.nlm.nih.gov/8994799/)]
20. Lim C, Takahashi E, Hongsuwan M, Wuthiekanun V, Thamlikitkul V, Hinjoy S, et al. Epidemiology and burden of multidrug-resistant bacterial infection in a developing country. Elife 2016 Sep 06;5:e18082 [FREE Full text] [doi: [10.7554/eLife.18082](https://doi.org/10.7554/eLife.18082)] [Medline: [27599374](https://pubmed.ncbi.nlm.nih.gov/27599374/)]
21. Gupta K, Scholes D, Stamm WE. Increasing prevalence of antimicrobial resistance among uropathogens causing acute uncomplicated cystitis in women. JAMA 1999 Feb 24;281(8):736-738. [doi: [10.1001/jama.281.8.736](https://doi.org/10.1001/jama.281.8.736)] [Medline: [10052444](https://pubmed.ncbi.nlm.nih.gov/10052444/)]
22. Apte M, Neidell M, Furuya E, Caplan D, Glied S, Larson E. Using electronically available inpatient hospital data for research. Clin Transl Sci 2011 Oct;4(5):338-345 [FREE Full text] [doi: [10.1111/j.1752-8062.2011.00353.x](https://doi.org/10.1111/j.1752-8062.2011.00353.x)] [Medline: [22029805](https://pubmed.ncbi.nlm.nih.gov/22029805/)]
23. Wyllie DH, Crook DW, Peto TE. Mortality after Staphylococcus aureus bacteraemia in two hospitals in Oxfordshire, 1997-2003: cohort study. BMJ 2006 Jun 23;333(7562):281. [doi: [10.1136/bmj.38834.421713.2f](https://doi.org/10.1136/bmj.38834.421713.2f)]
24. AMASS. URL: <https://www.amass.website/> [accessed 2020-09-07]
25. Figshare. URL: [https://figshare.com/articles/AMASS\\_zip/7763330](https://figshare.com/articles/AMASS_zip/7763330) [accessed 2020-09-07]
26. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. World Health Organization. 2017. URL: [https://www.who.int/medicines/publications/WHO-PPL-Short\\_Summary\\_25Feb-ET\\_NM\\_WHO.pdf?ua=1](https://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf?ua=1) [accessed 2019-12-01]
27. Deaths involving MRSA: England and Wales. Office for National Statistics. 2014. URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsinvolvingmrsaenglandandwales> [accessed 2019-12-01]
28. Deaths involving Clostridium difficile. Office for National Statistics. 2017. URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsinvolvingclostridiumdifficilereferencetables> [accessed 2019-12-01]
29. Angkor Hospital for Children, Siem Reap. Antimicrobial resistance surveillance report, Angkor Hospital for Children, Cambodia, 01 Jan 2016 to 31 Dec 2016. figshare. 2020. URL: <https://doi.org/10.6084/m9.figshare.12000225.v1> [accessed 2020-04-18]
30. Mahosot Hospital, Vientiane. Antimicrobial resistance surveillance report, Mahosot Hospital, Lao PDR, 01 Jul 2019 to 30 Sep 2019. figshare. 2020. URL: <https://doi.org/10.6084/m9.figshare.12000222.v1> [accessed 2020-04-18]
31. North Okkalapa General and Teaching Hospital, Yangon. Antimicrobial resistance surveillance report, North Okkalapa General and Teaching Hospital, Myanmar, 02 Jan 2019 to 27 Dec 2019. figshare. 2020. URL: <https://doi.org/10.6084/m9.figshare.12000237.v1> [accessed 2020-04-18]

32. Patan Hospital, Kathmandu. Antimicrobial resistance surveillance report, Patan Hospital, Nepal, 01 Jan 2018 to 30 Dec 2018. figshare. 2020. URL: <https://doi.org/10.6084/m9.figshare.12000231.v1> [accessed 2020-04-18]
33. Sunpasitthiprasong Hospital, Ubon Ratchathani. Antimicrobial resistance surveillance report, Sunpasitthiprasong Hospital, Thailand, 01 Jan 2015 to 31 Dec 2015. figshare. 2020. URL: <https://doi.org/10.6084/m9.figshare.12000240.v1> [accessed 2020-04-18]
34. St Thomas' Hospital, London. Antimicrobial resistance surveillance report, St Thomas' Hospital, United Kingdom, 01 Jan 2018 to 31 Dec 2018. figshare. 2020. URL: <https://doi.org/10.6084/m9.figshare.12000249.v1> [accessed 2020-04-18]
35. The Hospital for Tropical Diseases, Ho Chi Minh City. Antimicrobial resistance surveillance report, The Hospital for Tropical Diseases, Vietnam, 01 Jan 2017 to 30 Dec 2017. figshare. 2020. URL: <https://doi.org/10.6084/m9.figshare.12000252.v1> [accessed 2020-04-18]
36. Turner P, Ashley EA, Celhay OJ, Douangnouvong A, Hamers RL, Ling CL, et al. ACORN (A Clinically-Oriented Antimicrobial Resistance Surveillance Network): a pilot protocol for case based antimicrobial resistance surveillance. Wellcome Open Res 2020 Jan 27;5:13. [doi: [10.12688/wellcomeopenres.15681.1](https://doi.org/10.12688/wellcomeopenres.15681.1)]
37. Chang K, Rattanavong S, Mayxay M, Keoluangkhot V, Davong V, Vongsouvath M, et al. Bacteremia Caused by Extended-Spectrum Beta-Lactamase-Producing Enterobacteriaceae in Vientiane, Lao PDR: A 5-Year Study. Am J Trop Med Hyg 2020 May;102(5):1137-1143 [FREE Full text] [doi: [10.4269/ajtmh.19-0304](https://doi.org/10.4269/ajtmh.19-0304)] [Medline: [32157990](https://pubmed.ncbi.nlm.nih.gov/32157990/)]
38. Sriboonsong S, Boonchoo L. Automatic antimicrobial susceptibility testing system. In: Magjarevic R, Nagel JH, editors. World Congress on Medical Physics and Biomedical Engineering 2006. IFMBE Proceedings. Berlin, Heidelberg: Springer; 2007:2501-2504.
39. ICNET. URL: <https://www.icnetsoftware.com> [accessed 2020-09-07]
40. O'Brien TF, Clark A, Peters R, Stelling J. Why surveillance of antimicrobial resistance needs to be automated and comprehensive. J Glob Antimicrob Resist 2019 Jun;17:8-15. [doi: [10.1016/j.jgar.2018.10.011](https://doi.org/10.1016/j.jgar.2018.10.011)] [Medline: [30326273](https://pubmed.ncbi.nlm.nih.gov/30326273/)]
41. Turner P, Ashley EA. Standardising the reporting of microbiology and antimicrobial susceptibility data. The Lancet Infectious Diseases 2019 Nov;19(11):1163-1164. [doi: [10.1016/s1473-3099\(19\)30561-4](https://doi.org/10.1016/s1473-3099(19)30561-4)]
42. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health 2014 Nov 05;14(1):1144 [FREE Full text] [doi: [10.1186/1471-2458-14-1144](https://doi.org/10.1186/1471-2458-14-1144)] [Medline: [25377061](https://pubmed.ncbi.nlm.nih.gov/25377061/)]
43. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. PLoS One 2011 Jun 29;6(6):e21101 [FREE Full text] [doi: [10.1371/journal.pone.0021101](https://doi.org/10.1371/journal.pone.0021101)] [Medline: [21738610](https://pubmed.ncbi.nlm.nih.gov/21738610/)]

## Abbreviations

**AMASS:** AutoMated tool for Antimicrobial resistance Surveillance System

**AMR:** antimicrobial resistance

**AST:** antimicrobial susceptibility testing

**CLSI:** Clinical and Laboratory Standard Institute

**EUCAST:** European Committee on Antimicrobial Susceptibility Testing

**GLASS:** Global Antimicrobial Resistance Surveillance System

**LIS:** laboratory information system

**LMIC:** low and middle-income country

**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 03.05.20; peer-reviewed by O Las Vergnas, T Aslanidis; comments to author 10.07.20; revised version received 22.07.20; accepted 26.07.20; published 02.10.20.*

### *Please cite as:*

Lim C, Miliya T, Chansamouth V, Aung MT, Karkey A, Teparrukkul P, Rahul B, Lan NPH, Stelling J, Turner P, Ashley E, van Doorn HR, Lin HN, Ling C, Hinjoy S, Iamsirithaworn S, Dunachie S, Wangrangsimakul T, Hantrakun V, Schilling W, Yen LM, Tan LV, Hlaing HH, Mayxay M, Vongsouvath M, Basnyat B, Edgeworth J, Peacock SJ, Thwaites G, Day NPJ, Cooper BS, Limmathurotsakul D

*Automating the Generation of Antimicrobial Resistance Surveillance Reports: Proof-of-Concept Study Involving Seven Hospitals in Seven Countries*

*J Med Internet Res* 2020;22(10):e19762

URL: <https://www.jmir.org/2020/10/e19762>

doi: [10.2196/19762](https://doi.org/10.2196/19762)

PMID: [33006570](https://pubmed.ncbi.nlm.nih.gov/33006570/)

©Cherry Lim, Thyl Miliya, Vilada Chansamouth, Myint Thazin Aung, Abhilasha Karkey, Prapit Teparrukkul, Batra Rahul, Nguyen Phu Huong Lan, John Stelling, Paul Turner, Elizabeth Ashley, H Rogier van Doorn, Htet Naing Lin, Clare Ling, Soawapak Hinjoy, Sopon Iamsirithaworn, Susanna Dunachie, Tri Wangrangsimakul, Viriya Hantrakun, William Schilling, Lam Minh Yen, Le Van Tan, Htay Htay Hlaing, Mayfong Mayxay, Manivanh Vongsouvath, Buddha Basnyat, Jonathan Edgeworth, Sharon J Peacock, Guy Thwaites, Nicholas PJ Day, Ben S Cooper, Direk Limmathurotsakul. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 02.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Efficiency of Computer-Aided Facial Phenotyping (DeepGestalt) in Individuals With and Without a Genetic Syndrome: Diagnostic Accuracy Study

Jean Tori Pantel<sup>1,2\*</sup>; Nurulhuda Hajjir<sup>1,3\*</sup>; Magdalena Danyel<sup>1,4</sup>; Jonas Elsner<sup>1</sup>; Angela Teresa Abad-Perez<sup>1</sup>; Peter Hansen<sup>1,5</sup>, PhD; Stefan Mundlos<sup>1,6</sup>, Prof Dr; Malte Spielmann<sup>6,7</sup>, Prof Dr; Denise Horn<sup>1</sup>, Prof Dr; Claus-Eric Ott<sup>1</sup>, MD; Martin Atta Mensah<sup>1,8</sup>, MD

<sup>1</sup>Institute of Medical Genetics and Human Genetics, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health, Berlin, Germany

<sup>2</sup>Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

<sup>3</sup>Klinik für Pädiatrie mit Schwerpunkt Gastroenterologie, Nephrologie und Stoffwechselmedizin, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health, Berlin, Germany

<sup>4</sup>Berlin Center for Rare Diseases, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health, Berlin, Germany

<sup>5</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, United States

<sup>6</sup>RG Development & Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany

<sup>7</sup>Institute of Human Genetics, University of Lübeck, Lübeck, Germany

<sup>8</sup>Berlin Institute of Health, Berlin, Germany

\*these authors contributed equally

**Corresponding Author:**

Martin Atta Mensah, MD

Institute of Medical Genetics and Human Genetics

Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health

Berlin

Germany

Phone: 49 30 450 569 132

Fax: 49 30 450 569 914

Email: [martin-atta.mensah@charite.de](mailto:martin-atta.mensah@charite.de)

## Abstract

**Background:** Collectively, an estimated 5% of the population have a genetic disease. Many of them feature characteristics that can be detected by facial phenotyping. Face2Gene CLINIC is an online app for facial phenotyping of patients with genetic syndromes. DeepGestalt, the neural network driving Face2Gene, automatically prioritizes syndrome suggestions based on ordinary patient photographs, potentially improving the diagnostic process. Hitherto, studies on DeepGestalt's quality highlighted its sensitivity in syndromic patients. However, determining the accuracy of a diagnostic methodology also requires testing of negative controls.

**Objective:** The aim of this study was to evaluate DeepGestalt's accuracy with photos of individuals with and without a genetic syndrome. Moreover, we aimed to propose a machine learning-based framework for the automated differentiation of DeepGestalt's output on such images.

**Methods:** Frontal facial images of individuals with a diagnosis of a genetic syndrome (established clinically or molecularly) from a convenience sample were reanalyzed. Each photo was matched by age, sex, and ethnicity to a picture featuring an individual without a genetic syndrome. Absence of a facial gestalt suggestive of a genetic syndrome was determined by physicians working in medical genetics. Photos were selected from online reports or were taken by us for the purpose of this study. Facial phenotype was analyzed by DeepGestalt version 19.1.7, accessed via Face2Gene CLINIC. Furthermore, we designed linear support vector machines (SVMs) using Python 3.7 to automatically differentiate between the 2 classes of photographs based on DeepGestalt's result lists.



**Results:** We included photos of 323 patients diagnosed with 17 different genetic syndromes and matched those with an equal number of facial images without a genetic syndrome, analyzing a total of 646 pictures. We confirm DeepGestalt's high sensitivity (top 10 sensitivity: 295/323, 91%). DeepGestalt's syndrome suggestions in individuals without a craniofacially dysmorphic syndrome followed a nonrandom distribution. A total of 17 syndromes appeared in the top 30 suggestions of more than 50% of nondysmorphic images. DeepGestalt's top scores differed between the syndromic and control images (area under the receiver operating characteristic [AUROC] curve 0.72, 95% CI 0.68-0.76;  $P < .001$ ). A linear SVM running on DeepGestalt's result vectors showed stronger differences (AUROC 0.89, 95% CI 0.87-0.92;  $P < .001$ ).

**Conclusions:** DeepGestalt fairly separates images of individuals with and without a genetic syndrome. This separation can be significantly improved by SVMs running on top of DeepGestalt, thus supporting the diagnostic process of patients with a genetic syndrome. Our findings facilitate the critical interpretation of DeepGestalt's results and may help enhance it and similar computer-aided facial phenotyping tools.

(*J Med Internet Res* 2020;22(10):e19263) doi:[10.2196/19263](https://doi.org/10.2196/19263)

## KEYWORDS

facial phenotyping; DeepGestalt; facial recognition; Face2Gene; medical genetics; diagnostic accuracy; genetic syndrome; machine learning

## Introduction

### Background

Although individual genetic diseases are rare, they collectively affect an estimated 5% of a population [1]. Thus, these diseases represent a major challenge for health care systems, as it usually requires highly specialized knowledge to propose a specific genetic diagnosis. Assessing the facial phenotypes of patients with genetic syndromes is key to this diagnostic process [2]. Traditionally performed by a physician, the advents of computer vision and machine learning in medicine enable rapid and automated assessment of a patient's facial traits [3,4]. Numerous facial phenotyping systems have been developed with the potential to aid the diagnostic processes in medical genetics [5-12]. DeepGestalt, the neural network behind Face2Gene CLINIC, which was trained on more than 17,106 images, is thus far the best-investigated and most convenient to use application [11]. Several studies assessed the algorithm's sensitivity, suggesting that it is of a certain quality [11,13-38]. These tests predominantly analyzed images of patients diagnosed with a genetic disorder known to show characteristic facial features. This appears reasonable as DeepGestalt is designed to identify such syndromes. However, it might introduce a bias in conclusions of the system's everyday clinical use since not all individuals seen in a real-life setting belong to the group of patients included in previous studies of DeepGestalt. This may be because (1) the featured syndrome is yet to be analyzed by the system; (2) an individual features a syndrome not associated with a characteristic facies; or (3) an individual has no syndrome at all.

In addition to such evaluations of DeepGestalt's sensitivity, there is a need for studies on its specificity when tested on individuals without craniofacial dysmorphism. As DeepGestalt is not designed to suggest the class label "inconspicuous face" [11], evaluating its clinical specificity is not too trivial a task. Some studies tested the ability of DeepGestalt's methodology to distinguish between facial images with and without a genetic syndrome by constructing user-specific neural networks trained on healthy control images and on images of limited numbers of well-selected genetic disorders using Face2Gene RESEARCH

[20,26-28,30,32,34,39-41]. Their results suggested that neural networks such as DeepGestalt may have the potential to differentiate between the 2 classes and may thus be used in diagnosing patients in medical genetics. Such a test could be applied at different stages of the diagnostic process. Patients who want to know if genetic counseling is necessary could use it as a triage test to check whether a suspicion of a genetic disease is justified. Physicians and other medical professionals could similarly use such a test on patients suspected of having a genetic syndrome to narrow down the range of possible diagnoses. Geneticists could use it as an add-on test to further confirm a diagnosis, for example, in the presence of a variant of unknown significance.

### Objectives

We aimed to systematically benchmark DeepGestalt's power to discern images of individuals with a dysmorphic genetic syndrome from images of healthy control individuals. For this purpose, we tested the basic prerequisite for the diagnostic usefulness of DeepGestalt, that is, to yield different scores in persons with a conventionally established diagnosis of a genetic syndrome than in persons without a genetic syndrome ( $H_1: \mu_{\text{syndromic}} \neq \mu_{\text{healthy}}$ ). We also determined DeepGestalt's capacity to distinguish those images by measuring its area under the receiver operating characteristic (AUROC) curve. Furthermore, we aimed to develop and test a machine learning-based approach to improve DeepGestalt's accuracy.

## Methods

### Selection and Analysis of Portrait Photos

#### Study Design

To be included in this study, portrait photos had to depict the entire frontal face (from hairline to chin showing both eyes) and no artifact other than glasses. To achieve a vertical positioning of the face, the images were cropped and rotated if necessary. A convenience sample of online accessible images was collected between September 2019 and December 2019, using a methodology adjusted from Ferry et al [8]. Pictures photographed by us were taken at the 2018 meeting of the

Elterninitiative Apertsyndrom und Verwandte Fehlbildungen eV, a parents' initiative on Apert syndrome and related disorders in Germany, after obtaining written informed consents as approved by the ethics committee of the Charité – Universitätsmedizin Berlin (EA2/190/16). Image inclusion was planned before conducting analysis by DeepGestalt. A sample size of the positive and negative class of 105 (N=210) was calculated using G\*Power, version 3.1.9.7 (effect size 0.5;  $\alpha=.05$ ; power 0.95; allocation ratio 1).

### Defining Reference Phenotypes

Only images of individuals reported to be clinically or molecularly diagnosed with a genetic syndrome were labeled as syndromic. When no syndrome was reported and no facial gestalt suggestive of a syndrome was observed, as judged by physicians working in medical genetics, images were labeled as “healthy.”

### Computer-Aided Facial Phenotyping

Computer-aided facial phenotyping was performed using DeepGestalt version 19.1.7, accessed via Face2Gene CLINIC (FDNA Inc). Neither the class labels nor diagnoses were passed to DeepGestalt. No other phenotypic information but 1 portrait photo per case was entered into the system. DeepGestalt's training set was tested not to contain duplicates of images used in this study, as described previously [42].

### Danyel Cohort

The Danyel cohort, originally described by Danyel et al [30], comprises 116 healthy control images.

### Syndromic Cohort

This cohort comprises frontal facial images of 17 syndromes. We planned to collect the same number of images for each of these syndromes. A total of 16 of these syndromes were chosen from the 201 distinct suggestions in DeepGestalt's top 30 results lists of the Danyel cohort. Syndromes of different frequencies ranging from 76% (frequently suggested) to 1% (rarely suggested) were selected. In descending order of frequency, these syndromes are as follows: Fragile X syndrome (OMIM: #300624), Angelman syndrome (OMIM: #105830), Rett syndrome (OMIM: #312750), Phelan-McDermid syndrome (OMIM: #606232), Klinefelter syndrome, Beckwith-Wiedemann syndrome (OMIM: #130650), 22q11.2 deletion syndrome (OMIM: #611867), Sotos syndrome (OMIM: #117550), Noonan syndrome (OMIM: PS163950), Loeys-Dietz syndrome (OMIM: PS609192), Williams-Beuren syndrome (OMIM: #194050), Rubinstein-Taybi syndrome (OMIM: PS180849), achondroplasia (OMIM: #100800), Wolf-Hirschhorn syndrome (OMIM: #194190), Pallister-Killian syndrome (OMIM: #601803), and Treacher Collins syndrome (OMIM: PS154500). In addition, we chose Apert syndrome (OMIM: #101200), which was not implied in the Danyel cohort.

### Matched Control Cohort

Each photo of the syndromic cohort was matched to an image of an individual without a genetic syndrome by age, sex, and ethnicity to build a cohort of an equal number of control images.

## Statistical Evaluation and Classification Experiments

Face2Gene CLINIC returns DeepGestalt's top 30 syndrome suggestions. DeepGestalt associates each suggestion with a Gestalt score [11]. The syndrome suggestions' frequencies, scores, and ranks were statistically evaluated.

### Feature Extraction and Vector Construction

All images were labeled by class (syndromic vs healthy). Vectors were built to hold an attribute for any of the syndromes suggested at least once in DeepGestalt's top 30 suggestions. To construct a vector for a given photo, the 30 highest Gestalt scores were assigned to their respective attributes; and the remaining attributes were set to 0 (s. matrix.txt in [Multimedia Appendix 1](#)).

### Classification

To differentiate between syndromic and healthy portrait photos, we trained linear support vector machines (SVMs) using the LinearSVM class of scikit-learn, version 0.21.3, with default parameters in Python 3.7. To avoid overfitting, training and testing were performed using a leave-1-out classification scheme. Since ethnic background is a possible confounder of DeepGestalt [15,22,26,29,33], we designed classification experiments based on all images, images of White persons, and those of persons with other ethnicities, to benchmark the influence of ethnicity on SVM performance.

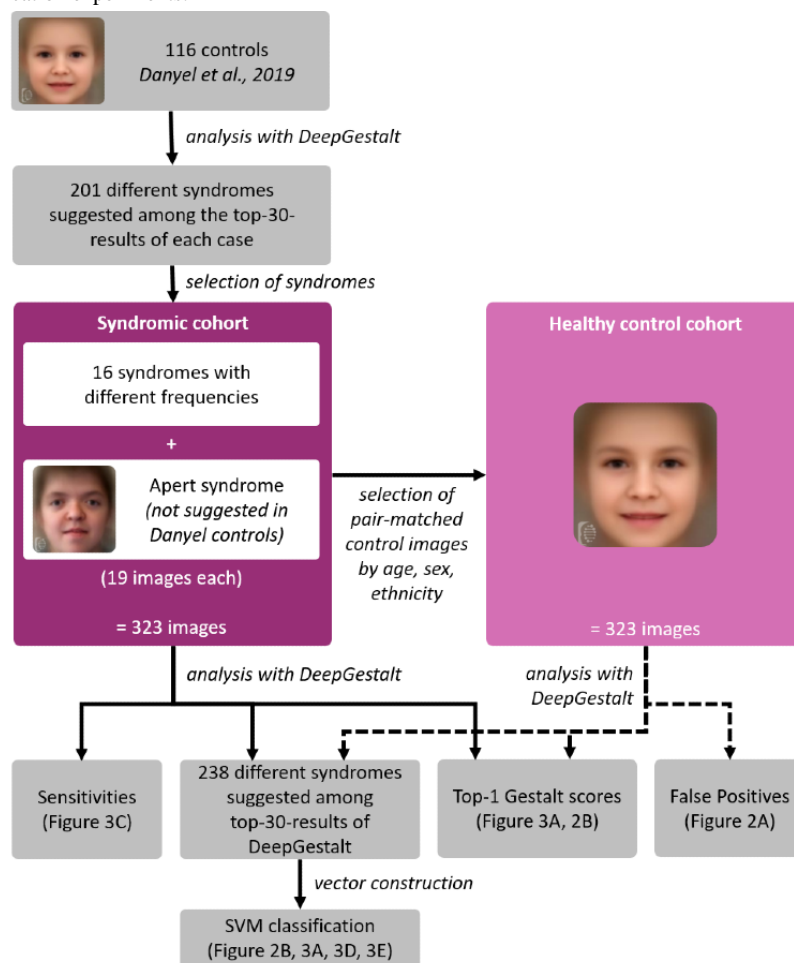
To test a possible influence of the number of top ranks considered, classification of all images was run 30 times with the number of considered top Gestalt ranks, ranging from 1 to 30.

### Statistical Analysis

Scores of the syndromic and healthy control cohort were tested to be different using a 2-sided, independent Welch *t* test. Difference of receiver operating characteristics (ROCs) was tested using a DeLong test. Classification performance was assessed using Matthews correlation coefficient (MCC). All statistical tests were performed in Python 3.7; the code can be found in [Multimedia Appendix 1](#).

### Data and Code Availability

The data and code can be found in [Multimedia Appendix 1](#). For reasons of data protection, all data were cumulated (where possible), deidentified, and minimized. Facial images depicted in [Figure 1](#) show computer-generated composite masks and not real individuals. In [Multimedia Appendix 1](#), file data.txt describes the diagnosis, age, sex, and ethnicity of persons in the analyzed set of images; and file matrix.txt contains DeepGestalt's output vectors as used for this study. Files differentiator.py and reproduce.py may be used for reproducing the statistical results of this study. Further information may be found in file readme.txt ([Multimedia Appendix 1](#)).

**Figure 1.** Workflow of classification experiments.

## Results

### Included Images

We could include 19 images for each of the 17 syndromes in the syndromic cohort. A total of 83% (272/323) of these images were of White persons (file data.txt of [Multimedia Appendix 1](#)). Images from the syndromic cohort were matched to 323 images forming the matched control cohort, resulting in a total number of 646 analyzed photos ([Figure 1](#)).

### Frequencies and Scores of Suggested Syndromes in Control Individuals

DeepGestalt suggested 238 different syndromes among the top 30 suggestions of the matched control cohort. One syndrome was suggested in more than 80% of the cases (Fragile X syndrome, 82%), 6 syndromes in 70%-80% of the cases; 4 syndromes in 60%-70% of the cases; 6 syndromes in 50%-60% of the cases; 6 syndromes in 40%-50% of the cases; 11 syndromes in 30%-40% of the cases; 15 syndromes in 20%-30% of the cases; 29 syndromes in 10%-20% of the cases; and 160 syndromes at least once in less than 10% of the cases ([Figure 2A](#)).

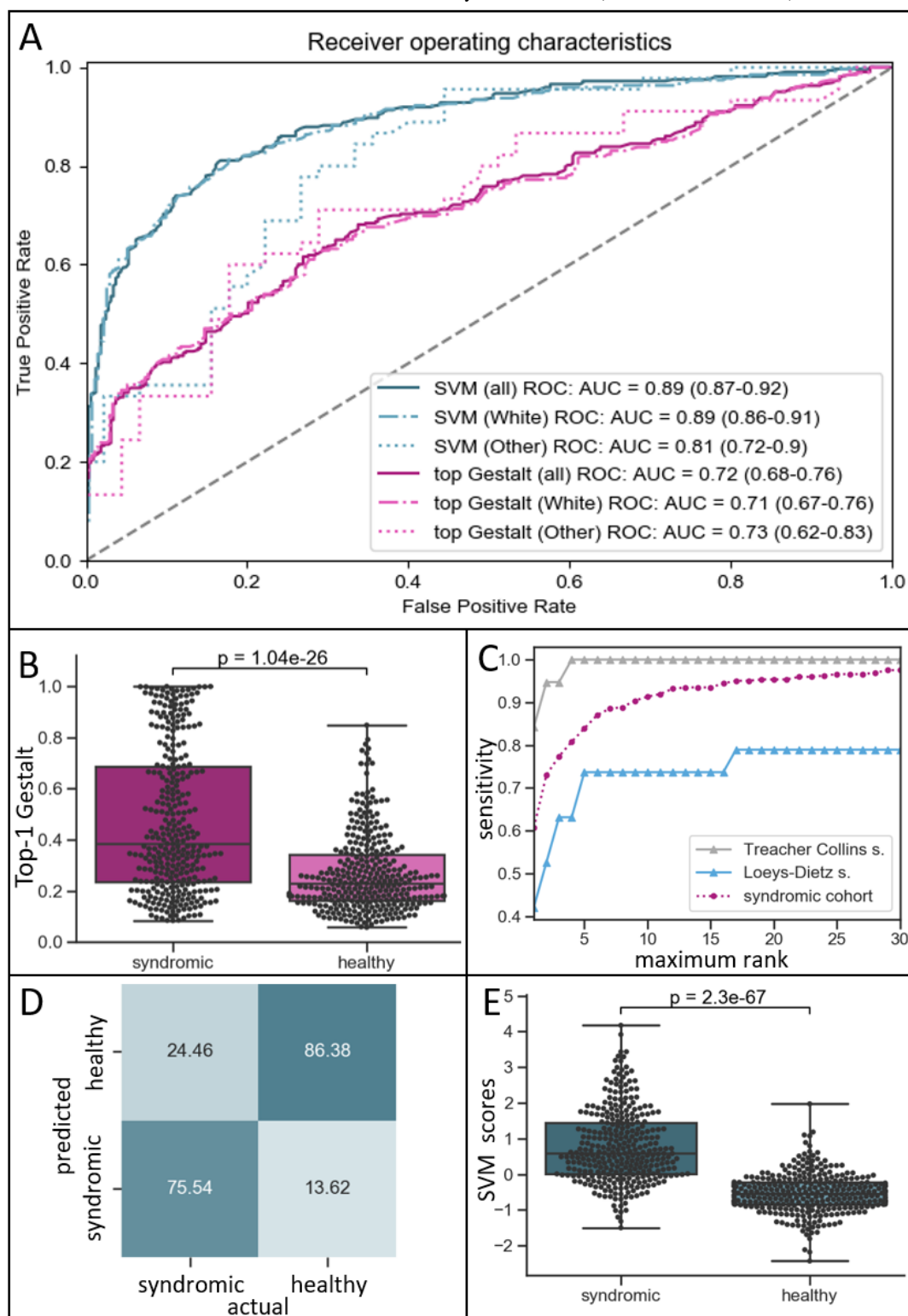
**Figure 2.** (A) Frequency of syndromes suggested by DeepGestalt in more than 20% of the matched control cohort's top-30-results lists. Colors indicate frequency percentages. (B) Number of images correctly classified as "syndromic"; colors relate to (A) and gray indicates <20%.



The highest first-rank Gestalt score of the matched control cohort amounted to 0.85, and the lowest, to 0.06, with a mean of 0.27 (SD 0.15). First-rank Gestalt scores of the syndromic cohort (highest 1.0; lowest 0.08; mean 0.47, SD 0.28) and the matched control cohort appeared to be separable with an AUROC of 0.72 (95% CI 0.68-0.76) (Figure 3A). Notably, this

was found for both tested ethnic groups (Figure 3A, Multimedia Appendix 2), White persons only (AUROC 0.71, 95% CI 0.67-0.76;  $P < .001$ ), and persons of other ethnicities only (AUROC 0.71, 95% CI 0.62-0.83;  $P < .001$ ). Separability of the 2 cohorts is evident and significant ( $P < .001$ ), as shown in Figure 3B.

**Figure 3.** (A) Receiver operating characteristic (ROC) curves: dashed line indicates random ROC curve; note that support vector machine (SVM) scores yield higher areas under the ROC curves (AUROCs) than their respective raw first-rank Gestalt scores. (B) Distribution of first-rank Gestalt scores in the syndromic cohort and the matched control cohort (healthy). (C) Sensitivities of DeepGestalt (X-axis: number of considered top ranks). Dark-purple circles: average of syndromic cohort; gray triangles: 19 images with Treacher-Collins syndrome; blue triangles: 19 images with Loeys-Dietz syndrome. (D) Distribution of SVM scores in the syndromic cohort and the matched control cohort; note: improved separability as compared to B. (E) SVM classification results based on the entire matched control cohort and syndromic cohort (threshold SVM score: 0).



### Sensitivity of DeepGestalt

DeepGestalt's average top 10 sensitivity in the syndromic cohort amounted to 91%, varying between the 17 tested syndromes (Figure 3C, Multimedia Appendix 3). Interestingly, DeepGestalt

was sensitive independent of ethnicity (White persons only, 90%; persons of other ethnicities only, 97%). A total of 7 syndromes reached a top 10 sensitivity of 100% (Fragile X, Noonan, Phelan-McDermid, Rett, Sotos, Treacher-Collins, and Williams-Beuren syndromes). DeepGestalt performed worst



for Loeys-Dietz syndrome, with a top 10 sensitivity of 74% (Figure 3C).

### Performance of the SVM

Sensitivities of binary SVM classification differed between syndromes (Figure 2B). All images of individuals with Apert syndrome, Wolf-Hirschhorn syndrome, and Williams-Beuren syndrome were correctly classified as being syndromic. The SVM performed worst on the 19 images of individuals with Klinefelter syndrome, correctly classifying only 7 of them as syndromic.

Binary SVM classification of DeepGestalt's output achieved an increased separability of syndromic images and healthy controls as compared to top Gestalt scores with an AUROC of 0.89 (95% CI 0.87-0.92) (Figure 3A). Again, this was true in both tested ethnic groups (Figure 3A), for photos of White persons (AUROC 0.88, 95% CI 0.86-0.91;  $P < .001$ ) and those of persons of other ethnicities (AUROC 0.79, 95% CI 0.62-0.83). However, difference in ROCs was not significant in the latter ( $P = .13$ ). SVM classification performance improved with an increasing number of considered ranks. Using the top 30 Gestalt scores showed the best MCC (0.63), as shown in Multimedia Appendix 4, with a sensitivity of 75.54% and a specificity of 86.38% (Figure 3D). Separability was significant ( $P < .001$ ) (Figure 3E).

## Discussion

### Classification of Images of Individuals Without a Genetic Syndrome

To our knowledge, this is the first study to systematically analyze DeepGestalt's behavior on portrait photos of individuals without a genetic syndrome. For these images, we show that DeepGestalt's syndrome suggestions follow an interesting distribution. Certain syndromes are implied as differential diagnoses with a considerably high likelihood. Among these were Fragile X, Klinefelter, Rett, and Angelman syndromes, which were suggested in more than 3 quarters of the matched control cohort. In contrast, syndromes such as Treacher-Collins syndrome and Wolf-Hirschhorn syndrome were implied very rarely.

DeepGestalt cannot assign the class label "inconspicuous." Yet, DeepGestalt's scores are used to help judge the presence of a given syndrome. Based on a high maximum Gestalt score, a user could assume that the individual depicted in an entered image is likely to have a syndrome. Likewise, one is tempted to assume that a low maximum Gestalt score makes an underlying syndrome unlikely. Indeed, the mean of first-rank Gestalt scores is higher in images depicting syndromic facies than in images of individuals without a genetic syndrome. Similarly, scores higher than 0.85 appear to be specific indicators of a syndromic facies, and those lower than 0.08 are not suggestive of a genetic syndrome. However, these specific values are very rare. Gestalt scores alone are only fairly sufficient for judging the presence or absence of a genetic syndrome with facial dysmorphism since the distributions of the highest Gestalt scores of the syndromic and matched control cohort greatly overlap. We show that this problem can be

reduced by considering both top Gestalt scores and the actual list of suggested syndrome matches. The boost in discriminatory power is illustrated by the increase of the respective AUROCs. Although DeepGestalt cannot directly assess the presence/absence of a syndromic facies, machine learning-based tools (eg, SVMs) built on top of DeepGestalt may be used for this purpose.

It is noteworthy that we achieved promising results with a comparably low number of samples and a low complexity classification model with default hyperparameters. We assume that the quality and complexity of future classifiers will improve as more data will become available. Increasing the number of top ranks considered for vector construction increased the performance of the SVM. However, the number of DeepGestalt's suggestions accessible via Face2Gene CLINIC is limited to 30 suggestions. We hypothesize that using more than just the 30 top ranks for vector construction might further boost classification performance. We classified DeepGestalt's output to predict the presence of a syndromic facies. We also suggest evaluating classification performance based on DeepGestalt's input vectors.

### Potential Confounders

Until now, differences in the diagnostic performance of DeepGestalt, which arise due to the ethnicity of the person depicted, have been evaluated using DeepGestalt's sensitivity. Studies of earlier versions of DeepGestalt showed that its sensitivity is dependent on the ethnic background in certain syndromes [15,22]. Studies of more recent versions of DeepGestalt suggested that ethnicity had no major influence on its sensitivity [26,29]. In our set of syndromic images, DeepGestalt's sensitivity is remarkably high, which is in line with the previous studies highlighting DeepGestalt's good general sensitivity [11,36,42]. This high sensitivity of DeepGestalt was confirmed for both groups of images, those of White persons and those of persons of other ethnicities. Improvement of distinguishability of images of individuals with and without a genetic syndrome appeared to be stronger in the group of photos of White persons than in the group of photos of persons of other ethnicities. However, we assume that this is caused by the limited sample size of images of non-White persons in our data set. We believe that our approach is also applicable to populations comprising predominantly other ethnicities.

The SVM had difficulties classifying images of patients with syndromes that were frequently suggested in healthy controls. Possible explanations for DeepGestalt's output to be similar in controls and individuals with these syndromes could be as follows: (1) such syndromes have only mild characteristic facial features; (2) they have a typical facial gestalt, which is present only in some but not all affected individuals; or (3) they have no typical facies at all. For example, not all patients with Loeys-Dietz syndrome exhibit distinctive facial features [43], and the facial appearances of males with Klinefelter syndrome show no commonly observed characteristics [44].

## Further Research

Further research is necessary to determine DeepGestalt's capacity to distinguish individuals with and without a genetic syndrome when combined with other sources of information, such as genetic test results and nonfacial phenotypic information. We suggest including additional scores that are based on both phenotype and genotype (eg, prioritization of exome data by image analysis [PEDIA] scores [42]) in future classifiers of the presence/absence of a syndromic facies.

The increasing use and quality of facial phenotyping software in clinical genetics should also be accompanied by an ethical evaluation of these systems [45]. This affects issues such as the automation of medical diagnostic action, the sharing of (potentially identifiable) data, and a potentially altered doctor-patient relationship. In particular, a systematic analysis

of the patient perspective on the use of computer-aided facial analysis methodologies in clinical genetics is lacking so far.

We believe that our findings will help improve future versions of DeepGestalt and similar systems and are crucial when interpreting Face2Gene's results in the clinical routine. In particular, we recommend providing users with the false-positive rates of each suggested syndrome.

## Conclusion

DeepGestalt is a computer-aided facial phenotyping tool that showed promising results for detecting a potentially syndromic facies. It yields higher first-rank scores in individuals with a genetic syndrome than in those without a diagnosis of a genetic syndrome. Its output may be classified to improve this detection. The exact stage to use DeepGestalt during the diagnostic makeup of individuals with a suspected genetic syndrome remains to be determined. Primarily, it should be used by expert geneticists.

## Acknowledgments

We thank the members of the Elterninitiative Apertsyndrom und Verwandte Fehlbildungen eV, a parents' initiative on Apert syndrome and related disorders in Germany, for the contribution of their images, and Yaron Gurovich and Nicole Fleischer of FDNA Inc for technical assistance in checking DeepGestalt's training set for duplicate images used in this study. MAM is a participant in the BIH Charité Digital Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin and the Berlin Institute of Health. We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Funds of Charité – Universitätsmedizin Berlin.

## Authors' Contributions

JTP, NH, and MAM designed the study. JTP, NH, MD, JE, ATAP, and MAM collected the data. SM, MS, DH, and CEO provided insights that were critical for the interpretation of data. MAM implemented the Python code with support from PH. PH and MAM performed the statistical analysis. JTP, NH, CEO, and MAM wrote the manuscript with approval of all the authors.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Code and data.

[ZIP File (Zip Archive), 137 KB - [jmir\\_v22i10e19263\\_app1.zip](#)]

### Multimedia Appendix 2

(A) Distribution of first-rank Gestalt scores for the images of White persons in the syndromic cohort and the matched control cohort (healthy). (B) Distribution of first-rank Gestalt scores for the images of persons with other ethnicities in the syndromic cohort and the matched control cohort (healthy).

[PNG File , 97 KB - [jmir\\_v22i10e19263\\_app2.png](#)]

### Multimedia Appendix 3

DeepGestalt's sensitivities: purple circles indicate the average of the entire syndromic cohort; for other symbols/coloring, see respective subfigure title.

[PNG File , 208 KB - [jmir\\_v22i10e19263\\_app3.png](#)]

### Multimedia Appendix 4

Performance of the SVM on the entire syndromic cohort and matched control cohort: X-axis number of top-rank Gestalt score used for vector construction per case. MCC: Matthews correlation coefficient. Note: rising tendency.

[PNG File , 46 KB - [jmir\\_v22i10e19263\\_app4.png](#)]

## References

1. Jackson M, Marks L, May GHW, Wilson JB. The genetic basis of disease. *Essays Biochem* 2018 Dec 03;62(5):643-723 [FREE Full text] [doi: [10.1042/EBC20170053](https://doi.org/10.1042/EBC20170053)] [Medline: [30509934](https://pubmed.ncbi.nlm.nih.gov/30509934/)]
2. Hart TC, Hart PS. Genetic studies of craniofacial anomalies: clinical implications and applications. *Orthod Craniofac Res* 2009 Aug;12(3):212-220 [FREE Full text] [doi: [10.1111/j.1601-6343.2009.01455.x](https://doi.org/10.1111/j.1601-6343.2009.01455.x)] [Medline: [19627523](https://pubmed.ncbi.nlm.nih.gov/19627523/)]
3. Xie Q, Faust K, Van Ommeren R, Sheikh A, Djuric U, Diamandis P. Deep learning for image analysis: Personalizing medicine closer to the point of care. *Crit Rev Clin Lab Sci* 2019 Jan;56(1):61-73. [doi: [10.1080/10408363.2018.1536111](https://doi.org/10.1080/10408363.2018.1536111)] [Medline: [30628494](https://pubmed.ncbi.nlm.nih.gov/30628494/)]
4. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med* 2019 Nov 19;11(1):70 [FREE Full text] [doi: [10.1186/s13073-019-0689-8](https://doi.org/10.1186/s13073-019-0689-8)] [Medline: [31744524](https://pubmed.ncbi.nlm.nih.gov/31744524/)]
5. Boehringer S, Vollmar T, Tasse C, Wurtz RP, Gillesen-Kaesbach G, Horsthemke B, et al. Syndrome identification based on 2D analysis software. *Eur J Hum Genet* 2006 Oct;14(10):1082-1089 [FREE Full text] [doi: [10.1038/sj.ejhg.5201673](https://doi.org/10.1038/sj.ejhg.5201673)] [Medline: [16773127](https://pubmed.ncbi.nlm.nih.gov/16773127/)]
6. Vollmar T, Maus B, Wurtz RP, Gillesen-Kaesbach G, Horsthemke B, Wiczorek D, et al. Impact of geometry and viewing angle on classification accuracy of 2D based analysis of dysmorphic faces. *Eur J Med Genet* 2008;51(1):44-53. [doi: [10.1016/j.ejmg.2007.10.002](https://doi.org/10.1016/j.ejmg.2007.10.002)] [Medline: [18054308](https://pubmed.ncbi.nlm.nih.gov/18054308/)]
7. Boehringer S, Guenther M, Sinigerova S, Wurtz RP, Horsthemke B, Wiczorek D. Automated syndrome detection in a set of clinical facial photographs. *Am J Med Genet A* 2011 Sep;155A(9):2161-2169. [doi: [10.1002/ajmg.a.34157](https://doi.org/10.1002/ajmg.a.34157)] [Medline: [21815261](https://pubmed.ncbi.nlm.nih.gov/21815261/)]
8. Ferry Q, Steinberg J, Webber C, FitzPatrick DR, Ponting CP, Zisserman A, et al. Diagnostically relevant facial gestalt information from ordinary photos. *Elife* 2014 Jun 24;3:e02020 [FREE Full text] [doi: [10.7554/eLife.02020](https://doi.org/10.7554/eLife.02020)] [Medline: [24963138](https://pubmed.ncbi.nlm.nih.gov/24963138/)]
9. Cerrolaza JJ, Porras AR, Mansoor A, Zhao Q, Summar M, Linguraru MG. Identification of dysmorphic syndromes using landmark-specific local texture descriptors Internet. 2016 Presented at: IEEE 13th International Symposium on Biomedical Imaging (ISBI); 13-16 April 2016; Prague, Czech Republic. [doi: [10.1109/isbi.2016.7493453](https://doi.org/10.1109/isbi.2016.7493453)]
10. Tu L, Porras A, Boyle A, Linguraru M. Analysis of 3D Facial Dysmorphology in Genetic Syndromes from Unconstrained 2D Photographs Internet. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. MICCAI 2018. Lecture Notes in Computer Science, vol 11070. Cham: Springer; 2018:347-355.
11. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med* 2019 Jan;25(1):60-64. [doi: [10.1038/s41591-018-0279-0](https://doi.org/10.1038/s41591-018-0279-0)] [Medline: [30617323](https://pubmed.ncbi.nlm.nih.gov/30617323/)]
12. Dudding-Byth T, Baxter A, Holliday EG, Hackett A, O'Donnell S, White SM, et al. Computer face-matching technology using two-dimensional photographs accurately matches the facial gestalt of unrelated individuals with the same syndromic form of intellectual disability. *BMC Biotechnol* 2017 Dec 19;17(1):90 [FREE Full text] [doi: [10.1186/s12896-017-0410-1](https://doi.org/10.1186/s12896-017-0410-1)] [Medline: [29258477](https://pubmed.ncbi.nlm.nih.gov/29258477/)]
13. Basel-Vanagaite L, Wolf L, Orin M, Larizza L, Gervasini C, Krantz ID, et al. Recognition of the Cornelia de Lange syndrome phenotype with facial dysmorphology novel analysis. *Clin Genet* 2016 May;89(5):557-563. [doi: [10.1111/cge.12716](https://doi.org/10.1111/cge.12716)] [Medline: [26663098](https://pubmed.ncbi.nlm.nih.gov/26663098/)]
14. Gripp KW, Baker L, Telegrafi A, Monaghan KG. The role of objective facial analysis using FDNA in making diagnoses following whole exome analysis. Report of two patients with mutations in the BAF complex genes. *Am J Med Genet A* 2016 Jul;170(7):1754-1762. [doi: [10.1002/ajmg.a.37672](https://doi.org/10.1002/ajmg.a.37672)] [Medline: [27112773](https://pubmed.ncbi.nlm.nih.gov/27112773/)]
15. Lumaka A, Cosemans N, Lulebo Mampasi A, Mubungu G, Mvuama N, Lubala T, et al. Facial dysmorphism is influenced by ethnic background of the patient and of the evaluator. *Clin Genet* 2017 Aug;92(2):166-171. [doi: [10.1111/cge.12948](https://doi.org/10.1111/cge.12948)] [Medline: [27925162](https://pubmed.ncbi.nlm.nih.gov/27925162/)]
16. Hady-Rabia S, Schneider H, Navarro E, Klein O, Kirby N, Huttner K, et al. Automatic recognition of the XLHED phenotype from facial images. *Am J Med Genet A* 2017 Sep;173(9):2408-2414. [doi: [10.1002/ajmg.a.38343](https://doi.org/10.1002/ajmg.a.38343)] [Medline: [28691769](https://pubmed.ncbi.nlm.nih.gov/28691769/)]
17. Gardner OK, Haynes K, Schweitzer D, Johns A, Magee WP, Urata MM, et al. Familial Recurrence of 3MC Syndrome in Consanguineous Families: A Clinical and Molecular Diagnostic Approach With Review of the Literature. *Cleft Palate Craniofac J* 2017 Nov;54(6):739-748. [doi: [10.1597/15-151](https://doi.org/10.1597/15-151)] [Medline: [27356087](https://pubmed.ncbi.nlm.nih.gov/27356087/)]
18. Valentine M, Bihm DCJ, Wolf L, Hoyne HE, May PA, Buckley D, et al. Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum Disorders. *Pediatrics* 2017 Dec;140(6):e20162028. [doi: [10.1542/peds.2016-2028](https://doi.org/10.1542/peds.2016-2028)] [Medline: [29187580](https://pubmed.ncbi.nlm.nih.gov/29187580/)]
19. Knaus A, Pantel JT, Pendziwiat M, Hajjir N, Zhao M, Hsieh T, et al. Characterization of glycosylphosphatidylinositol biosynthesis defects by clinical features, flow cytometry, and automated image analysis. *Genome Med* 2018 Jan 09;10(1):3 [FREE Full text] [doi: [10.1186/s13073-017-0510-5](https://doi.org/10.1186/s13073-017-0510-5)] [Medline: [29310717](https://pubmed.ncbi.nlm.nih.gov/29310717/)]
20. Liehr T, Acquarola N, Pyle K, St-Pierre S, Rinholm M, Bar O, et al. Next generation phenotyping in Emanuel and Pallister-Killian syndrome using computer-aided facial dysmorphology analysis of 2D photos. *Clin Genet* 2018 Feb;93(2):378-381. [doi: [10.1111/cge.13087](https://doi.org/10.1111/cge.13087)] [Medline: [28661575](https://pubmed.ncbi.nlm.nih.gov/28661575/)]

21. Zarate YA, Smith-Hicks CL, Greene C, Abbott M, Siu VM, Calhoun ARUL, et al. Natural history and genotype-phenotype correlations in 72 individuals with SATB2-associated syndrome. *Am J Med Genet A* 2018 Apr;176(4):925-935. [doi: [10.1002/ajmg.a.38630](https://doi.org/10.1002/ajmg.a.38630)] [Medline: [29436146](https://pubmed.ncbi.nlm.nih.gov/29436146/)]
22. Pantel JT, Zhao M, Mensah MA, Hajjir N, Hsieh T, Hanani Y, et al. Advances in computer-assisted syndrome recognition by the example of inborn errors of metabolism. *J Inherit Metab Dis* 2018 May;41(3):533-539 [FREE Full text] [doi: [10.1007/s10545-018-0174-3](https://doi.org/10.1007/s10545-018-0174-3)] [Medline: [29623569](https://pubmed.ncbi.nlm.nih.gov/29623569/)]
23. Ferreira CR, Altassan R, Marques-Da-Silva D, Francisco R, Jaeken J, Morava E. Recognizable phenotypes in CDG. *J Inherit Metab Dis* 2018 May;41(3):541-553 [FREE Full text] [doi: [10.1007/s10545-018-0156-5](https://doi.org/10.1007/s10545-018-0156-5)] [Medline: [29654385](https://pubmed.ncbi.nlm.nih.gov/29654385/)]
24. Jiang Y, Wangler MF, McGuire AL, Lupski JR, Posey JE, Khayat MM, et al. The phenotypic spectrum of Xia-Gibbs syndrome. *Am J Med Genet A* 2018 Jun;176(6):1315-1326 [FREE Full text] [doi: [10.1002/ajmg.a.38699](https://doi.org/10.1002/ajmg.a.38699)] [Medline: [29696776](https://pubmed.ncbi.nlm.nih.gov/29696776/)]
25. Graul-Neumann LM, Mensah MA, Klopocki E, Uebe S, Ekici AB, Thiel CT, et al. Biallelic intragenic deletion in MASPI in an adult female with 3MC syndrome. *Eur J Med Genet* 2018 Jul;61(7):363-368. [doi: [10.1016/j.ejmg.2018.01.016](https://doi.org/10.1016/j.ejmg.2018.01.016)] [Medline: [29407414](https://pubmed.ncbi.nlm.nih.gov/29407414/)]
26. Vorravanprecha N, Lertboonnum T, Rodjanadit R, Sriplienchan P, Rojnueangnit K. Studying Down syndrome recognition probabilities in Thai children with de-identified computer-aided facial analysis. *Am J Med Genet A* 2018 Sep;176(9):1935-1940. [doi: [10.1002/ajmg.a.40483](https://doi.org/10.1002/ajmg.a.40483)] [Medline: [30070762](https://pubmed.ncbi.nlm.nih.gov/30070762/)]
27. Martinez-Monseny A, Cuadras D, Bolasell M, Muchart J, Arjona C, Borregan M, et al. From gestalt to gene: early predictive dysmorphic features of PMM2-CDG. *J Med Genet* 2019 Apr;56(4):236-245. [doi: [10.1136/jmedgenet-2018-105588](https://doi.org/10.1136/jmedgenet-2018-105588)] [Medline: [30464053](https://pubmed.ncbi.nlm.nih.gov/30464053/)]
28. Pascolini G, Fleischer N, Ferraris A, Majore S, Grammatico P. The facial dysmorphology analysis technology in intellectual disability syndromes related to defects in the histones modifiers. *J Hum Genet* 2019 Aug;64(8):721-728. [doi: [10.1038/s10038-019-0598-0](https://doi.org/10.1038/s10038-019-0598-0)] [Medline: [31086247](https://pubmed.ncbi.nlm.nih.gov/31086247/)]
29. Mishima H, Suzuki H, Doi M, Miyazaki M, Watanabe A, Matsumoto T, et al. Evaluation of Face2Gene using facial images of patients with congenital dysmorphic syndromes recruited in Japan. *J Hum Genet* 2019 Aug;64(8):789-794. [doi: [10.1038/s10038-019-0619-z](https://doi.org/10.1038/s10038-019-0619-z)] [Medline: [31138847](https://pubmed.ncbi.nlm.nih.gov/31138847/)]
30. Danyel M, Cheng Z, Jung C, Boschann F, Pantel JT, Hajjir N, et al. Differentiation of MISSLA and Fanconi anaemia by computer-aided image analysis and presentation of two novel MISSLA siblings. *Eur J Hum Genet* 2019 Dec;27(12):1827-1835. [doi: [10.1038/s41431-019-0469-3](https://doi.org/10.1038/s41431-019-0469-3)] [Medline: [31320746](https://pubmed.ncbi.nlm.nih.gov/31320746/)]
31. Pascolini G, Valiante M, Bottillo I, Laino L, Fleischer N, Ferraris A, et al. Striking phenotypic overlap between Nicolaidis-Baraitser and Coffin-Siris syndromes in monozygotic twins with ARID1B intragenic deletion. *Eur J Med Genet* 2020 Mar;63(3):103739. [doi: [10.1016/j.ejmg.2019.103739](https://doi.org/10.1016/j.ejmg.2019.103739)] [Medline: [31421289](https://pubmed.ncbi.nlm.nih.gov/31421289/)]
32. Kruszka P, Hu T, Hong S, Signer R, Cogné B, Isidor B, et al. Phenotype delineation of ZNF462 related syndrome. *Am J Med Genet A* 2019 Oct;179(10):2075-2082 [FREE Full text] [doi: [10.1002/ajmg.a.61306](https://doi.org/10.1002/ajmg.a.61306)] [Medline: [31361404](https://pubmed.ncbi.nlm.nih.gov/31361404/)]
33. Fung JLF, Rethanavelu K, Luk H, Ho MSP, Lo IFM, Chung BHY. Coffin-Lowry syndrome in Chinese. *Am J Med Genet A* 2019 Oct;179(10):2043-2048. [doi: [10.1002/ajmg.a.61323](https://doi.org/10.1002/ajmg.a.61323)] [Medline: [31400053](https://pubmed.ncbi.nlm.nih.gov/31400053/)]
34. Weiss K, Lazar HP, Kurolap A, Martinez AF, Paperna T, Cohen L, et al. The CHD4-related syndrome: a comprehensive investigation of the clinical spectrum, genotype-phenotype correlations, and molecular basis. *Genet Med* 2020 Feb;22(2):389-397. [doi: [10.1038/s41436-019-0612-0](https://doi.org/10.1038/s41436-019-0612-0)] [Medline: [31388190](https://pubmed.ncbi.nlm.nih.gov/31388190/)]
35. Zarate YA, Bosanko KA, Gripp KW. Using facial analysis technology in a typical genetic clinic: experience from 30 individuals from a single institution. *J Hum Genet* 2019 Dec;64(12):1243-1245. [doi: [10.1038/s10038-019-0673-6](https://doi.org/10.1038/s10038-019-0673-6)] [Medline: [31551534](https://pubmed.ncbi.nlm.nih.gov/31551534/)]
36. Narayanan DL, Ranganath P, Aggarwal S, Dalal A, Phadke SR, Mandal K. Computer-aided Facial Analysis in Diagnosing Dysmorphic Syndromes in Indian Children. *Indian Pediatr* 2019 Dec 15;56(12):1017-1019 [FREE Full text] [Medline: [31884430](https://pubmed.ncbi.nlm.nih.gov/31884430/)]
37. Latorre-Pellicer A, Ascaso Á, Trujillano L, Gil-Salvador M, Arnedo M, Lucia-Campos C, et al. Evaluating Face2Gene as a Tool to Identify Cornelia de Lange Syndrome by Facial Phenotypes. *Int J Mol Sci* 2020 Feb 04;21(3):1042 [FREE Full text] [doi: [10.3390/ijms21031042](https://doi.org/10.3390/ijms21031042)] [Medline: [32033219](https://pubmed.ncbi.nlm.nih.gov/32033219/)]
38. Arora V, Puri RD, Bijarnia-Mahay S, Verma IC. Expanding the phenotypic and genotypic spectrum of Wiedemann-Steiner syndrome: First patient from India. *Am J Med Genet A* 2020 May;182(5):953-956. [doi: [10.1002/ajmg.a.61534](https://doi.org/10.1002/ajmg.a.61534)] [Medline: [32128942](https://pubmed.ncbi.nlm.nih.gov/32128942/)]
39. Carli D, Giorgio E, Pantaleoni F, Bruselles A, Barresi S, Riberi E, et al. NBAS pathogenic variants: Defining the associated clinical and facial phenotype and genotype-phenotype correlations. *Hum Mutat* 2019 Jun;40(6):721-728. [doi: [10.1002/humu.23734](https://doi.org/10.1002/humu.23734)] [Medline: [30825388](https://pubmed.ncbi.nlm.nih.gov/30825388/)]
40. Stauffer C, Peters B, Wagner M, Alameer S, Barić I, Broué P, et al. Defining clinical subgroups and genotype-phenotype correlations in NBAS-associated disease across 110 patients. *Genet Med* 2020 Mar;22(3):610-621. [doi: [10.1038/s41436-019-0698-4](https://doi.org/10.1038/s41436-019-0698-4)] [Medline: [31761904](https://pubmed.ncbi.nlm.nih.gov/31761904/)]



41. Myers L, Anderlid B, Nordgren A, Lundin K, Kuja-Halkola R, Tammimies K, et al. Clinical versus automated assessments of morphological variants in twins with and without neurodevelopmental disorders. *Am J Med Genet A* 2020 May 12;182(5):1177-1189. [doi: [10.1002/ajmg.a.61545](https://doi.org/10.1002/ajmg.a.61545)] [Medline: [32162839](https://pubmed.ncbi.nlm.nih.gov/32162839/)]
42. Hsieh T, Mensah MA, Pantel JT, Aguilar D, Bar O, Bayat A, et al. PEDIA: prioritization of exome data by image analysis. *Genet Med* 2019 Dec;21(12):2807-2814 [FREE Full text] [doi: [10.1038/s41436-019-0566-2](https://doi.org/10.1038/s41436-019-0566-2)] [Medline: [31164752](https://pubmed.ncbi.nlm.nih.gov/31164752/)]
43. MacCarrick G, Black JH, Bowdin S, El-Hamamsy I, Frischmeyer-Guerrerio PA, Guerrerio AL, et al. Loeys-Dietz syndrome: a primer for diagnosis and management. *Genet Med* 2014 Aug;16(8):576-587 [FREE Full text] [doi: [10.1038/gim.2014.11](https://doi.org/10.1038/gim.2014.11)] [Medline: [24577266](https://pubmed.ncbi.nlm.nih.gov/24577266/)]
44. Bird RJ, Hurren BJ. Anatomical and clinical aspects of Klinefelter's syndrome. *Clin Anat* 2016 Jul;29(5):606-619. [doi: [10.1002/ca.22695](https://doi.org/10.1002/ca.22695)] [Medline: [26823086](https://pubmed.ncbi.nlm.nih.gov/26823086/)]
45. Martinez-Martin N. What Are Important Ethical Implications of Using Facial Recognition Technology in Health Care? *AMA J Ethics* 2019 Mar 01;21(2):E180-E187 [FREE Full text] [doi: [10.1001/amajethics.2019.180](https://doi.org/10.1001/amajethics.2019.180)] [Medline: [30794128](https://pubmed.ncbi.nlm.nih.gov/30794128/)]

## Abbreviations

**AUROC:** area under the receiver operating characteristic

**MCC:** Matthews correlation coefficient

**PEDIA:** prioritization of exome data by image analysis

**ROC:** receiver operating characteristic

**SVM:** support vector machine

*Edited by G Eysenbach; submitted 10.04.20; peer-reviewed by T Liehr, G Pascolini, M Pradhan, D Szinay; comments to author 12.06.20; revised version received 26.06.20; accepted 26.07.20; published 22.10.20.*

*Please cite as:*

*Pantel JT, Hajjir N, Danyel M, Elsner J, Abad-Perez AT, Hansen P, Mundlos S, Spielmann M, Horn D, Ott CE, Mensah MA Efficiency of Computer-Aided Facial Phenotyping (DeepGestalt) in Individuals With and Without a Genetic Syndrome: Diagnostic Accuracy Study*

*J Med Internet Res* 2020;22(10):e19263

URL: <http://www.jmir.org/2020/10/e19263/>

doi: [10.2196/19263](https://doi.org/10.2196/19263)

PMID: [33090109](https://pubmed.ncbi.nlm.nih.gov/33090109/)

©Jean Tori Pantel, Nurulhuda Hajjir, Magdalena Danyel, Jonas Elsner, Angela Teresa Abad-Perez, Peter Hansen, Stefan Mundlos, Malte Spielmann, Denise Horn, Claus-Eric Ott, Martin Atta Mensah. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Stress Tracker—Detecting Acute Stress From a Trackpad: Controlled Study

Rahul Goel<sup>1\*</sup>, PhD; Michael An<sup>1,2\*</sup>; Hugo Alayrangues<sup>1,3</sup>; Amirhossein Koneshloo<sup>1,4</sup>, PhD; Emmanuel Thierry Lincoln<sup>1,3</sup>; Pablo Enrique Paredes<sup>5</sup>, MBA, PhD

<sup>1</sup>Department of Radiology, School of Medicine, Stanford University, Palo Alto, CA, United States

<sup>2</sup>Department of Computer Science, University of California Los Angeles, Los Angeles, CA, United States

<sup>3</sup>Department of Computer Science, Institut supérieur d'électronique de Paris, Paris, France

<sup>4</sup>Department of Industrial, Manufacturing, and Systems Engineering, Texas Tech University, Lubbock, TX, United States

<sup>5</sup>Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Palo Alto, CA, United States

\*these authors contributed equally

**Corresponding Author:**

Pablo Enrique Paredes, MBA, PhD

Department of Psychiatry and Behavioral Sciences

School of Medicine

Stanford University

1701 Page Mill Road

Palo Alto, CA,

United States

Phone: 1 4152723787

Email: [paredes@cs.stanford.edu](mailto:paredes@cs.stanford.edu)

## Abstract

**Background:** Stress is a risk factor associated with physiological and mental health problems. Unobtrusive, continuous stress sensing would enable precision health monitoring and proactive interventions, but current sensing methods are often inconvenient, expensive, or suffer from limited adherence. Prior work has shown the possibility to detect acute stress using biomechanical models derived from passive logging of computer input devices.

**Objective:** Our objective is to detect acute stress from passive movement measurements of everyday interactions on a laptop trackpad: (1) *click*, (2) *steer*, and (3) *drag and drop*.

**Methods:** We built upon previous work, detecting acute stress through the biomechanical analyses of canonical computer mouse interactions and extended it to study similar interactions with the trackpad. A total of 18 participants carried out 40 trials each of three different types of movement—(1) *click*, (2) *steer*, and (3) *drag and drop*—under both relaxed and stressed conditions.

**Results:** The mean and SD of the contact area under the finger were higher when clicking trials were performed under stressed versus relaxed conditions (mean area:  $P=.009$ , effect size=0.76; SD area:  $P=.01$ , effect size=0.69). Further, our results show that as little as 4 clicks on a trackpad can be used to detect binary levels of acute stress (ie, whether it is present or not).

**Conclusions:** We present evidence that scalable, inexpensive, and unobtrusive stress sensing can be done via repurposing passive monitoring of computer trackpad usage.

(*J Med Internet Res* 2020;22(10):e22743) doi:[10.2196/22743](https://doi.org/10.2196/22743)

**KEYWORDS**

precision health; well-being; trackpad; computer input device; computer interaction; stress sensing; affective interfaces; mental health

## Introduction

### Overview

Several health risks such as cardiovascular disease and immune deficiencies, which can diminish the quality of life and shorten life expectancy [1], are linked to repetitive daily acute stress (ie, short-term response to a perceived threat or challenge [2,3]). Stress can also have a profound impact on cognitive and emotional well-being. Advances in wearable technology enable affective and cognitive state measurements. Still, wearable devices can suffer from high attrition and low adoption, in general, and getting stable stress measurements “in the wild” (eg, heart rate variability [HRV] and electrodermal activity [EDA]) remains challenging.

Previous lab studies have shown that data from interactions with everyday devices, such as computer mouse movements, keyboard pressure, smartphone touch, and key swipes, or even the steering wheel of a car, can be used in an unobtrusive and scalable way to detect the presence of acute stress [4-8]. To the best of our knowledge, this paper presents the first work showing the feasibility of detecting acute stress from interactions with laptop computer trackpads.

There are several billion personal computing devices (PCDs) deployed in the world, and this number increases by around 300 million every year [9]. Laptops, or notebooks, represent the largest (42%) and the fastest-growing segment [9] among PCDs.

Most laptop users prefer the trackpad over an external mouse, for improved mobility or due to space limitations. Investigating biomechanics and motor control of finger dynamics during trackpad use can lead to reliable detection of acute stress, especially for long-term unobtrusive continuous monitoring. We propose two research questions (RQs):

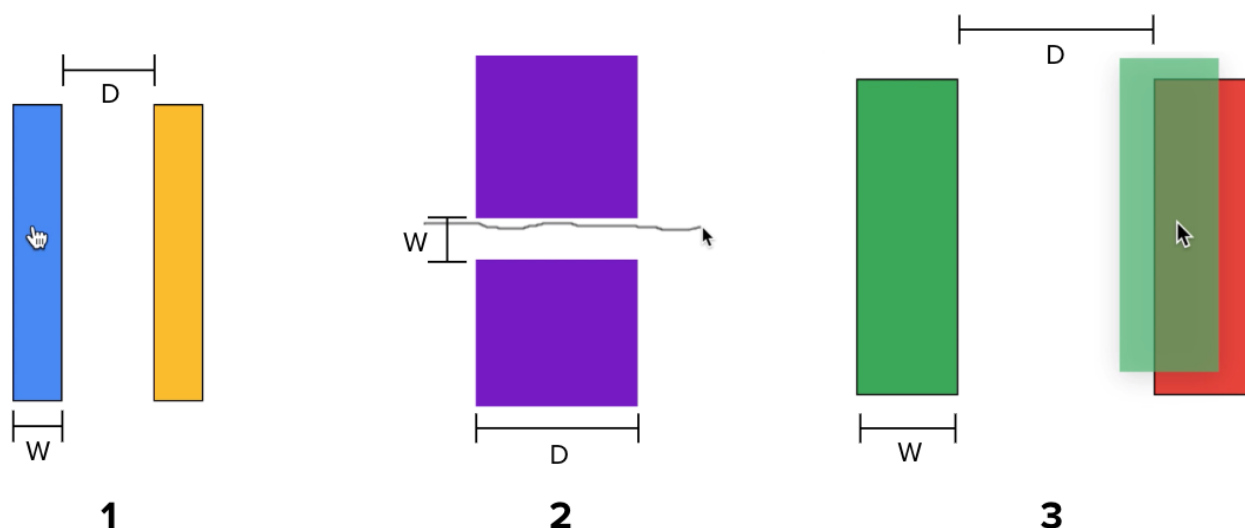
1. RQ1: Can we assess differences between stress versus relaxation through changes in the damping ratio ( $\Gamma$ ) and damping frequency ( $\omega$ ) of a mass-spring-damper (MSD) model derived from finger strokes on a laptop trackpad, similar to our previous study on a computer mouse [4]?
2. RQ2: Are there differences between stress and relaxation in other metrics of finger dynamics, such as contact area, velocity, and acceleration, which are theoretically linked to muscle tension?

We conducted a within-subjects study (N=18) counterbalancing relaxed and stressed conditions. Subjects performed 40 trials each of three canonical computer tasks using a trackpad, namely, *clicking*, *steering*, and *drag and drop* (see Table 1 and Figure 1), similar to what Sun et al did with a computer mouse [4]. We focused our analyses on metrics derived from finger dynamics. Our results confirm a significantly higher mean and SD of the contact area under the finger for the tasks performed under stress compared to relaxed conditions, even among the initial 10% of the data (ie, 4 out of 40 *clicks*) from *clicking* trials. To the best of our knowledge, this is the first systematic study that links acute stress to finger strokes on the trackpad of a laptop device.

**Table 1.** Different task configurations by varying the distance and width parameters.

Task type	Distance by configuration, pixels					Width by configuration, pixels			
	1	2	3	4	5	1	2	3	4
Click	64	128	256	512	1024	8	16	32	64
Drag and drop	64	128	256	512	1024	16	32	64	128
Steer	64	128	256	512	1024	8	16	32	64

**Figure 1.** Computer tasks used in the study: (1) *click*, (2) *steer*, and (3) *drag and drop*. Different combinations of width (W) and distance (D) were randomly presented (see Table 1). In the *click* trials, subjects had to click anywhere on the blue rectangle first and then the yellow rectangle. In the *steer* trials, subjects had to hold down the left key and steer across the distance between the two rectangles. In the *drag-and-drop* trials, subjects had to click on the green rectangle, hold it, drag it, and then drop it over the red rectangle.



## Background

### Stress Measurements

Stress can be measured via self-reports or physiological signals. Self-reported stress (SRS) can be measured through the Perceived Stress Scale (PSS) [10,11]. Typically, a simplified version with a single 11-item scale of stress, ranging from 0 to 10, is used in repeated-measure studies [4,12]. Although subjective responses are commonly used, they are not precise and suffer from response bias and noncompliance [13], making them impractical for long-term monitoring.

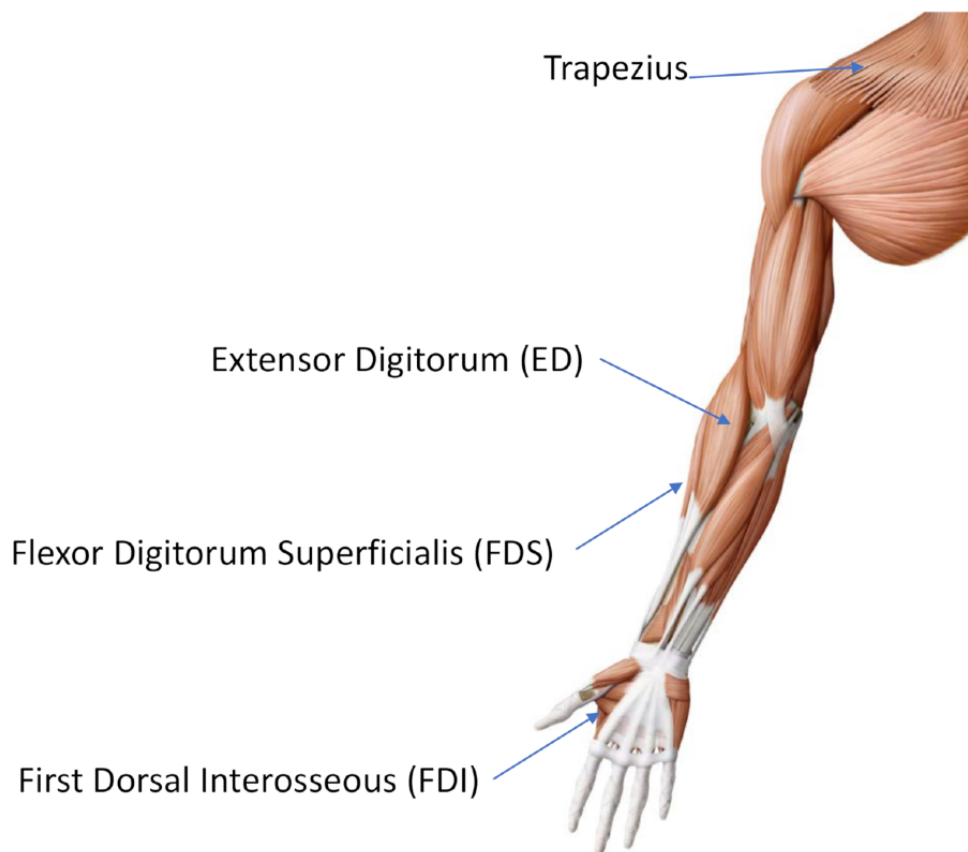
Complementarily, physiological measurement of stress can be done by an indirect observation of the autonomic nervous system (ANS) signal [14]. A popular metric is EDA [15,16]. EDA, formerly known as the galvanic skin response, is a measurement of skin conductance due to the activation of the eccrine sweat glands, which are only innervated by the sympathetic nervous system (SNS). The SNS is one of the branches of the ANS linked to the “fight or flight” response, the coordinated response by the body organs to a threat signal (ie, stressor). High average levels and an increase in skin conductance responses (SCRs) (ie, the number of EDA peaks) are associated with stress [16]. Another popular measure is HRV, which is the measurement of the time variation between R-R peaks of an electrocardiogram (ECG) signal due to the sinusoidal arrhythmia [14]. EDA and ECG can be obtained in ambulatory settings using wearable devices. However, these sensors are obtrusive, leading to dropout, and are also affected by motion artifacts. In this paper, we validated our stressor with SRS and EDA.

### Links Between Stress, Motor Control, and Muscle Tension

Previous research suggests that hormones and neurotransmitters released during heightened stress arousal can influence motor output [17]. Noteboom et al [17] carried out a precision task that required a submaximal isometric pinch grip. Their research found that the steadiness in electromyography (EMG) activity in the first dorsal interosseous (FDI) muscle of the hand (see Figure 2), and the flexor digitorum superficialis muscle of the forearm, was increased during stress, especially in those with moderate anxiety trait compared to those with low anxiety trait. Laidlaw et al [18] measured the motor unit activity using intramuscular EMG recordings in the FDI during a pinch grip task. They concluded that increased variability of motor unit discharge is associated with reductions in steadiness. Coombes et al [19] showed that the force production in wrist and finger extensor muscles was increased during continuous exposure to unpleasant stimuli, compared to neutral stimuli, for a voluntary bimanual maximal isometric contraction task. Arnich et al [20] showed that nervous or anxious individuals have higher levels of movement variability while sitting (ie, a higher center of pressure dispersion on the seat). These studies provide evidence on how acute stress can influence variability while performing a motor task.

Additionally, various studies have shown that acute mental stress increases muscle tension in the neck, forehead, and arms [21-24]. The shoulder's trapezius muscle [25], as well as biceps and triceps [26], have shown direct changes due to mental arithmetic tasks. In this study, we used two types of biomechanical metrics: those that are influenced by changes in muscle tension and motor control variability due to stress.

**Figure 2.** Main muscles activated when using computer input devices: mouse and trackpad.



### ***Effects of Stress on the Manipulation of Computer Input Devices***

Sun et al [4] showed a direct correlation between stress, muscle tension, and movement of a computer mouse by approximating the mouse movement trajectories as a step response of an MSD system. They calculated the  $\Gamma$  and  $\omega$ , both of which were higher under stress. Hernandez et al showed a higher contact area with the surface of a capacitive mouse under stressful conditions compared to relaxed conditions [6]. Wahlstrom et al [27] showed that muscle activity in the right extensor digitorum (ED) and right trapezius muscles were greater in stressed situations evoked by time pressure and verbal provocation while working with a computer mouse. Carneiro et al [28] showed that acceleration and mean and maximum intensity (aka, pressure) of touch when interacting with touchscreen devices were higher under stressful conditions. While not the same as stress, Gao et al [7] showed that the velocity and contact area of the finger strokes during the *frustrated* state in comparison to the *relaxed* state were among the main features used in a machine learning (ML) prediction model of *frustration* during game-playing interactions on an iPod. To the best of our knowledge, there is no prior research on systematic stress detection using finger dynamics during laptop trackpad use.

### ***Biomechanics of Trackpad Versus Mouse Usage***

It is important to understand the biomechanical differences between trackpad and mouse use [29]. Direct comparisons between methods and metrics used with a computer mouse may not be feasible for a trackpad. Mouse use primarily triggers the large trapezius muscles for moving the entire arm leading to

horizontal and vertical movement of the cursor on the screen [29]. Although both mouse and trackpad movements use the ED muscle for *clicking*, trackpad displacement mainly involves the rather small FDI hand muscle [29]. Figure 2 shows the relative sizes of the muscles of the arm and hand.

In comparison to the trackpad, mouse use induces larger shoulder abduction and smaller elbow flexion [29]. Trackpad use requires a more static posture of the upper arm to maintain stabilization of the hand but a more rigorous movement of fingers [29]. Stress detection via an approximation of muscle tension [4] benefits from more muscle activity. However, the absolute level of muscle activity in the FDI during trackpad usage is low compared to that in the bigger trapezius muscle during computer mouse usage [29,30]. As a result, it could be hard to rely on muscle tension and stiffness to infer stress using trackpad versus mouse data [4,27].

## ***Methods***

### ***Subjects***

We recruited 22 subjects (10 males [45%] and 12 females [55%]), with ages ranging from 19 to 68 (mean 37, SD 16), not screened for preferences in using a computer mouse or trackpad. We eliminated 4 subjects for whom our stressor did not have any effect on SRS, leaving a total of 18 subjects (8 males [44%] and 10 females [56%]). Of the remaining 18 subjects, 11 (61%) reported using a computer mouse and 9 (50%) reported using a laptop trackpad daily. The mean self-reported daily usage duration of the laptop trackpad (mean 2.6, SD 3.1 hours) was significantly lower ( $t_{17}=2.1$ ,  $P=.03$ ) than that of the computer



mouse (mean 5.2, SD 3.1 hours). Subjects provided written informed consent before participation and were given US \$20 gift cards for their participation. The study was approved by the Institutional Review Board of Stanford University.

### Apparatus

The experiment was performed in an office setting without any distractions. We used a 15-inch 2015 MacBook Pro (Apple Inc) (see Figure 3) equipped with a 140×70-mm Force Touch trackpad with sensitivity preset to a default value of 1.0. An

overhead camera captured hand movement, and our logging software captured trackpad and cursor activity. The logger was implemented in C and Swift (Apple Inc) using Apple's *MultiTouchSupport* framework. It recorded coordinates rounded to the nearest hundredth of a millimeter with an average sampling rate of 8.17 milliseconds (SD 3.75), multiple finger touches marked with ID numbers, contact shape under the finger (ie, major and minor axes of the ellipse), interaction type (ie, touching, dragging, etc), and pressed state (ie, active or inactive).

**Figure 3.** Apparatus: a 15-inch 2015 MacBook Pro (Apple Inc) with a passive movement logger teamed with an overhead camera for the ancillary recording of hand movement.



### Stimuli

The stimuli consisted of *Relaxation* and *Stressor* phases replicated from prior studies [4,5]. During the *Relaxation* phase, we instructed subjects to breathe deeply while viewing a soothing video, which is recommended over doing nothing [31]. The acute *Stressor* phase consisted of an arithmetic cognitive stressor combined with social evaluation (ie, the presence of an experimenter continuously observing and correcting if the subject made a mistake), derived from the Trier Social Stress Test [32], and enhanced with a biased financial stimulus. The test required subjects to perform a series of subtractions out loud (eg, subtract 13 from 2017, and so forth) under the scrutiny

of the experimenter. If the subject made a mistake, or if he or she took more than four seconds to respond, the experimenter asked the subject to start again from 2017. The subject was offered an additional US \$1 for every 10 consecutive arithmetic operations performed correctly, but only US \$1 was to be deducted for each mistake. All subjects made mistakes that made them “lose” money, but after debriefing, they all received full compensation. The financial part of the stimulus based on performance could have been more frustrating than stressful for some subjects. However, it was added, as was done in our previous study [4], since frustration can further exacerbate acute stress.



## Experimental Tasks

All subjects performed three types of tasks—*click*, *steer*, and *drag and drop* (see Figure 1)—as in our previous study using the computer mouse [4]. Five different distance (D) and four different width (W) configurations were used for each of the three task types, resulting in a total of 20 configurations per

task (see Table 1) [4]. Users performed two trials per configuration within each task type (see Figure 4). Tasks of the same type were grouped to prevent performance differences due to subjects needing to adjust to different task types. In total, each experimental condition, *tStressed* (stressed task) and *tRelaxed* (relaxed task), had 120 trials (3 task types × 40 trials).

**Figure 4.** Example of distribution of different task types across different subjects.

Subject No.	Task Type 1	Task Type 2	Task Type 3
A1	Click1, Click2, ....., Click40	Drag1, Drag2, ....., Drag40	Steer1, Steer2, ....., Steer40
A2	Drag1, Drag2, ....., Drag40	Steer1, Steer2, ....., Steer40	Click1, Click2, ....., Click40
⋮			
B9	Steer1, Steer2, ....., Steer40	Click1, Click2, ....., Click40	Drag1, Drag2, ....., Drag40

## Experimental Design

### Testing Procedure

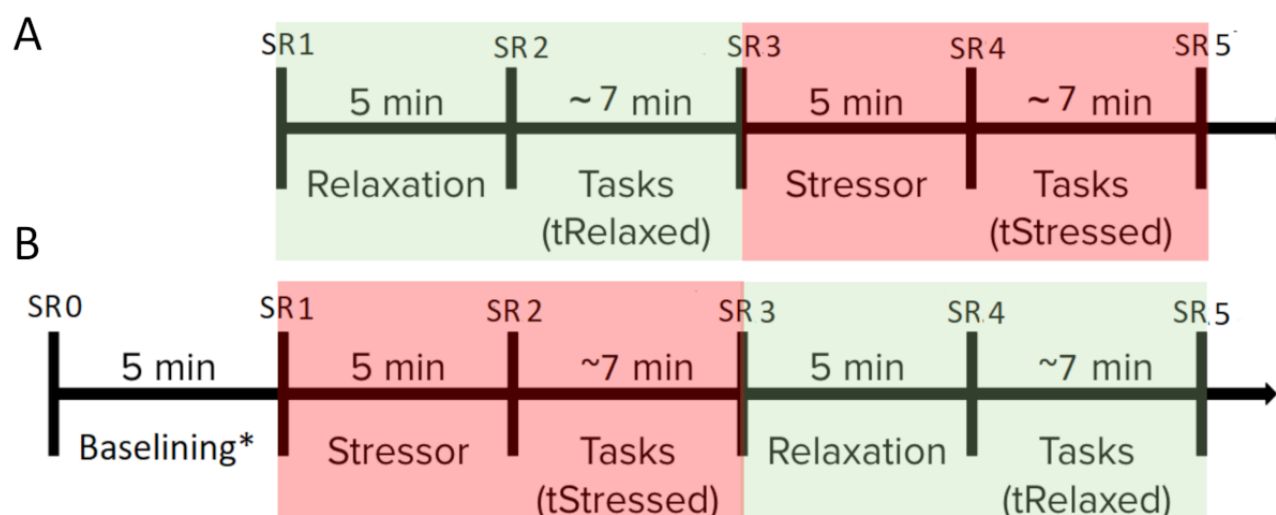
The experiment consisted of four distinct phases (see Figure 5):

1. *Baselining and Relaxation*: a 5-minute phase designed to normalize the effects of any stressors, including any external factors, and to bring the subject to a baseline level.
2. *tRelaxed*: an approximately 7-minute phase (mean 6.8, SD 1.7 minutes) during which subjects were instructed to perform the randomized tasks—*click*, *drag and drop*, and *steer* (see Figure 1)—as quickly and accurately as possible.
3. *Stressor*: a 5-minute stress-inducing phase.

4. *tStressed*: an approximately 7-minute phase (mean 6.8, SD 1.5 minutes) during which subjects were instructed to perform the randomized tasks—*click*, *drag and drop*, and *steer* (see Figure 1)—as quickly and accurately as possible.

The 18 subjects were assigned to one of the two counterbalanced arms: *Relax-Stress* (3 males [17%] and 5 females [28%]) and *Stress-Relax* (5 males [28%] and 5 females [28%]) (see Figure 5). The *Stress-Relax* arm required an additional baselining phase, before *tRelaxed*, to normalize the effects of external factors. Subjects wore an E4 wristband (Empatica Inc) to measure EDA in their nondominant hand and provided perceived self-reported *stress*, *tension*, and *concentration* levels on a scale of 0-10, before and after each phase (see Figure 5).

**Figure 5.** Study procedure with (A) *Relax-Stress* and (B) *Stress-Relax* counterbalanced arms. The *Relax* part is shown with a green transparent box and the *Stress* part is shown with a red transparent box. Data were collected at self-report marker (SR) time points 0-5. \*For the *Stress-Relax* arm, before the stressor phase, a baselining phase was carried out to normalize the effects of external factors. *tRelaxed*: relaxed task; *tStressed*: stressed task.



## Data Processing

### Self-Report and Ancillary Data

SRS was assessed using a simplified version of the PSS [11], an 11-point scale question—“What is your current level of stress?” where responses ranged from *Low* (0) to *High*

(10)—immediately after completion of each phase (see Figure 5). To verify that the stressor did not induce cognitive alterations, we assessed perceived *concentration*, ranging from *Low* (0) to *High* (10), and measured *trial completion time*. Finally, we captured ancillary data on perceived muscle *tension*, ranging from *None* (0) to *High* (10). Self-report markers (SRs)

in Figure 5 show the time points during the experiment where these data were collected. To estimate the self-reported metrics *during* any phase, the before and after values were averaged for that phase. For example, the average of SR4 and SR5 values for stress for subjects in arm A (see Figure 5) gave us an estimate of the SRS value *during* the *tStressed* phase.

All self-reported metrics were min-max normalized across the four phases—*Relaxation*, *Stressor*, *tRelaxed*, and *tStressed*—unless stated otherwise. For normally distributed data, where the Shapiro-Wilk test was not significant, we applied a 1-tailed paired *t* test. Otherwise, we applied a 1-tailed Wilcoxon signed-rank test. We used 1-sided comparisons, as we had prior knowledge about the expected direction of changes in the different metrics, and the stressor task we used has already been shown to be effective in inducing stress in prior studies [4,5,32]. Effect sizes (Cohen *d*) are provided for within-group changes for primary measures, where Cohen *d* values between 0.20 and 0.49 indicate a small effect, values between 0.50 and 0.79 indicate a medium effect, and values of 0.80 and above indicate a large effect [33]. For any bivariate correlations, we used the Pearson correlation coefficient for normally distributed data; otherwise, we used Spearman rho.

Out of 22 subjects, 4 (18%) showed no change in SRS across the two extreme phases—*Stressor* and *Relaxation*—and were excluded from all analyses, leaving a working set of 18 subjects.

### EDA Processing

We eliminated 3 of the 18 subjects (17%) before the analysis, leaving a total of 15 subjects (83%) for EDA analysis: 1 subject did not have a signal due to loose contact, another had an anomalous signal above 20  $\mu$ S [34], and 1 had more than 5% noisy 5-second epochs estimated with the *EDA Explorer* [35]. For the remaining 15 subjects, EDA signals were filtered using a 6th-order 1-Hz low-pass Butterworth filter [36]. Then, using Ledalab, version 3.2.5 [37], we obtained the tonic mean, phasic

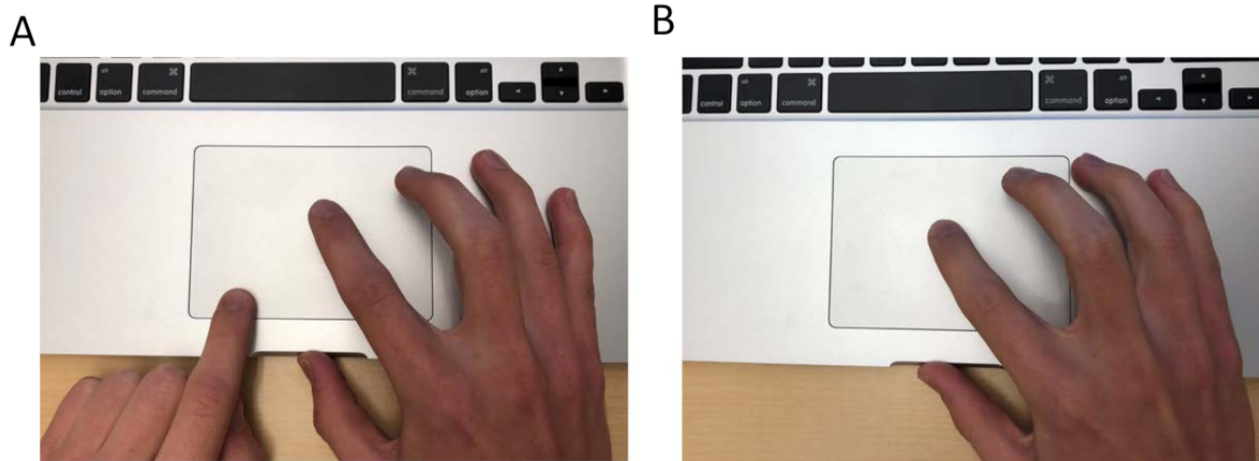
mean, and the number of SCR peaks with an amplitude above 0.01  $\mu$ S. Only data from *tRelaxed* and *tStressed* phases were considered for the EDA analyses.

### RQ1: MSD Model Evaluation

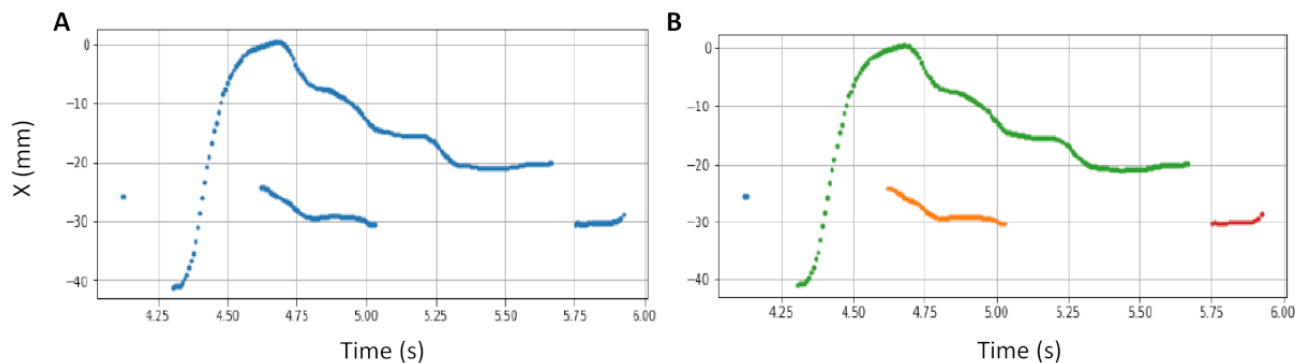
We modeled trackpad finger trajectories using a linear predictive coding (LPC) inverse filtering technique [38] as in Sun et al, who successfully modeled the arm movement as an approximation of a step response of a single-degree-of-freedom MSD system [4]. The LPC approximation for the MSD model has been used in several other studies, like in our previous work using steering wheel data [5] and also by Guo et al [39], using data from touchpad interactions. However, unlike mouse movements, which are primarily continuous and more standardized across subjects, there are many quirks in the way different individuals interact with the trackpad. Since subjects were given the freedom to use the trackpad as normally as possible, we observed that 17 out of 18 subjects (94%) used multiple fingers from their dominant hand, and/or used fingers from both hands (see Figure 6), creating many discontinuities in the finger strokes during each trial (see Figure 7). Across all subjects, a mean of 38.0% (SD 13.9) of the trials had at least one gap in the continuity of the finger stroke because subjects either used multiple hands (see Figure 6, A), multiple fingers (see Figure 6, B), or even the same finger, but the finger hovered on the trackpad. We used our ancillary videos to observe this behavior, as it was impossible to use logger data alone to verify that the subject used the same finger multiple times or separate fingers during one trial.

We identify gaps (see Figure 7, A) by searching for large variations in x or y coordinates and keeping points that represented the largest continuous movement of the underlying stroke (see Figure 7, B—green trace). We fitted an approximation to the MSD model [4] to this stroke and extracted  $\Gamma$  and  $\omega$  for the horizontal x-direction, which contained most of the displacement information for our tasks.

**Figure 6.** (A) multiple hands or (B) multiple fingers being used while interacting with the trackpad.



**Figure 7.** Example traces from a *click* trial. (A) Multiple traces at the same time (see between 4.6 and 5.0 seconds) suggests the use of multiple fingers from the same or different hands at the same time. The gap between 5.70 and 5.75 seconds indicates that the subject lifted his or her finger and then placed the same or another finger back. (B) Identifying the largest stroke (green trace) and other ancillary strokes due to the use of different fingers (orange and red traces). X represents the horizontal movement on the trackpad.



## RQ2: Measurements of Finger Dynamics From Trackpad Interactions

Under stress, wrist and forearm muscles present changes in mean force production [19] and motor control variability [17]. Therefore, we measured mean and SD values for three typical finger dynamics measurements: velocity, acceleration, and contact area under the finger. We kept all strokes and sub-strokes for every trial by identifying the gaps between partial movements (see Figure 7, B) and averaged their metrics for each trial. We estimated the contact area of the ellipse ( $\pi \times a \times b$ ) between the finger and the trackpad using the major axis (a) and minor axis (b) provided by the logger. We calculated finger velocity by differentiating position information and finger acceleration by differentiating velocity information. We used only horizontal (x-direction) displacement data, which contained most of the information for our tasks.

## Statistical Modeling and Sensitivity Analysis

First, we evaluated a *Mixed Task* omnibus model independent of specific task types—*click*, *drag and drop*, and *steer*—and task configurations ( $W \times D$ ). Values for each task type were obtained by averaging across all configurations and repetitions (ie, 40 trials) for each of the *tStressed* and *tRelaxed* phases for each subject. Then, a single value for each *tStressed* and *tRelaxed* phase was obtained after averaging across all three task types for each subject. Variables with significant differences across the *tStressed* and *tRelaxed* phases in the *Mixed Task* model were further examined using *Task-Specific* models. To

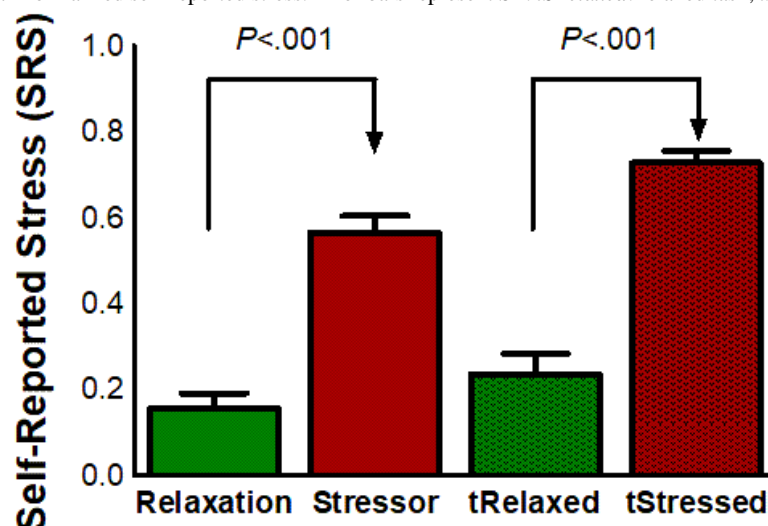
avoid inflation, we applied Bonferroni correction ( $0.05/3 = 0.017$ ) for *Task-Specific* models. Finally, for the type of task with the most significant difference, we applied a sensitivity analysis comparing the *initial* 10% (ie, 4 trials out of 40) of data across the *tStressed* and *tRelaxed* phases. We also carried out post hoc power ( $\beta$ ) analyses on all finger dynamics measures that were found to be significant. Additionally, to check if the order of stress-relax condition has any effect, we investigated the *Order* (ie, A vs B; see Figure 5)  $\times$  *Condition* (ie, stress vs relax) interaction effect using analyses of variance on all finger dynamics measures that were found to be significant.

## Results

### Mental Stress Validation

#### SRS and Tension Metrics

On average, the SRS was significantly higher during the *Stressor* phase (mean 0.56, SE 0.04) compared to the *Relaxation* phase (mean 0.15, SE 0.03) ( $Z_{17}=3.72$ ,  $P<.001$ ). The SRS was also significantly higher during the *tStressed* phase (mean 0.72, SE 0.03) compared to the *tRelaxed* phase (mean 0.23, SE 0.04) ( $t_{17}=13.06$ ,  $P<.001$ ). Figure 8 shows a comparison of min-max normalized SRS values for the four phases of the experiment. Table 2 shows the average raw SRS values at the end of the *Relaxation*, *tRelaxed*, *Stressor*, and *tStressed* phases. It can be observed that SRS levels go down over time during the *tStressed* phase.

**Figure 8.** Levels of min-max normalized self-reported stress. Error bars represent SE. *tSRelaxed*: relaxed task; *tStressed*: stressed task.**Table 2.** Average raw self-reported stress values across all subjects at the end of different phases.

Time point	Mean (SE)
End of <i>Relaxation</i> phase	2.27 (0.42)
End of <i>tRelaxed</i> <sup>a</sup> phase	3.78 (0.42)
End of <i>Stressor</i> phase	6.22 (0.61)
End of <i>tStressed</i> <sup>b</sup> phase	4.50 (0.42)

<sup>a</sup>*tRelaxed*: relaxed task.<sup>b</sup>*tStressed*: stressed task.

In addition to SRS, we found that the perceived *tension* was significantly higher *during Stressor* (mean 0.52, SE 0.05) versus *Relaxation* (mean 0.09, SE 0.02) ( $Z_{17}=3.59$ ,  $P<.001$ ) phases and *during tStressed* (mean 0.78, SE 0.03) versus *tRelaxed* (mean 0.29, SE 0.04) ( $t_{17}=11.79$ ,  $P<.001$ ) phases. Furthermore, perceived *tension* and SRS were highly correlated ( $r=.74$ ,  $P<.001$ ).

No statistical differences ( $t_{17}=0.58$ ,  $P=.29$ ) were observed for *trial completion time* between *tRelaxed* (mean 2.04, SE 0.14 seconds) and *tStressed* (mean 2.10, SE 0.13 seconds). Additionally, we found no statistically significant differences for perceived *concentration* *during Stressor* versus *Relaxation* phases ( $Z_{17}=1.46$ ,  $P=.07$ ), and *tStressed* versus *tRelaxed* phases ( $Z_{17}=0.56$ ,  $P=.28$ ). This shows that our stressor elicited an affective response as opposed to a response due to changes in cognitive performance.

### Physiological Stress

All metrics for EDA were significantly higher. Mean phasic EDA *during tStressed* (mean 0.040, SE 0.015 S) compared to *tRelaxed* (mean 0.020, SE 0.006 S) was higher ( $Z_{14}=2.33$ ,  $P<.01$ ); mean tonic EDA *during tStressed* (mean 4.223, SE 1.389 S) compared to *tRelaxed* (mean 2.057, SE 0.806 S) was

higher ( $Z_{14}=3.181$ ,  $P<.001$ ); and the number of EDA *peaks* (ie, SCR events) *during tStressed* (mean 68.2, SE 16.9) compared to *tRelaxed* (mean 40.4, SE 11.7) was higher as well ( $Z_{14}=2.86$ ,  $P<.01$ ).

Overall, both SRS and EDA evaluations showed that our stressor was effective in inducing acute stress, as planned.

### Trackpad Interactions

#### Mixed Task Models

Both the mean and SD of contact area under the finger were significantly different between *tRelaxed* and *tStressed* phases with large effect sizes (Cohen  $d>0.8$ ) and large post hoc power ( $\beta>97\%$ ) (see Table 3). The mean and SD of velocity and acceleration, as well as the two LPC-approximated biomechanical MSD parameters ( $\Gamma$  and  $\omega$ ), were not significantly different. Paired *t* tests were carried out for all these measures except for  $\Gamma$ , for which a Wilcoxon signed-rank test was carried out. For mean contact area, the *Order*  $\times$  *Condition* interaction effect was not significant ( $F_{1,32}=1.87$ ,  $P=.18$ ). However, for the SD of contact area, the interaction effect was significant ( $F_{1,32}=11.17$ ,  $P<.01$ ), suggesting that the *Order* of the conditions did have a significant effect on the overall SD of the contact area across the two *Conditions*.



**Table 3.** Summary of descriptive statistics between *tRelaxed* and *tStressed* phases for the Mixed Task model.

Measure	<i>tRelaxed</i> <sup>a</sup> , mean (SE)	<i>tStressed</i> <sup>b</sup> , mean (SE)	<i>t</i> test (17)	Z test (17)	<i>P</i> value	Cohen <i>d</i>
<b>Normalized velocity</b>						
Mean	0.47 (0.16)	0.49 (0.12)	0.50	N/A <sup>c</sup>	.31	0.15
SD	0.39 (0.14)	0.47 (0.20)	1.42	N/A	.09	0.48
<b>Normalized acceleration</b>						
Mean	0.35 (0.15)	0.43 (0.18)	1.36	N/A	.10	0.49
SD	0.31 (0.17)	0.42 (0.17)	1.54	N/A	.07	0.64
<b>Normalized area</b>						
Mean	0.40 (0.18)	0.55 (0.16)	2.18	N/A	.02 <sup>d</sup>	0.87
SD	0.39 (0.17)	0.53 (0.13)	2.08	N/A	.03	0.92
<b>Mass-spring-damper model</b>						
$\Gamma$ (damping ratio)	0.538 (0.007)	0.540 (0.006)	N/A	0.11	.46	0.09
$\omega$ (damping frequency) (rad/s)	0.256 (0.006)	0.261 (0.005)	1.21	N/A	.12	0.17

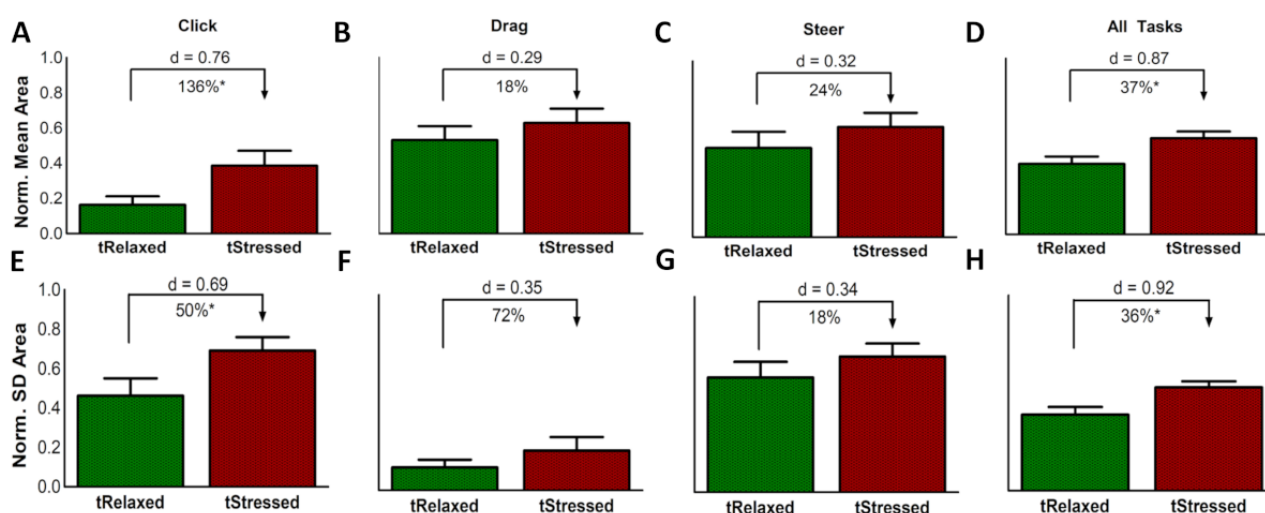
<sup>a</sup>*tRelaxed*: relaxed task.<sup>b</sup>*tStressed*: stressed task.<sup>c</sup>N/A: not applicable.<sup>d</sup>Italicized values indicate significance ( $P < .05$ ).

### Task-Specific Models

Task-Specific models for mean and SD of the contact area under the finger (see Figure 9) were performed for all three task types. Significant differences ( $P < .02$ ) between *tStressed* and *tRelaxed*

were observed for *click* tasks only, both for mean area ( $Z_{17}=2.37$ ,  $P=.009$ ,  $\beta=91\%$ ) and for SD of the area ( $Z_{17}=2.33$ ,  $P=.01$ ,  $\beta=86\%$ ). Overall, *click* tasks (see Figure 9, A and E) show large percentage differences between *tRelaxed* and *tStressed* phases.

**Figure 9.** Bar plots (Mean  $\pm$  SE), effect size (Cohen *d*), and the percentage difference between *tRelaxed* (relaxed task) and *tStressed* (stressed task) phases of (A, E) *click*, (B, F) *drag*, (C, G) *steer* tasks, and overall combining (D, H) *All Tasks* (Mixed Task model) for both mean and SD of contact area, respectively. (\*) for click indicates  $P < .05/3 = .017$  (Bonferroni correction), and (\*) for *All Tasks* indicates  $P < .05$ .



The interaction effects for mean and SD of contact area during *click* tasks were not significant (mean area:  $F_{1,32}=0.00$ ,  $P=.98$ ; SD area:  $F_{1,32}=1.39$ ,  $P=.25$ ). We also evaluated the interaction effects for SD of the contact area for *drag* and *drop* ( $F_{1,32}=1.82$ ,  $P=.19$ ) and *steer* ( $F_{1,32}=4.48$ ,  $P=.04$ ), the latter of which was significant, suggesting that the significant interaction in SD of the contact area variable in the Mixed Task model (see the Mixed Task Models section) is mainly due to significant interaction

effect only in *steer* task trials. Overall, our results suggest that finger dynamics from *click* tasks alone, which are the most common computer interactions (see Discussion section), and are unaffected by *Order* effects, can be used as robust measurements to detect acute stress.



## Sensitivity Analyses

### Data Reduction

In this section, we focus on *click* data, which was the only task type contributing to the overall differences in the *Mixed Task* model. We started looking for data closest to the stressor (ie, the *initial* 10% [4 *clicks* out of 40]), and we intended to explore bigger segments (ie, 20%, 30%, etc) in case we did not find a signal sufficient to detect acute stress across the two conditions.

To our surprise, with only 4 *clicks*, we could see a significant difference across the two phases (see [Table 4](#)). This result may be because the effect of the stressor is higher earlier during the *tStressed* phase, and it decays over time (see [Table 2](#)). Similar sensitivity results have been observed in other biomechanical models of stress [5]. Neither of the two variables mentioned in [Table 4](#) showed any significant interaction effect (mean area:  $F_{1,32}=0.00$ ,  $P=.99$ ; SD area:  $F_{1,32}=0.50$ ,  $P=.49$ ), which reconfirms that *click* tasks are not influenced by *Order* effects.

**Table 4.** Summary of descriptive statistics between *tRelaxed* and *tStressed* phases pertaining to only the initial 10% of click trials.

Measure	No. of <i>clicks</i> (N=40), n (%)	<i>tRelaxed</i> <sup>a</sup> , mean (SE)	<i>tStressed</i> <sup>b</sup> , mean (SE)	<i>t</i> test (17)	<i>P</i> value	Cohen <i>d</i>	$\beta$ , %
Mean area (mm <sup>2</sup> )	4 (10)	270.9 (6.3)	284.6 (8.1)	2.81	<.01 <sup>c</sup>	0.44	56
SD area (mm <sup>2</sup> )	4 (10)	16.7 (1.6)	21.2 (1.3)	2.88	<.01	0.73	91

<sup>a</sup>*tRelaxed*: relaxed task.

<sup>b</sup>*tStressed*: stressed task.

<sup>c</sup>Italicized values indicate significance ( $P<.05$ ).

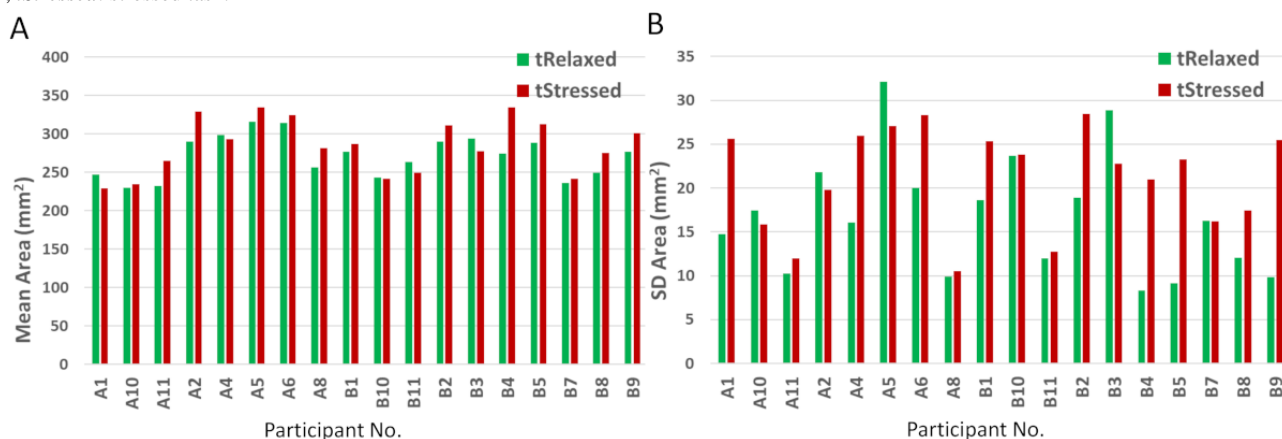
### Click Rate Manipulation

Our original manipulation generated *click* trials at a rate of roughly 1 *click* per 3.5 seconds: 40 *click* trials were completed in an average of 136 seconds (SD 31). To create a scenario in which the *clicks* are obtained at a slower rate, which could also be plausible in the wild, we tested the artificially slowest rate by decimating the samples, using only 10% of the data distributed across the entire range of the production of the *click* trials. We picked the first, the last, and 2 random *click* trials from 1 to 40, in this way simulating a scenario in which *clicks* are generated once every ~45 seconds. As it turns out, both area measures were still significantly different across two stress conditions. The mean of the mean area for *tStressed* was 277 mm<sup>2</sup> (SE 8.61) and for *tRelaxed* was 267 mm<sup>2</sup> (SE 6.4) ( $t_{17}=2.07$ ,  $P=.03$ ,  $d=0.32$ ,  $\beta=37\%$ ). The mean of the SD area for *tStressed* was 22 mm<sup>2</sup> (SE 1.8) and for *tRelaxed* was 18 mm<sup>2</sup> (SE 1.5) ( $t_{17}=1.86$ ,  $P=.04$ ,  $d=0.54$ ,  $\beta=71\%$ ). Further, no significant interaction effect was found for either of these two measures as well (mean area:  $F_{1,32}=0.14$ ,  $P=.72$ ; SD area:  $F_{1,32}=1.98$ ,  $P=.17$ ).

### Individual Differences

When looking at individual differences for all tasks, 12 out of 18 (67%) participants showed an expected increase in either the mean or the SD of the area under the finger. For the initial 10% (ie, 4 out of 40 *clicks*), 13 out of 18 (72%) subjects had changes in the expected direction for either mean or SD (see [Figure 10](#)). Participants A1, A4, B10, B1, and B3 showed an opposite trend for the mean area (see [Figure 10, A](#)), while participants A10, A2, A5, B3, and B7 showed an opposite trend for SD area (see [Figure 10, B](#)). Subject B3 showed an opposite trend for both measures. If we were to consider a hard evaluation where both metrics must occur at the same time, we would see 11 out of 18 (61%) subjects satisfying the expected change. However, if we were to detect stress whenever either of the two measures has the right trend, we would detect stress changes in 17 out of 18 (94%) subjects. This latter approach is plausible, as both measurements are theoretically linked to changes in either muscle tension or motor control alterations due to stress (see section RQ2: Measurements of Finger Dynamics From Trackpad Interactions).

**Figure 10.** Individual differences for (A) mean and (B) SD of the contact area under the finger across the initial 10% of the *click* trials. *tRelaxed*: relaxed task; *tStressed*: stressed task.



## Discussion

### Principal Findings

Enabling precision health and effective stress management regimes are dependent on the evolution of ubiquitous real-time stress sensors and algorithms. An ideal scenario would be to passively obtain data from everyday use devices, rather than having people wear sensors or answer survey questions. In this study, we aim at assessing the feasibility of using passively obtained laptop trackpad data to detect acute stress. We present evidence that binary levels of acute stress (ie, whether it is present or not) can be detected from finger strokes alone.

In response to RQ1, we found out that our approximation to an MSD model, previously used with data from the computer mouse [4], cannot be used directly on trackpad data. Our analyses provide new insights into the complexity of the use of finger stroke data from trackpad operations for applying an MSD model. Unlike a computer mouse, handling a laptop trackpad is subject dependent. Some people may employ multiple hands and/or multiple fingers to perform the same tasks. This creates challenges to process the data due to the gaps present in the strokes and the difficulty of obtaining a clean intentional segment of data, with a clear *beginning* and *end*, required to get meaningful MSD results. These limitations leave little useful data for analysis, and we did not observe a difference in either of the two MSD parameters ( $\omega$  and  $\Gamma$ ) between the *tRelaxed* versus *tStressed* phases. However, our analyses were restricted to the largest stroke per trial. It is possible that in the wild, where we can collect larger amounts of data that would allow us to extract valid displacement segments, an MSD approach may still be useful to assess changes due to muscle tension under acute stress.

Our RQ2's results are encouraging, showing the robustness of the mean and SD of the finger contact area as a stress sensor. Both metrics were higher during the *stressed* tasks as compared to the *relaxed* tasks. We show that as few as 4 *click* trials per subject are sufficient to see a significant difference, which is promising for longitudinal studies. Overall, we present evidence of the use of the finger contact area from laptop trackpad measurements to detect acute stress while performing common computer tasks.

### Acute Stress Detected Using Click Finger Dynamics

This is the first study showing that the mean and SD of the contact area under the finger on a trackpad are higher under stress compared to a relaxed state (ie, *Mixed Task* model). Theoretically, the mean value is related to mean force production [19] and higher variability [17] under stressful situations in the wrist and forearm muscles. There may be other underlying mechanisms (eg, changes in hand or body posture [40]) through which stress may influence finger dynamics, and future studies should monitor those variables as well. In the *Task-Specific* model, differences in mean and SD of the contact area under the finger were different for *click* trials. If combined, mean *or* SD, we could detect up to 94% of acute stress events. This is encouraging because *clicking* tasks represent nearly 70% of all computer mouse events in a typical day [41].

With just the initial 10% of *clicking* trials (ie, 4 *clicks* out of 40), we could still see small and medium effect size differences in mean and SD of the contact area under the finger. Since *click* trials were not always carried out right after the stressor, the difference is not due to ordering effects; here we are referring to ordering of *click*, *steer*, and *drag and drop*, and not the ordering of *stress-relax* conditions. We attribute the strength of using so little data to the proximity to the stressor, as there is a clear decay of stress over time, which would clearly affect values averaged over longer periods. Furthermore, we simulated a scenario where *clicks* are produced at a different rate (ie, going from 1 *click* produced every ~3.5 seconds to every ~45 seconds). We obtained significant results even for the lowest *click* productivity we could simulate with our data (1 *click*/45 s). In a separate pilot study, collecting computer mouse data “in the wild,” we have seen that a typical user may generate a *click* every 10-30 seconds, and at least 1000 *clicks* are generated in any given working day. Thus, we feel confident we can use our findings as the basis for a passive, continuous, scalable, inexpensive, and unobtrusive stress sensor. In the wild, of course, changes in muscle tension would not only be due to affective (ie, ANS) processes but also due to cognitive processes (see Potential Interaction Effects With Task Performance section). Ultimately, a reliable “in the wild” stress detection system would have to combine multiple modalities, such as biomechanical, behavioral, and physiological sensors, as well as contextual information.

### Potential Interaction Effects With Task Performance

Mean and SD of finger velocity and acceleration were similar across the two phases, as well as *trial completion time*. This reconfirms that our stressor did not change the cognitive performance of the subjects but only their affective state. However, the relationships between stress, velocity, and acceleration may not be simple. For example, one may say that under a stressful situation, a user may want to move the fingers with higher velocity and/or acceleration; however, after a certain point, it will also start affecting the accuracy of this fine motor control task [42]. Thus, the user may not use higher velocities and/or acceleration under stress. Further studies with different elicitations to modify the performance rewards must be performed to observe interaction effects between these variables.

### Contrasting With Other Touch Technologies

At least two studies have shown higher finger pressure under stress [28] and under frustrating [7] conditions while using touchscreen devices. However, the usage and handling of touchscreen devices are not the same as that of a trackpad, in terms of fine and precise interaction [43]. The capacitive technologies used in touchscreens and trackpads are different in terms of how much pressure needs to be applied: typically, the trackpad requires stronger pressure [43]. Additional research would be needed to determine if our finger dynamics from contact area (ie, mean and SD) can be translated to other types of touch-sensitive devices.

### Limitations

Our study has four limitations related to sample size, stressor effect size, apparatus, and assumptions for modeling. First,

because of the small sample size ( $N=18$ ) we could not investigate the effects of age, gender, length of experience with trackpads, and frequency of day-to-day usage of trackpads. We also could not build reliable predictive models (eg, using different ML techniques), which could parse out the importance of various features in predicting acute stress. Second, given the mild nature of the stressor, not all users (18/22, 82%) responded to our manipulation. Nevertheless, this level of efficacy is expected for a mild stressor and in line with our prior study [6]. Third, although we do not expect the behavior of users to be fundamentally different across different laptops, our experiment should be revalidated on other models of trackpads. Fourth, using finger contact area as a proxy to pressure assumes a linear relationship between the two. However, factors like the angle of the finger can also influence this relationship. Since the time of the data collection for this study, Apple has updated the *MultiTouchSupport* framework, and it now also provides information about pressure and capacitance density underneath the fingers. Additional models collecting this new data would be quite useful to strengthen the precision of our detection method further.

### Future Directions

We plan to expand current results by running lab experiments with larger sample sizes; by logging other actions performed using trackpads, such as scrolling and multiple finger gestures; and by collecting other behavioral metrics, such as *click* production rate, the timing between *clicks*, etc. Once we collect more data, both in the laboratory settings and “in the wild,” we aim at building predictive models that include both biomechanical and behavioral metrics to increase the accuracy and specificity of stress prediction in real time.

As we move out of the lab, we will explore ways to carry out longitudinal studies collecting data from our passive loggers across devices (ie, mice, trackpads, touchpads), contextual data, and stress “labels” obtained from empirical sampling methods or physiological measurements. As we accumulate richer and bigger datasets, we plan to investigate data-intensive unsupervised and supervised ML methods to optimize data collection, preprocessing, and prediction. For example, we plan to investigate algorithms to parse types of events and movements (eg, *clicks*, *drag and drop*, etc) or to select segments that may have the right morphology to derive biomechanical features (eg, MSD parameters, contact area, etc). Ultimately, we hope that our lab and field findings can help select the appropriate behavioral, biomechanical, contextual, and perceptual features (ie, feature engineering) to train ML systems to be able to detect stress levels throughout the day.

### Conclusions

The current methods for sensing stress are often inconvenient, expensive, or suffer from limited adherence. In this lab study, we show the efficacy of repurposing signals from a laptop trackpad to detect acute mental stress. We showed that with a handful of *click* events, we could use the mean and SD of the finger contact area to obtain a binary estimation of acute stress (ie, whether it is present or not). We validated our sensing models against well-known stress metrics such as self-reports and EDA. Complementarily, we validated that we did not elicit performance or cognitive confounds that may affect our affective stress elicitation. These results provide a firm baseline toward our future goal of deploying our algorithms in the wild, leveraging the trackpad as a potential unobtrusive, scalable, continuous, and inexpensive stress sensor.

### Acknowledgments

We appreciate the financial support provided by the late Prof Sam Gambhir, the Precision Health and Integrated Diagnostics Center, Prof Sarah L Billington, and Prof James A Landay, as well as the advice provided by Dr Jean Costa. [Figure 2](#) was licensed from QA International, science photo library, and labels were added to the four muscles.

### Conflicts of Interest

None declared.

### References

1. Epel ES, Blackburn EH, Lin J, Dhabhar FS, Adler NE, Morrow JD, et al. Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci U S A* 2004 Dec 07;101(49):17312-17315. [doi: [10.1073/pnas.0407162101](https://doi.org/10.1073/pnas.0407162101)] [Medline: [15574496](https://pubmed.ncbi.nlm.nih.gov/15574496/)]
2. Lazarus RS. *Psychological Stress and the Coping Process*. New York, NY: McGraw-Hill; 1966.
3. Tomaka J, Blascovich J, Kelsey RM, Leitten CL. Subjective, physiological, and behavioral effects of threat and challenge appraisal. *J Pers Soc Psychol* 1993;65(2):248-260. [doi: [10.1037/0022-3514.65.2.248](https://doi.org/10.1037/0022-3514.65.2.248)]
4. Sun D, Paredes PE, Canny J. MouStress: Detecting stress from mouse motion. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM; 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, Canada p. 61-70. [doi: [10.1145/2556288.2557243](https://doi.org/10.1145/2556288.2557243)]
5. Paredes PE, Ordóñez F, Ju W, Landay J. Fast & Furious: Detecting Stress with a Car Steering Wheel. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM; 2018 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, Canada p. 1-12 URL: <https://dl.acm.org/doi/10.1145/3173574.3174239>

6. Hernandez J, Paredes PE, Roseway A, Czerwinski M. Under pressure: Sensing stress of computer users. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.; ACM; 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, Canada p. 51-60. [doi: [10.1145/2556288.2557165](https://doi.org/10.1145/2556288.2557165)]
7. Gao Y, Bianchi-Berthouze N, Meng H. What does touch tell us about emotions in touchscreen-based gameplay? ACM Trans Comput Hum Interact 2012 Dec 01;19(4):1-30. [doi: [10.1145/2395131.2395138](https://doi.org/10.1145/2395131.2395138)]
8. Exposito M, Hernandez J, Picard RW. Affective keys: Towards unobtrusive stress sensing of smartphone users. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services.; ACM; 2018 Presented at: 20th International Conference on Human-Computer Interaction with Mobile Devices and Services; September 3-6, 2018; Barcelona, Spain p. 139-145. [doi: [10.1145/3236112.3236132](https://doi.org/10.1145/3236112.3236132)]
9. Tankovska H. Global shipments of tablets, notebook and desktop PCs 2017-2022. Statista. 2020 Aug 27. URL: <https://www.statista.com/statistics/866044/global-shipments-forecast-for-tablets-notebook-and-desktop-pcs/> [accessed 2020-04-22]
10. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: [6668417](https://pubmed.ncbi.nlm.nih.gov/6668417/)]
11. Roberti J, Harrington L, Storch E. Further psychometric support for the 10-item version of the perceived stress scale. J Coll Couns 2006;9(2):135-147. [doi: [10.1002/j.2161-1882.2006.tb00100.x](https://doi.org/10.1002/j.2161-1882.2006.tb00100.x)]
12. Balters S, Mauriello ML, Park SY, Landay JA, Paredes PE. Calm commute: Guided slow breathing for daily stress management in drivers. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar 18;4(1):1-19. [doi: [10.1145/3380998](https://doi.org/10.1145/3380998)]
13. Sarker H, Tyburski M, Rahman MM, Hovsepian K, Sharmin M, Epstein DH, et al. Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2016 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 4489-4501 URL: <https://doi.org/10.1145/2858036.2858218> [doi: [10.1145/2858036.2858218](https://doi.org/10.1145/2858036.2858218)]
14. Kim H, Cheon E, Bai D, Lee YH, Koo B. Stress and heart rate variability: A meta-analysis and review of the literature. Psychiatry Investig 2018 Mar;15(3):235-245 [FREE Full text] [doi: [10.30773/pi.2017.08.17](https://doi.org/10.30773/pi.2017.08.17)] [Medline: [29486547](https://pubmed.ncbi.nlm.nih.gov/29486547/)]
15. Setz C, Arnrich B, Schumm J, La Marca R, Tröster G, Ehlert U. Discriminating stress from cognitive load using a wearable EDA device. IEEE Trans Inf Technol Biomed 2010 Mar;14(2):410-417. [doi: [10.1109/TITB.2009.2036164](https://doi.org/10.1109/TITB.2009.2036164)] [Medline: [19906598](https://pubmed.ncbi.nlm.nih.gov/19906598/)]
16. Boucsein W. Electrodermal Activity. New York, NY: Springer; 1992.
17. Noteboom JT, Fleshner M, Enoka RM. Activation of the arousal response can impair performance on a simple motor task. J Appl Physiol (1985) 2001 Aug;91(2):821-831. [doi: [10.1152/jappl.2001.91.2.821](https://doi.org/10.1152/jappl.2001.91.2.821)] [Medline: [11457799](https://pubmed.ncbi.nlm.nih.gov/11457799/)]
18. Laidlaw DH, Kornatz KW, Keen DA, Suzuki S, Enoka RM. Strength training improves the steadiness of slow lengthening contractions performed by old adults. J Appl Physiol (1985) 1999 Nov;87(5):1786-1795. [doi: [10.1152/jappl.1999.87.5.1786](https://doi.org/10.1152/jappl.1999.87.5.1786)] [Medline: [10562623](https://pubmed.ncbi.nlm.nih.gov/10562623/)]
19. Coombes SA, Cauraugh JH, Janelle CM. Emotion and movement: Activation of defensive circuitry alters the magnitude of a sustained muscle contraction. Neurosci Lett 2006 Apr 03;396(3):192-196. [doi: [10.1016/j.neulet.2005.11.048](https://doi.org/10.1016/j.neulet.2005.11.048)] [Medline: [16376016](https://pubmed.ncbi.nlm.nih.gov/16376016/)]
20. Arnrich B, Setz C, La Marca R, Tröster G, Ehlert U. What does your chair know about your stress level? IEEE Trans Inf Technol Biomed 2010 Mar;14(2):207-214. [doi: [10.1109/TITB.2009.2035498](https://doi.org/10.1109/TITB.2009.2035498)] [Medline: [19887325](https://pubmed.ncbi.nlm.nih.gov/19887325/)]
21. Davis M, Robbins Eshelman E, McKay M. The Relaxation and Stress Reduction Workbook. 7th edition. Oakland, CA: New Harbinger Publications; 2019.
22. Cacioppo JT, Tassinary LG. Inferring psychological significance from physiological signals. Am Psychol 1990 Jan;45(1):16-28. [doi: [10.1037//0003-066x.45.1.16](https://doi.org/10.1037//0003-066x.45.1.16)] [Medline: [2297166](https://pubmed.ncbi.nlm.nih.gov/2297166/)]
23. Sainsbury P, Gibson JG. Symptoms of anxiety and tension and the accompanying physiological changes in the muscular system. J Neurol Neurosurg Psychiatry 1954 Aug;17(3):216-224 [FREE Full text] [doi: [10.1136/jnnp.17.3.216](https://doi.org/10.1136/jnnp.17.3.216)] [Medline: [13192498](https://pubmed.ncbi.nlm.nih.gov/13192498/)]
24. Hoehn-Saric R, McLeod DR, Funderburk F, Kowalski P. Somatic symptoms and physiologic responses in generalized anxiety disorder and panic disorder: An ambulatory monitor study. Arch Gen Psychiatry 2004 Sep;61(9):913-921. [doi: [10.1001/archpsyc.61.9.913](https://doi.org/10.1001/archpsyc.61.9.913)] [Medline: [15351770](https://pubmed.ncbi.nlm.nih.gov/15351770/)]
25. Lundberg U, Kadefors R, Melin B, Palmerud G, Hassmen P, Engstrom M, et al. Psychophysiological stress and EMG activity of the trapezius muscle. Int J Behav Med 1994;1(4):354-370. [doi: [10.1207/s15327558ijbm0104\\_5](https://doi.org/10.1207/s15327558ijbm0104_5)] [Medline: [16250795](https://pubmed.ncbi.nlm.nih.gov/16250795/)]
26. van Loon EM, Masters RS, Ring C, McIntyre DB. Changes in limb stiffness under conditions of mental stress. J Mot Behav 2001 Jun;33(2):153-164. [doi: [10.1080/00222890109603147](https://doi.org/10.1080/00222890109603147)] [Medline: [11404211](https://pubmed.ncbi.nlm.nih.gov/11404211/)]
27. Wahlström J, Hagberg M, Johnson PW, Svensson J, Rempel D. Influence of time pressure and verbal provocation on physiological and psychological reactions during work with a computer mouse. Eur J Appl Physiol 2002 Jul;87(3):257-263. [doi: [10.1007/s00421-002-0611-7](https://doi.org/10.1007/s00421-002-0611-7)] [Medline: [12111287](https://pubmed.ncbi.nlm.nih.gov/12111287/)]
28. Carneiro D, Castillo JC, Novais P, Fernández-Caballero A, Neves J. Multimodal behavioral analysis for non-invasive stress detection. Expert Syst Appl 2012 Dec;39(18):13376-13389. [doi: [10.1016/j.eswa.2012.05.065](https://doi.org/10.1016/j.eswa.2012.05.065)]



29. Conte C, Ranavolo A, Serrao M, Silvetti A, Orengo G, Mari S, et al. Kinematic and electromyographic differences between mouse and touchpad use on laptop computers. *Int J Ind Ergon* 2014 May;44(3):413-420. [doi: [10.1016/j.ergon.2014.01.001](https://doi.org/10.1016/j.ergon.2014.01.001)]
30. Dennerlein JT, Johnson PW. Different computer tasks affect the exposure of the upper extremity to biomechanical risk factors. *Ergonomics* 2006 Jan 15;49(1):45-61. [doi: [10.1080/00140130500321845](https://doi.org/10.1080/00140130500321845)] [Medline: [16393803](https://pubmed.ncbi.nlm.nih.gov/16393803/)]
31. Piferi RL, Kline KA, Younger J, Lawler KA. An alternative approach for achieving cardiovascular baseline: Viewing an aquatic video. *Int J Psychophysiol* 2000 Aug;37(2):207-217. [doi: [10.1016/s0167-8760\(00\)00102-1](https://doi.org/10.1016/s0167-8760(00)00102-1)]
32. Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'--A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 1993;28(1-2):76-81. [doi: [10.1159/000119004](https://doi.org/10.1159/000119004)] [Medline: [8255414](https://pubmed.ncbi.nlm.nih.gov/8255414/)]
33. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Front Psychol* 2013 Nov 26;4:863 [FREE Full text] [doi: [10.3389/fpsyg.2013.00863](https://doi.org/10.3389/fpsyg.2013.00863)] [Medline: [24324449](https://pubmed.ncbi.nlm.nih.gov/24324449/)]
34. Hernandez J, Riobo I, Rozga A, Abowd G, Picard R. Using electrodermal activity to recognize ease of engagement in children during social interactions. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, WA. [doi: [10.1145/2632048.2636065](https://doi.org/10.1145/2632048.2636065)]
35. Taylor S, Jaques N, Chen W, Fedor S, Sano A, Picard RW. Automatic identification of artifacts in electrodermal activity data. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015 Presented at: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 25-29, 2015; Milan, Italy p. 1934-1937 URL: <http://europepmc.org/abstract/MED/26736662> [doi: [10.1109/EMBC.2015.7318762](https://doi.org/10.1109/EMBC.2015.7318762)]
36. Topogulu Y, Watson J, Suri R, Ayaz H. Electrodermal activity in ambulatory settings: A narrative review of literature. In: Proceedings of the AHFE 2019 International Conference on Neuroergonomics and Cognitive Engineering, and the AHFE International Conference on Industrial Cognitive Ergonomics and Engineering Psychology. Cham, Switzerland: Springer Nature; 2019 Presented at: AHFE 2019 International Conference on Neuroergonomics and Cognitive Engineering, and the AHFE International Conference on Industrial Cognitive Ergonomics and Engineering Psychology; July 24-28, 2019; Washington, DC p. 91-102. [doi: [10.1007/978-3-030-20473-0\\_10](https://doi.org/10.1007/978-3-030-20473-0_10)]
37. Benedek M, Kaernbach C. A continuous measure of phasic electrodermal activity. *J Neurosci Methods* 2010 Jun 30;190(1):80-91 [FREE Full text] [doi: [10.1016/j.jneumeth.2010.04.028](https://doi.org/10.1016/j.jneumeth.2010.04.028)] [Medline: [20451556](https://pubmed.ncbi.nlm.nih.gov/20451556/)]
38. O'Shaughnessy D. Linear predictive coding. *IEEE Potentials* 1988 Feb;7(1):29-32. [doi: [10.1109/45.1890](https://doi.org/10.1109/45.1890)]
39. Guo W, Wang J. Understanding mobile reading via camera based gaze tracking kinematic touch modeling. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI). 2018 Presented at: 20th ACM International Conference on Multimodal Interaction (ICMI); October 16-20, 2018; Boulder, CO. [doi: [10.1145/3242969.3243011](https://doi.org/10.1145/3242969.3243011)]
40. Qin J, Chen H, Dennerlein JT. Wrist posture affects hand and forearm muscle stress during tapping. *Appl Ergon* 2013 Nov;44(6):969-976. [doi: [10.1016/j.apergo.2013.03.013](https://doi.org/10.1016/j.apergo.2013.03.013)] [Medline: [23591089](https://pubmed.ncbi.nlm.nih.gov/23591089/)]
41. Antal M, Egyed-Zsigmond E. Intrusion detection using mouse dynamics. *IET Biometrics* 2019;8(5):285-294. [doi: [10.1049/iet-bmt.2018.5126](https://doi.org/10.1049/iet-bmt.2018.5126)]
42. Kowatsch T, Wahle F, Filler A. stressOUT: Design, implementation and evaluation of a mouse-based stress management service. In: Proceedings of the 12th International Conference on Design Science Research in Information Systems and Technology (DESIST 2017). Karlsruhe, Germany: Karlsruher Institut für Technologie; 2017 Presented at: 12th International Conference on Design Science Research in Information Systems and Technology (DESIST 2017); May 30-June 1, 2017; Karlsruhe, Germany p. 37-45. [doi: [10.3929/ethz-b-000218657](https://doi.org/10.3929/ethz-b-000218657)]
43. Maleckar A, Kljun M, Rogelj P, Pucihar K. Evaluation of common input devices for web browsing: Mouse vs touchpad vs touchscreen. In: Proceedings of the Human-Computer Interaction in Information Society. 2016 Presented at: Human-Computer Interaction in Information Society; October 9-13, 2016; Ljubljana, Slovenia URL: [https://hci.si/wp-content/uploads/2016/10/HCI-IS\\_2016\\_paper\\_6.pdf](https://hci.si/wp-content/uploads/2016/10/HCI-IS_2016_paper_6.pdf)

## Abbreviations

**ANS:** autonomic nervous system  
**β:** post hoc power  
**D:** distance  
**ECG:** electrocardiogram  
**ED:** extensor digitorum  
**EDA:** electrodermal activity  
**EMG:** electromyography  
**ESM:** empirical sampling methods  
**FDI:** first dorsal interosseous  
**Γ:** damping ratio  
**HRV:** heart rate variability  
**LPC:** linear predictive coding



**ML:** machine learning  
**MSD:** mass spring damper  
 $\omega$ : damping frequency  
**PCD:** personal computing device  
**PSS:** Perceived Stress Scale  
**RQ:** research question  
**SCR:** skin conductance response  
**SNS:** sympathetic nervous system  
**SR:** self-report marker  
**SRS:** self-reported stress  
**tRelaxed:** relaxed task  
**tStressed:** stressed task  
**W:** width

*Edited by G Eysenbach; submitted 22.07.20; peer-reviewed by P Hsiao, P Jimenez; comments to author 08.08.20; revised version received 20.08.20; accepted 07.09.20; published 23.10.20.*

Please cite as:

Goel R, An M, Alayrangues H, Koneshloo A, Lincoln ET, Paredes PE  
Stress Tracker—Detecting Acute Stress From a Trackpad: Controlled Study  
J Med Internet Res 2020;22(10):e22743  
URL: <http://www.jmir.org/2020/10/e22743/>  
doi: [10.2196/22743](https://doi.org/10.2196/22743)  
PMID: [33095176](https://pubmed.ncbi.nlm.nih.gov/33095176/)

©Rahul Goel, Michael An, Hugo Alayrangues, Amirhossein Koneshloo, Emmanuel Thierry Lincoln, Pablo Enrique Paredes. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Influenza Screening via Deep Learning Using a Combination of Epidemiological and Patient-Generated Health Data: Development and Validation Study

Hyunwoo Choo<sup>1</sup>, BSc; Myeongchan Kim<sup>2</sup>, MD; Jiyun Choi<sup>2</sup>; Jaewon Shin<sup>2</sup>, MD; Soo-Yong Shin<sup>1,3</sup>, PhD

<sup>1</sup>Department of Digital Health, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea

<sup>2</sup>Mobile Doctor, Seoul, Republic of Korea

<sup>3</sup>Big Data Research Center, Samsung Medical Center, Seoul, Republic of Korea

**Corresponding Author:**

Soo-Yong Shin, PhD

Department of Digital Health

Samsung Advanced Institute for Health Sciences and Technology

Sungkyunkwan University

115, Irwon-ro, Gangnam-gu

Seoul

Republic of Korea

Phone: 82 2 3410 1449

Email: [sy.shin@skku.edu](mailto:sy.shin@skku.edu)

## Abstract

**Background:** Screening for influenza in primary care is challenging due to the low sensitivity of rapid antigen tests and the lack of proper screening tests.

**Objective:** The aim of this study was to develop a machine learning-based screening tool using patient-generated health data (PGHD) obtained from a mobile health (mHealth) app.

**Methods:** We trained a deep learning model based on a gated recurrent unit to screen influenza using PGHD, including each patient's fever pattern and drug administration records. We used meteorological data and app-based surveillance of the weekly number of patients with influenza. We defined a single episode as the set of consecutive days, including the day the user was diagnosed with influenza or another disease. Any record a user entered 24 hours after his or her last record was considered to be the start of a new episode. Each episode contained data on the user's age, gender, weight, and at least one body temperature record. The total number of episodes was 6657. Of these, there were 3326 episodes within which influenza was diagnosed. We divided these episodes into 80% training sets (2664/3330) and 20% test sets (666/3330). A 5-fold cross-validation was used on the training set.

**Results:** We achieved reliable performance with an accuracy of 82%, a sensitivity of 84%, and a specificity of 80% in the test set. After the effect of each input variable was evaluated, app-based surveillance was observed to be the most influential variable. The correlation between the duration of input data and performance was not statistically significant ( $P=.09$ ).

**Conclusions:** These findings suggest that PGHD from an mHealth app could be a complementary tool for influenza screening. In addition, PGHD, along with traditional clinical data, could be used to improve health conditions.

(*J Med Internet Res* 2020;22(10):e21369) doi:[10.2196/21369](https://doi.org/10.2196/21369)

**KEYWORDS**

influenza; screening tool; patient-generated health data; mobile health; mHealth; deep learning

## Introduction

With the increasing popularity of mobile health (mHealth), a considerable amount of health-related data are now generated and accumulated outside of hospitals [1-3]. These health-related

data cover a wide range of quantitative variables, such as physical activity, blood glucose levels, blood pressure, heart rate/rhythm, and oxygen saturation along with a range of qualitative data, such as mood-related symptoms, food intake, medication use, and sleep patterns. Even data from social media posts or search engine queries may be included [4]. These kinds

of health-related data are categorized as patient-generated health data (PGHD) and defined by the Office of the National Coordinator for Health Information Technology as “health-related data—including health history, symptoms, biometric data, treatment history, lifestyle choices, and other information—created, recorded, gathered, or inferred by or from patients or their designees (i.e., care partners or those who assist them) to help address a health concern” [5].

Many studies have shown that PGHD have various potential benefits for health care. For example, PGHD may help patients with chronic diseases like diabetes or hypertension take better care of themselves by delivering continuous monitoring and support with more personalized treatment planning [6-9]. PGHD are also beneficial for remote monitoring of patients’ postsurgical pain or chronic pain and have been found to more accurately assess the psychoemotional status of patients [10-12]. Another example of PGHD use is forecasting contagious diseases. Some research has shown that influenza [13-15] and Middle East respiratory syndrome (MERS) [16] outbreaks could be predicted using search engine query data, including Google Flu Trends and social media posts. In addition to these indirect methods, a website or smartphone app through which patients directly report their symptoms can also be used to detect epidemics [17,18].

Although influenza outbreaks can be predicted using PGHD, the diagnosis or screening of individual patients has been conducted using traditional medical devices, such as the rapid influenza antigen test or reverse transcription–polymerase chain reaction (RT-PCR). The rapid influenza diagnostic test (RIDT) has mainly been used as a diagnostic test because of its reduced processing time and easy accessibility [19]. However, due to the low sensitivity of the RIDT, it is insufficient to serve as a screening test for influenza [20-22]. Due to this concern, influenza treatment with antiviral medication has been prescribed for suspected influenza cases, based on clinical judgment, even when the RIDT showed a negative result. Influenza-like illness (ILI) case definition is one of the symptom-based screening methods of suspected cases, but it has been reported to have limited sensitivity despite its loss of substantial specificity [23].

Fever is regarded as the most distinctive symptom of influenza. Due to the lack of other distinguishable symptoms, it can be challenging to differentiate influenza from other diseases [24,25]. Recently, deep learning approaches have been reported to exceed classical statistical methods for predicting the outcomes of an individual patient using time series data, such as inpatient data [23,26]. In this study, we propose a deep learning method for influenza screening by combining epidemiological information and PGHD from an mHealth app. These results were then compared with the patients’ diagnostic findings.

## Methods

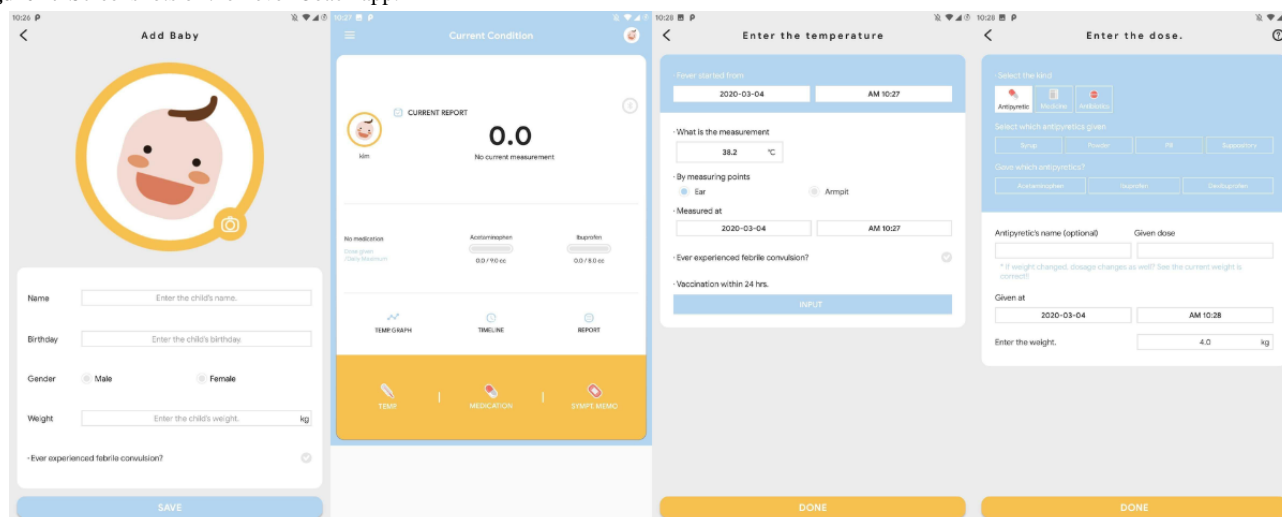
### Data Collection

We retrospectively collected log data from the Fever Coach app, which is available on Android and iOS [27]. Fever Coach is a fever management app that uses the self-reported data of its users (Figure 1).

The data were collected from January 2017 to December 2018. A total of 480,793 users entered 28,010,112 records. During the same period, the number of users diagnosed with influenza at a clinic was 16,432. In 2017 and 2018, 3583 and 12,849 users were diagnosed with influenza, respectively. The log data included body temperature, volume, type and form of antipyretic drugs or antibiotic drugs, sex, age, weight, symptoms, and memos. The users of Fever Coach agreed that their deidentified data could be used for research purposes, and the institutional review board of Samsung Medical Center waived informed consent.

We collected the daily mean temperature, daily maximum temperature, daily minimum temperature, daily mean dew point, daily mean relative humidity, and daily mean pressure data between January 2017 and December 2018 from the Korea Meteorological Administration information portal. The observation point was Seoul 108 [28].

Korea Center for Disease Control (KCDC) produces a weekly influenza-like illness report every Tuesday using data received from public health centers during the previous week. These data were collected for the period of January 2017–December 2018 [29].

**Figure 1.** Screenshots of the Fever Coach app.

## Data Preprocessing

All of the log data, separated by user ID and year, were then split into episodes. The episodes were defined as the set of consecutive days containing the day the user was diagnosed with influenza or another disease. For example, if a user was diagnosed with influenza on February 23, 2018, and recorded his or her body temperature between February 21, 2018, and February 24, 2018, these days were considered to be 1 episode. If the user logged another record 24 hours after his or her previous record, it was considered to be a new episode. [Table 1](#) shows examples of episode separation.

Each episode must contain information about the user's age, gender, and weight. Users were divided into 4 age groups—0-2 years, 2-5 years, 6-12 years, and  $\geq 13$  years—to avoid possible overfitting according to age, as age is one of the key factors of influenza propagation. Any episode without age, gender, and weight was excluded. Moreover, any episode not containing at least 1 fever data point was excluded.

We then calculated the app-based weekly influenza surveillance from the influenza-diagnosed episodes each year. The app-based weekly influenza surveillance was defined by the weekly number of reported influenza cases divided by the total number of annually reported influenza cases in the same year. For example, if there were 3000 reported influenza cases in 2018 and 300 weekly reported influenza cases in week 49 of 2018, the app-based surveillance for week 49 of 2018 was 0.1. We calculated this value every week for each year and then added

this value to the corresponding episode. If each episode had multiple days, we used the first day of each episode as the representative value, considering that the incubation period of influenza is 1 to 4 days [30,31]. Our week-numbering was based on the ISO week-date system [32]. The app-based weekly influenza surveillance data are in [Multimedia Appendix 1](#).

We also added meteorological data from the Korean Meteorological Administration. As before, we used values corresponding to the first day of each episode. We added KCDC laboratory surveillance as well, but this time we used values corresponding to 1 week before the first day of each episode. Due to the reporting delay of the KCDC surveillance, we could not use values corresponding to the same week.

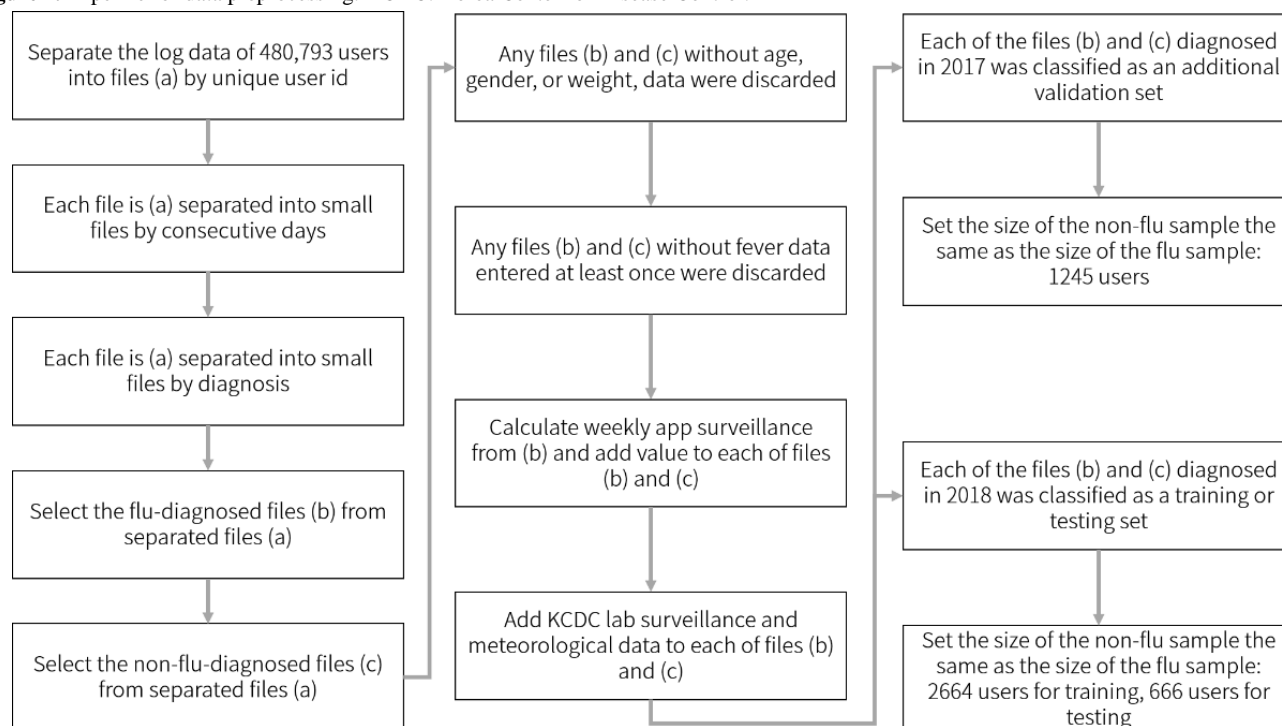
Finally, as the log data we collected had more noninfluenza episodes than the influenza episodes, we set the number of the noninfluenza episodes to be the same as the influenza episodes each year. Data from 2018 were used for training and hyperparameter tuning, and those data were randomly split into the training set (2664/3330, 80%) and the test set (666/3330, 20%). A 5-fold cross-validation was used on the training set. Considering that the influenza epidemic is slightly different each year, we prepared an additional validation set. Although our training/test sets included the data collected in 2018, the additional validation set included the data collected in 2017 that had a different distribution of weekly reported influenza cases. As with the training/test set, the additional validation set was also adjusted to 50:50 for influenza and noninfluenza episodes. [Figure 2](#) summarizes the overall pipeline for data preprocessing.

**Table 1.** Examples of episode separation.

Episodes and the user-added date and time log	Time elapsed since the previous log
<b>Episode 1</b>	
2018-09-06 22:25	N/A <sup>a</sup>
2018-09-06 22:37	0 h 12 min
2018-09-06 23:53	0 h 16 min
2018-09-07 1:01	0 h 8 min
2018-09-07 2:49	1 h 48 min
2018-09-07 10:00	7 h 11 min
2018-09-07 15:56	5 h 56 min
2018-09-07 21:15	5 h 19 min
2018-09-08 11:20	14 h 5 min
2018-09-08 12:10	0 h 50 min
2018-09-08 21:10	9 h 0 min
2018-09-09 12:14	15 h 4 min
2018-09-09 21:38	9 h 24 min
2018-09-10 9:40	12 h 2 min
2018-09-10 21:30	11 h 50 min
2018-09-11 9:14	11 h 44 min
2018-09-11 19:14	10 h 0 min
<b>Episode 2</b>	
2018-10-03 22:11	> 24 h
2018-10-03 22:12	0 h 1 min
2018-10-03 22:26	0 h 14 min
2018-10-03 23:31	1 h 5 min
2018-10-04 0:31	1 h 0 min
2018-10-04 2:38	2 h 7 min
<b>Episode 3</b>	
2018-10-11 8:30	> 24 h
2018-10-11 10:10	1 h 40 min
2018-10-11 10:12	0 h 2 min
2018-10-11 10:14	0 h 2 min
2018-10-11 11:35	1 h 21 min

<sup>a</sup>N/A: not applicable.



**Figure 2.** Pipeline for data preprocessing. KCDC: Korea Center for Disease Control.

## Deep Learning Model and Training Hyperparameters

We used GRU-D as our baseline model [26]. GRU-D is a modified design of the gated recurrent unit (GRU) neural network structure based on a recurrent neural network. Unlike in the GRU, the mask and timestamp were combined together, and input was manipulated to 3-channel data. Since Fever Coach data were characterized by a variety of missing values, we considered that the mask system of the GRU-D structure would be effective in our experiment. Backpropagation was not performed for the masked data; therefore, it did not update parameters. The input data were manipulated to 3-channel data, which were concatenated with a timestamp and masked as previously described. Thus, the shape of the matrix  $X_{\text{input}}$  was  $3 \times D \times T$ , where  $D$  is the number of variables for each experiment, and  $T$  is the maximum number of time series. We used  $T=70$  in the experiment in that the maximum count of the input data in 1 episode was 70. The maximum number of variable dimensions in our experiment was 16 (4 for age, 6 for meteorological data, and 1 each for sex, weight, influenza surveillance, app-based surveillance, antibiotic administration, and antipyretics administration). We performed 3 experiments using different combinations of variables. First, we used the

entire 16 dimensions (7 variables) for inputting the model, and 2 additional experiments were performed to evaluate the effect of the input variables on performance. The second experiment was performed with the same conditions as the first, except a single variable was removed, which brought the number of variables to 6. The third experiment was similar to the first as well, except for the addition of 1 variable out of the 3 (body temperature, antipyretics administration, and antibiotic administration). We used binary cross-entropy as a loss function, and we used accuracy as an evaluation metric to choose the best model. All hidden states were initialized to 0. We used the optimizer, rectified adaptive moment estimation, with a learning rate of 0.0001 [33]. The total number of epochs was 50. The softmax function was used as an activation function. We used a dropout of 0.01 to prevent overfitting. All the input variables were normalized to have a mean of 0 (SD 1). The codes are publicly available at a GitHub repository [34].

## Results

The total number of episodes obtained was 6657. Out of these 6657 episodes, 3326 were diagnosed with influenza. The average and SD of each episode length were 29.24 (SD 21.79). Table 2 summarizes the general characteristics of the processed data.

**Table 2.** General characteristics of the data set.

Variables	Year 2017	Year 2018
<b>Body temperature</b>		
Average number of inputs	15.05	20
Variance in the number of inputs	16.32	18.29
<b>Antipyretic administration</b>		
Average number of inputs	4.578	6.040
Variance in the number of inputs	4.685	24.03
<b>Antibiotic administration, n</b>		
At least 1 antibiotic administration	372	4705
No antibiotic administration	2118	1952
<b>Age (years), n</b>		
0 to 2	886	2529
2 to 5	1328	3564
5 to 12	262	479
Older than 12	14	85
<b>Sex, n</b>		
Male	1246	3348
Female	1244	3309

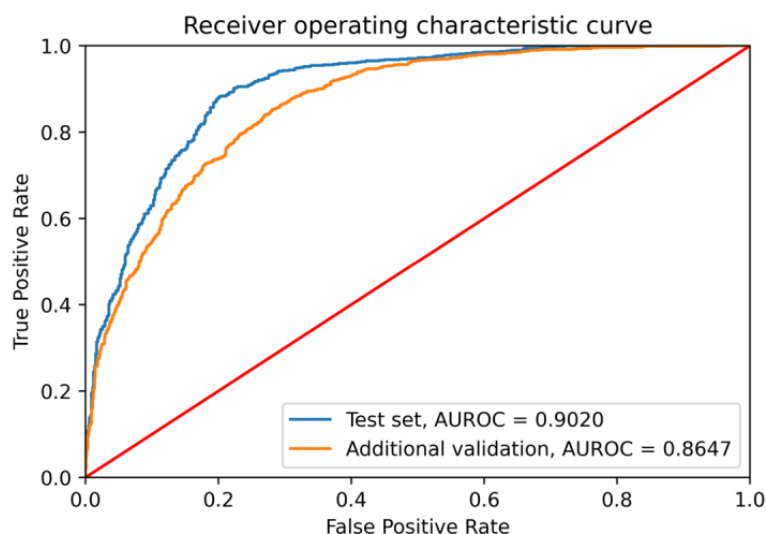
Based on the GRU-D, the proposed screening algorithm used PGHD (body temperature records, antipyretic drug administration records, and antibiotic drug administration records), app-based surveillance, and meteorological data as the input variables. The area under the receiver operating characteristic (AUROC) curve of the test data set was 0.902, with an accuracy of 82.43% (95% CI 80.28%-84.44%), a

sensitivity of 84.20% (95% CI 81.07%-87.00%), a specificity of 80.92% (95% CI 77.85%-83.73%), a positive predictive value (PPV) of 79.05% (95% CI 76.38%-81.50%), and a negative predictive value (NPV) of 85.69% (95% CI 83.26%-87.83%). The confusion matrix and the receiver operating characteristic (ROC) curve are shown in [Figures 3](#) and [4](#), respectively.

**Figure 3.** Confusion matrix for the test set and the additional validation set.

Confusion Matrix					
Test Set			Additional Validation Set		
	Predicted Positive	Predicted Negative		Predicted Positive	Predicted Negative
Actual Positive	581	97	Actual Positive	951	172
Actual Negative	137	517	Actual Negative	241	976

**Figure 4.** Receiver operating characteristic (ROC) curve illustrating the screening ability of the model. The red line shows a random guess, the blue line is the result of the test set collected in 2018, and the orange line is the result of additional validation using data from 2017. AUROC curve: area under the receiver operating characteristic curve.



Considering that the influenza epidemic is slightly different each year, we prepared additional validation set as described in the methods section. For the additional validation set, we achieved an area under the curve (AUC) of 0.8647, an accuracy of 77.99% (95% CI 76.31%-79.61%), a sensitivity of 82.35% (95% CI 79.91%-84.61%), a specificity of 74.79% (95% CI 72.46%-77.02%), a PPV of 70.57% (95% CI 68.59%-72.47%), and an NPV of 85.24% (95% CI 83.47%-86.84%).

We also attempted to evaluate the effect of the input variables on performance in 2 ways. First, we removed them one at a time from all variables. Second, we added them one at a time from baseline variables. To remove them one by one, we first trained the model using all 10 input variables and measured the

performance at that time. We then removed 1 input variable and trained the model on the same data set using a total of 9 input variables and measured the performance. We obtained a total of 10 results and summarized them in Table 3. For example, the second row means all variables except fever were used. As a result, the app-based surveillance turned out to be the most influential variable, even though it had little effect on specificity. The second most influential variable was the meteorological observation data. Interestingly, KCDC surveillance data did not seem to have a significant impact. The meteorological factors and app-based surveillance seemed to compensate for the exclusion of the KCDC surveillance data from the input variables.

**Table 3.** The effects of the removal of each variable from the analysis. “—<Variable>” means that the variable was singularly removed from the list of variables for the corresponding experiment.

Variable	Sensitivity	Specificity	AUROC <sup>a</sup>	Accuracy	NPV <sup>b</sup>	F <sub>1</sub>
All	0.8171	0.8425	0.8931	0.8296	0.8163	0.8300
—Sex	0.8510	0.8028	0.8960	0.8273	0.8387	0.8338
—Weight	0.8171	0.8150	0.8832	0.8161	0.8113	0.8189
—Age	0.8333	0.8346	0.8911	0.8339	0.8333	0.8339
—Fever	0.8083	0.8287	0.8882	0.8183	0.8065	0.8191
—Antipyretics	0.8510	0.8058	0.8744 <sup>c</sup>	0.8288	0.8392	0.8350
—Anti-viral agent	0.8304	0.8211	0.8892	0.8258	0.8236	0.8292
—App-based surveillance	0.8215	0.7905	0.8775	0.8063 <sup>c</sup>	0.8103	0.8120 <sup>c</sup>
—KCDC <sup>d</sup> surveillance	0.8614	0.7813 <sup>c</sup>	0.8892	0.8221	0.8446	0.8313
—Meteorological	0.7950 <sup>c</sup>	0.8486	0.8900	0.8213	0.7997 <sup>c</sup>	0.8191

<sup>a</sup>AUROC: area under the receiver operating characteristic.

<sup>b</sup>NPV: negative predictive value.

<sup>c</sup>The highest decrease in the value for the corresponding column.

<sup>d</sup>KCDC: Korea Center for Disease Control.

Another experiment was conducted to observe the performance changes by defining the base features and adding the variables one at a time (Table 4). The baseline features used were body temperature and the antipyretic and antibiotic drug data. We repeated the analysis by adding each variable to the base features and observing the performance. In each experiment, a total of 4 input variables was used. Consequently, gender data were found to slightly decrease the AUC performance ( $-0.02$ ), but there was no significant difference between the baseline performance and the performance modified by the addition of

gender. Weight and age also displayed no significant differences. For the variables of meteorological data, app surveillance, and KCDC laboratory surveillance, each significantly improved the performance ( $P < .001$ ). There was no significant difference between the performance of "baseline features + app surveillance" and that of "baseline features + meteorological data" ( $P = .48$ ). Similarly, there was no significant difference between the performance of "baseline features + app surveillance" and that of "baseline features + KCDC laboratory surveillance" ( $P = .46$ ).

**Table 4.** Effect of each variable on the analysis. The baseline included body temperature, antipyretic drug, and antibiotic drug data. "+<variable>" means that the variable was added to the baseline for the analysis and then removed for the next analysis (noncumulative addition).

Variable	Sensitivity	Specificity	AUROC <sup>a</sup>	Accuracy	NPV <sup>b</sup>	F <sub>1</sub>
Baseline	0.6018	0.7187	0.7221	0.6592	0.6351	0.6425
+sex	0.5678	0.7401	0.7087	0.6524	0.6229	0.6245
+weight	0.5734	0.7523	0.7232	0.6619	0.6332	0.6315
+age	0.5634	0.7477	0.7201	0.6539	0.6229	0.6237
+app surveillance	0.8673 <sup>c</sup>	0.7599	0.8808 <sup>c</sup>	0.8146 <sup>c</sup>	0.8467 <sup>c</sup>	0.8264 <sup>c</sup>
+KCDC <sup>d</sup> surveillance	0.7670	0.7936 <sup>c</sup>	0.8607	0.7800	0.7666	0.7802
+meteorological	0.8127	0.7470	0.8712	0.7802	0.7961	0.7888

<sup>a</sup>AUROC: area under the receiver operating characteristic.

<sup>b</sup>NPV: negative predictive value.

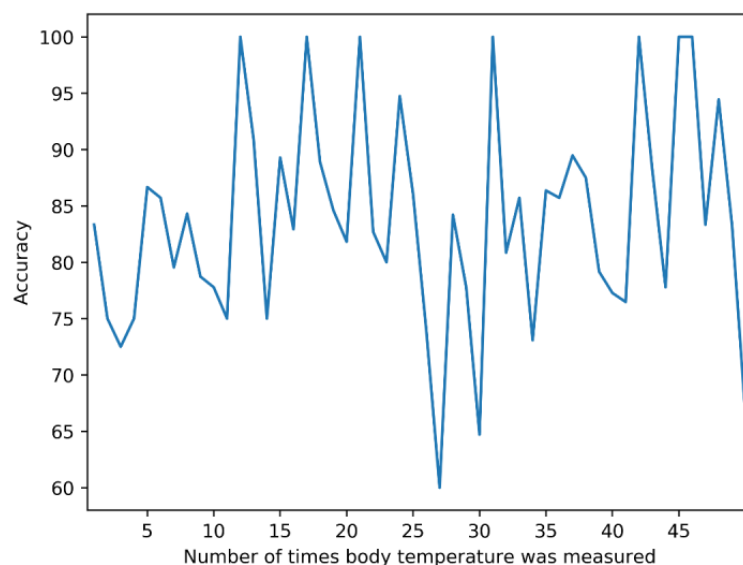
<sup>c</sup>The highest increase in the value for the corresponding column.

<sup>d</sup>KCDC: Korea Center for Disease Control.

Finally, we looked at the correlation between the duration of the input data and the screening performance. Figure 5 describes the association between the duration of body temperature records and the screening performance. We initially assumed that the prediction would be more accurate if the user entered more data.

However, in reality, no correlation was found between the duration of the input data and the screening performance. Spearman rank correlation coefficient was 0.0916. Thus, the association was not considered to be statistically significant.

**Figure 5.** Screening performance versus the number of body temperature records. The y-axis shows the percentage of accuracy, and the x-axis refers to the number of body temperatures entered by the user.



## Discussion

With this study, we investigated the possibility of screening for influenza using PGHD, such as body temperature and medication records collected from an mHealth app.

At the beginning of this study, we did not know whether body temperature would change when antipyretics were administered, or if body temperature alone was more important. Although fever is a major symptom of influenza, it is impossible to diagnose influenza using only body temperature changes [24,25]. Therefore, we hypothesized that patients with influenza would respond more slowly to antipyretics. To test this hypothesis, we specifically looked at the difference between the performance of the model with and without antipyretic administration. There was a greater change in performance when the antipyretic dose records were removed from the input variable than when only the body temperature was removed. Based on these results, we conclude that the model works as expected. Antibiotic administration records are another variable that we considered important. We expected that the antibiotic administration records and antipyretic administration records would have similar effects, but antibiotic administration records appeared to limit the performance. This might have been due to the ineffectiveness of antibiotics or unnecessary prescription of antibiotics. In our data, 1952 of all 6657 users were prescribed antibiotics, and 674 of those who were prescribed drugs were diagnosed with diseases other than influenza.

Body temperature is known to be one of the most important symptoms of influenza. However, its effect on the model was not as strong as we expected. A temperature higher than 38.3 °C was recorded at least once during 97.42% (6485/6657) episodes in our data. This shows that the majority of users used the app when their children had a fever, which was the original purpose of the app. Among the episodes, 50.82% (3296/6485) were those of influenza, and 49.18% (3189/6485) were due to other conditions. The mean and variance of body temperature in the patient group diagnosed with influenza were 38.1519 °C and 0.8611 °C, respectively; and the mean and variance of body temperature with other conditions were 38.0449 °C and 0.8367 °C. There was a significant difference between the 2 groups ( $P < .001$ ). We speculate that because the app focused on fever, the predictive power of body temperature for influenza was diminished.

One interesting finding was the effect that sex had on specificity. Although some studies have shown that there is a difference in

influenza prevalence by gender, our data found that the sex ratio was almost equal, with 1677 males and 1660 females diagnosed with influenza. Moreover, when we excluded sex from the input variables, the accuracy and  $F_1$  measure did not significantly change. We obtained similar results by repeating the ablation study. Therefore, further research may be needed to clarify this point.

In summary, age, weight, and gender had little effect on the screening performance. App-based surveillance has greatly improved the screening performance and is nearly identical to using KCDC laboratory surveillance or meteorological data, which are frequently used as indicators of influenza outbreaks.

This study has several limitations. First, the training and validation data used were self-reported by the patients. Most users reported their diagnosis using their smartphones; thus, these data were not reported by clinicians. Therefore, we cannot ascertain that the same results would be recorded if hospital-generated data were used. Also, primary care doctors usually use the RIDT instead of RT-PCR to diagnose influenza. As the RIDT has low reliability, our ground truth label may be noisy. For the deep learning model, if the character of the data on deployment is slightly different from that of the training data, it is difficult to achieve the expected performance on validation due to the difficulties in analyzing the effect of the data distribution and input variables on the model [35]. Since the data did not include laboratory results, they are difficult to use in a clinical setting or for general epidemiological analysis; and we expect that the application of limited screening tests through the Fever Coach app will be possible with further research. We are planning to conduct a prospective observational study to address these limitations. Second, various methods were used to measure body temperature. Some of the app users used axillary instead of tympanic temperatures. As there are no primary blood vessels in the axilla, the axillary temperatures are less accurate. This may have influenced the performance of the model.

Screening for influenza can be challenging due to the low sensitivity of rapid antigen tests and the lack of proper screening tests. In this study, we developed a deep learning-based screening tool using PGHD obtained from an mHealth app. The experimental results confirm that PGHD from an mHealth app can be a complementary tool for screening for influenza in individual patients. Since our digital approach can screen patients without physical contact, this approach could be quite beneficial in screening new contagious diseases.

## Authors' Contributions

HC implemented the code and performed the experiments. MK and JC manipulated the raw data for preprocessing and designed the experiments. JS provided the data and designed the experiments. SYS designed the experiments and supervised the study. All authors wrote the manuscript and discussed the results. HC and MK equally contributed to this work. JS and SYS are co-corresponding authors.

## Conflicts of Interest

SYS holds stocks in Mobile Doctor, which created the app Fever Coach. SYS also holds stocks in Hurraypositive and Mune, serves as an outside director in Life Semantics, and is a partner of Digital Healthcare Partners. MK is Chief Medical Information



Officer of Mobile Doctor and has equity in Mobile Doctor. SJW is the founding member and Chief Executive Officer of Mobile Doctor. SJW is also Chief Executive Officer of Aim Med and a partner in Digital Healthcare Partners. JC was an internship researcher for Mobile Doctor.

## Multimedia Appendix 1

App-based surveillance calculated from user input data.

[[XLSX File \(Microsoft Excel File\), 11 KB - jmir\\_v22i10e21369\\_app1.xlsx](#)]

## References

1. Nittas V, Lun P, Ehrler F, Puhon MA, Mütsch M. Electronic Patient-Generated Health Data to Facilitate Disease Prevention and Health Promotion: Scoping Review. *J Med Internet Res* 2019 Oct 14;21(10):e13320 [[FREE Full text](#)] [doi: [10.2196/13320](#)] [Medline: [31613225](#)]
2. Bhavnani SP, Narula J, Sengupta PP. Mobile technology and the digitization of healthcare. *Eur Heart J* 2016 May 07;37(18):1428-1438 [[FREE Full text](#)] [doi: [10.1093/eurheartj/ehv770](#)] [Medline: [26873093](#)]
3. Lai AM, Hsueh PS, Choi YK, Austin RR. Present and Future Trends in Consumer Health Informatics and Patient-Generated Health Data. *Yearb Med Inform* 2017 Aug;26(1):152-159 [[FREE Full text](#)] [doi: [10.15265/IY-2017-016](#)] [Medline: [29063559](#)]
4. Kallinikos J, Tempini N. Patient Data as Medical Facts: Social Media Practices as a Foundation for Medical Knowledge Creation. *Information Systems Research* 2014 Dec;25(4):817-833. [doi: [10.1287/isre.2014.0544](#)]
5. Shapiro M, Johnston D, Wald J, Mon D. Patient-Generated Health Data: White Paper. 2012 Apr. URL: [https://www.healthit.gov/sites/default/files/rti\\_pghd\\_whitepaper\\_april\\_2012.pdf](https://www.healthit.gov/sites/default/files/rti_pghd_whitepaper_april_2012.pdf) [accessed 2019-01-04]
6. Nundy S, Lu CE, Hogan P, Mishra A, Peek ME. Using Patient-Generated Health Data From Mobile Technologies for Diabetes Self-Management Support: Provider Perspectives From an Academic Medical Center. *J Diabetes Sci Technol* 2014 Jan;8(1):74-82 [[FREE Full text](#)] [doi: [10.1177/1932296813511727](#)] [Medline: [24876541](#)]
7. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. *J Diabetes Sci Technol* 2017 Sep;11(5):1015-1027 [[FREE Full text](#)] [doi: [10.1177/1932296817713506](#)] [Medline: [28560898](#)]
8. Milani RV, Lavie CJ, Wilt JK, Bober RM, Ventura HO. New Concepts in Hypertension Management: A Population-Based Perspective. *Prog Cardiovasc Dis* 2016;59(3):289-294. [doi: [10.1016/j.pcad.2016.09.005](#)] [Medline: [27693860](#)]
9. Lv N, Xiao L, Simmons ML, Rosas LG, Chan A, Entwistle M. Personalized Hypertension Management Using Patient-Generated Health Data Integrated With Electronic Health Records (EMPOWER-H): Six-Month Pre-Post Study. *J Med Internet Res* 2017 Sep 19;19(9):e311 [[FREE Full text](#)] [doi: [10.2196/jmir.7831](#)] [Medline: [28928111](#)]
10. Kim BY, Lee J. Smart Devices for Older Adults Managing Chronic Disease: A Scoping Review. *JMIR Mhealth Uhealth* 2017 May 23;5(5):e69 [[FREE Full text](#)] [doi: [10.2196/mhealth.7141](#)] [Medline: [28536089](#)]
11. Wongkoblap A, Vadillo MA, Curcin V. Researching Mental Health Disorders in the Era of Social Media: Systematic Review. *J Med Internet Res* 2017 Jun 29;19(6):e228 [[FREE Full text](#)] [doi: [10.2196/jmir.7215](#)] [Medline: [28663166](#)]
12. Seyhan AA, Carini C. Are innovation and new technologies in precision medicine paving a new era in patients centric care? *J Transl Med* 2019 Apr 05;17(1):114 [[FREE Full text](#)] [doi: [10.1186/s12967-019-1864-9](#)] [Medline: [30953518](#)]
13. Alessa A, Faezipour M. A review of influenza detection and prediction through social networking sites. *Theor Biol Med Model* 2018 Feb 01;15(1):2. [doi: [10.1186/s12976-017-0074-5](#)] [Medline: [29386017](#)]
14. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Feb 19;457(7232):1012-1014. [doi: [10.1038/nature07634](#)] [Medline: [19020500](#)]
15. Kim M, Yune S, Chang S, Jung Y, Sa SO, Han HW. The Fever Coach Mobile App for Participatory Influenza Surveillance in Children: Usability Study. *JMIR Mhealth Uhealth* 2019 Oct 17;7(10):e14276 [[FREE Full text](#)] [doi: [10.2196/14276](#)] [Medline: [31625946](#)]
16. Shin S, Seo D, An J, Kwak H, Kim S, Gwack J, et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. *Sci Rep* 2016 Sep 06;6:32920 [[FREE Full text](#)] [doi: [10.1038/srep32920](#)] [Medline: [27595921](#)]
17. Seo D, Jo M, Sohn CH, Shin S, Lee J, Yu M, et al. Cumulative query method for influenza surveillance using search engine data. *J Med Internet Res* 2014 Dec 16;16(12):e289 [[FREE Full text](#)] [doi: [10.2196/jmir.3680](#)] [Medline: [25517353](#)]
18. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu Near You: Crowdsourced Symptom Reporting Spanning 2 Influenza Seasons. *Am J Public Health* 2015 Oct;105(10):2124-2130. [doi: [10.2105/AJPH.2015.302696](#)] [Medline: [26270299](#)]
19. Centers for Disease Control and Prevention. Rapid Influenza Diagnostic Tests. 2016 Oct 25. URL: [https://www.cdc.gov/flu/professionals/diagnosis/clinician\\_guidance\\_ridt.htm](https://www.cdc.gov/flu/professionals/diagnosis/clinician_guidance_ridt.htm) [accessed 2020-09-22]
20. Grijalva CG, Poehling KA, Edwards KM, Weinberg GA, Staat MA, Iwane MK, et al. Accuracy and interpretation of rapid influenza tests in children. *Pediatrics* 2007 Jan;119(1):e6-11. [doi: [10.1542/peds.2006-1694](#)] [Medline: [17200259](#)]

21. Chartrand C, Leeftang MM, Minion J, Brewer T, Pai M. Accuracy of rapid influenza diagnostic tests: a meta-analysis. *Ann Intern Med* 2012 Apr 03;156(7):500-511. [doi: [10.7326/0003-4819-156-7-201204030-00403](https://doi.org/10.7326/0003-4819-156-7-201204030-00403)] [Medline: [22371850](https://pubmed.ncbi.nlm.nih.gov/22371850/)]
22. Uyeki T, Prasad R, Vukotich C, Stebbins S, Rinaldo C, Ferng Y, et al. Low sensitivity of rapid diagnostic test for influenza. *Clin Infect Dis* 2009 May 01;48(9):e89-e92. [doi: [10.1086/597828](https://doi.org/10.1086/597828)] [Medline: [19323628](https://pubmed.ncbi.nlm.nih.gov/19323628/)]
23. Chen S, Chen Y, Chiang W, Kung H, King C, Lai M, et al. Field performance of clinical case definitions for influenza screening during the 2009 pandemic. *Am J Emerg Med* 2012 Nov;30(9):1796-1803. [doi: [10.1016/j.ajem.2012.02.016](https://doi.org/10.1016/j.ajem.2012.02.016)] [Medline: [22633701](https://pubmed.ncbi.nlm.nih.gov/22633701/)]
24. Call SA, Vollenweider MA, Hornung CA, Simel DL, McKinney WP. Does this patient have influenza? *JAMA* 2005 Feb 23;293(8):987-997. [doi: [10.1001/jama.293.8.987](https://doi.org/10.1001/jama.293.8.987)] [Medline: [15728170](https://pubmed.ncbi.nlm.nih.gov/15728170/)]
25. Monto AS, Gravenstein S, Elliott M, Colopy M, Schweinle J. Clinical signs and symptoms predicting influenza infection. *Arch Intern Med* 2000 Nov 27;160(21):3243-3247. [doi: [10.1001/archinte.160.21.3243](https://doi.org/10.1001/archinte.160.21.3243)] [Medline: [11088084](https://pubmed.ncbi.nlm.nih.gov/11088084/)]
26. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Sci Rep* 2018 Apr 17;8(1):6085 [FREE Full text] [doi: [10.1038/s41598-018-24271-9](https://doi.org/10.1038/s41598-018-24271-9)] [Medline: [29666385](https://pubmed.ncbi.nlm.nih.gov/29666385/)]
27. Fever Coach. URL: <https://fevercoach.net/?lang=en> [accessed 2020-09-22]
28. Korea Meteorological Administration. 종관기 상관측(ASOS). URL: <https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36> [accessed 2020-09-22]
29. Korean Centre for Disease Control and Prevention. Infectious Disease Portal. URL: <http://www.cdc.go.kr/npt/biz/npp/iss/influenzaStatisticsMain.do> [accessed 2020-09-22]
30. Centers for Disease Control and Prevention. How Flu Spreads. 2018. URL: <https://www.cdc.gov/flu/about/disease/spread.htm> [accessed 2020-09-22]
31. Lessler J, Reich NG, Brookmeyer R, Perl TM, Nelson KE, Cummings DA. Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect Dis* 2009 May;9(5):291-300 [FREE Full text] [doi: [10.1016/S1473-3099\(09\)70069-6](https://doi.org/10.1016/S1473-3099(09)70069-6)] [Medline: [19393959](https://pubmed.ncbi.nlm.nih.gov/19393959/)]
32. ISO. ISO 8601 Date And Time Format. URL: <https://www.iso.org/iso-8601-date-and-time-format.html> [accessed 2020-09-22]
33. Liu, Liyuan, Jiang, Chen, Weizhu, Liu. On the Variance of the Adaptive Learning Rate and Beyond Internet. arXiv. URL: <https://arxiv.org/abs/1908.03265> [accessed 2020-09-22]
34. bmiskkuedu/FeverCoach. GitHub. URL: <https://github.com/bmiskkuedu/FeverCoach> [accessed 2020-09-24]
35. Mani S, Sankaran A, Tamilselvan S, Sethi A. Coverage Testing of Deep Learning Models using Dataset Characterization. arXiv. 2019. URL: <https://arxiv.org/abs/1911.07309> [accessed 2020-09-14]

## Abbreviations

**AUC:** area under the curve  
**AUROC:** area under the receiver operating characteristic  
**GRU:** gated recurrent unit  
**ILI:** influenza-like illness  
**KCDC:** Korea Center for Disease Control  
**MERS:** Middle East respiratory syndrome  
**mHealth:** mobile health  
**NPV:** negative predictive value  
**PGHD:** patient-generated health data  
**PPV:** positive predictive value  
**RIDT:** rapid influenza diagnostic test  
**ROC:** receiver operating characteristic  
**RT-PCR:** reverse transcription-polymerase chain reaction

*Edited by G Eysenbach; submitted 12.06.20; peer-reviewed by IH Kwon, CM Choi; comments to author 20.07.20; revised version received 16.08.20; accepted 18.08.20; published 29.10.20.*

### *Please cite as:*

Choo H, Kim M, Choi J, Shin J, Shin SY  
 Influenza Screening via Deep Learning Using a Combination of Epidemiological and Patient-Generated Health Data: Development and Validation Study  
*J Med Internet Res* 2020;22(10):e21369  
 URL: <http://www.jmir.org/2020/10/e21369/>  
 doi: [10.2196/21369](https://doi.org/10.2196/21369)  
 PMID: [33118941](https://pubmed.ncbi.nlm.nih.gov/33118941/)

©Hyunwoo Choo, Myeongchan Kim, Jiyun Choi, Jaewon Shin, Soo-Yong Shin. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Information Access and Use by Patients With Cancer and Their Friends and Family: Development of a Grounded Theory

Maclean Thiessen<sup>1,2,3</sup>, MD, MN, FRCPC; Shane Sinclair<sup>3,4</sup>, PhD; Patricia A Tang<sup>4</sup>, MD, FRCPC; Shelley Raffin Bouchal<sup>3</sup>, RN, PhD

<sup>1</sup>Research Institute in Oncology and Hematology, CancerCare Manitoba, Winnipeg, MB, Canada

<sup>2</sup>Department of Internal Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada

<sup>3</sup>Faculty of Nursing, University of Calgary, Calgary, AB, Canada

<sup>4</sup>Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

**Corresponding Author:**

Maclean Thiessen, MD, MN, FRCPC

Research Institute in Oncology and Hematology

CancerCare Manitoba

675 McDermot Avenue

Winnipeg, MB, R3E 0V9

Canada

Phone: 1 2047878776

Email: [macthiessen@gmail.com](mailto:macthiessen@gmail.com)

## Abstract

**Background:** Information has been identified as a commonly unmet supportive care need for those living with cancer (ie, patients and their friends and family). The information needed to help individuals plan their lives around the consequences of cancer, such as the receipt of health care, is an example of an important informational need. A suitable theory to guide the development of interventions designed to meet this informational need has not been identified by the authors.

**Objective:** The aim of this study is to generate a grounded theory capable of guiding the development of interventions designed to assist those living with cancer in meeting their informational needs.

**Methods:** Classic grounded theory was used to analyze data collected through digitally recorded one-on-one audio interviews with 31 patients with cancer and 29 friends and family members. These interviews focused on how the participants had accessed and used information to plan their lives and what barriers they faced in obtaining and using this information.

**Results:** The theory that emerged consisted of 4 variables: personal projects, cancer as a source of disruption to personal projects, information as the process of accessing and interpreting cancer-related data (CRD) to inform action, and CRD quality as defined by accessibility, credibility, applicability, and framing. CRD quality as a moderator of personal project disruption by cancer is the core concept of this theory.

**Conclusions:** Informational resources providing accessible, credible, applicable, and positively framed CRD are likely key to meeting the information needs of those affected by cancer. Web-based informational resources delivering high-quality CRD focused on assisting individuals living with cancer in maintaining and planning their personal projects are predicted to improve quality of life. Research is needed to develop and integrate resources informed by this theoretical framework into clinical practice.

(*J Med Internet Res* 2020;22(10):e20510) doi:[10.2196/20510](https://doi.org/10.2196/20510)

**KEYWORDS**

persons; personal autonomy; patient-centered care; health education; health information-seeking behavior; grounded theory; empowerment; cancer; qualitative research; adaptation, psychological; mobile phone

## Introduction

### Background

One of the most commonly reported unmet supportive care needs of those facing cancer, including patients [1-4] and their friends and family [5-7], is for information. Common informational needs include those related to prognosis, how to care for someone with cancer, and the benefits and toxicities of treatment [2,8-14]. Unmet information needs are associated with decreased treatment adherence [15,16], increased health care costs [17-19], anxiety [20,21], and depression [20].

In a previous study conducted by the lead author (MT) [22] it was identified that the diagnosis of malignancy resulted in widespread interruption to the lives of those affected by cancer. This study used grounded theory [23,24] to analyze data from 43 semistructured interviews conducted with 18 patients who had been recently diagnosed with cancer and 15 friends and family in Manitoba, Canada. One conclusion from this study was that to support individuals affected by cancer, information is needed that supports them in planning the activities necessary for maintaining participation in the relationships and projects they established before diagnosis—such as those with family, friends, and their work—both in the short and long term. Addressing this finding was a major motivator for the work presented in this study.

Health information-seeking behavior (HISB) is a field of research encompassing how those affected by illness utilize health information. Important areas of study within HISB include how individuals seek, use, and share information [25-27]. The HISB literature can be broadly divided into 3 main categories: (1) coping with health situations, (2) involvement in shared medical decisions, and (3) behavior change and preventative behavior [25,27]. In their concept analysis of HISB, Lambert and Loiselle [27] identified that one goal of the individual engaging in HISB is to better understand what to expect. They identify that HISB serves to “increase predictability” [27] and may assist with “anticipating the sequence of events” [27] that will likely take place. These goals of HISB are consistent with the informational need identified in the study leading up to this work: that information is essential to helping individuals know what to expect in the future so that they can effectively plan how to live their lives [22].

The factors that affect HISB are complex, as illustrated by the examination of the Miller Behavioral Style Scale (MBSS) [26-28] and the model of health information acquisition (HIA) [26,27,29], both of which are utilized in the HISB literature [26,27]. The MBSS is a validated scale [28,30] useful for characterizing individual information-seeking styles in response to both *physical and psychological stress* [28]. The MBSS categorizes individuals as either information seekers (high monitors) or information avoiders (low monitors) [28,30]. High monitors are more likely to seek out information to cope with stressors, whereas the response of low monitors is to avoid information [28,31,32]. The HIA, developed by Freimuth et al [29] from their work with the National Cancer Institute's telephone-based Cancer Information Service [26], predicts that the decision to seek information involves a cost-benefit analysis.

The expected benefit of information is weighed against the effort to seek additional information [26,27,29] in terms of cost considerations such as “financial and time expenditures, frustration, confusion, [and] emotional distress” [29]. Together, the MBSS and the HIA suggest that both intrinsic and external variables impact an individual's ability to access and use health information to plan around the receipt of health care. This conclusion is supported by the wealth of empirical evidence, including multiple systematic reviews [1,5,7,33-40], consistently correlating specific demographic factors (eg, age, gender, education) [33,41,42] with various types of information-seeking behavior [27,43] and health information needs [5,35].

In the context of cancer, it is not clear how to optimally design interventions to support individuals in obtaining the health information that is most useful for them. In general, using theory to guide intervention design results in better outcomes [44,45]. The application of theory in intervention design facilitates the identification of key constructs to be included in the intervention, potentially resulting in a stronger effect [46]. In addition, the results of testing theory-based interventions provide valuable feedback about the accuracy of the theory [23,46], furthering the understanding of the contextual area under study and facilitating modification of the theory to enhance its accuracy [23].

Multiple theories have been used to guide both the understanding of HISB [26,27] and the development of interventions [37]. Importantly, besides the model of HISB by Longo [47,48] and the HIA by Freimuth et al [29,48], few theoretical frameworks regarding HISB have been developed within the context of cancer. In addition, although existing theories in the HISB literature facilitate understanding and explanations of HISB patterns, the utility of these theories for developing interventions is not clear. For instance, although existing theoretical frameworks employed in the HISB cancer context [29,47-49] describe a cost-benefit relationship in terms of whether an individual will search for additional information, they do not provide guidance in terms of how to structure interventions for those affected by cancer to minimize the cost and maximize the benefit of information seeking. A theoretical framework that addresses this gap in the literature is thought to be valuable for developing interventions that address the informational needs of those living with cancer [1,4,11,12,14,15,35,38,48]. Such a framework would be capable of informing the development of interventions that support individuals in planning their lives around the short- and long-term consequences of cancer, including the receipt of treatment and altered life expectancy [22,50].

### Objectives

The objective of this study is to develop a grounded theory capable of guiding the creation of informational resources designed to assist individuals living with cancer in meeting their informational needs by minimizing the cost and maximizing the benefit of information seeking.



## Methods

### Study Approach

This study used classic grounded theory (CGT), a method for discovering theory through iterative data collection and analysis [51-53]. CGT has been identified as a method for uncovering latent behavioral patterns and generating theory capable of guiding practical action for problem solving [54]. This was one reason that CGT was considered the ideal method for this study as the objective required a theory that operationalized (1) why information related to cancer is important for supporting individuals affected by it and (2) how information about cancer can be optimally provided to improve the lives of those affected by it. Both questions assume that shared patterns of behavior exist among those affected by cancer.

### Study Procedures

The study procedures, including the study design, data collection, analysis, and drafting of the report, were conducted primarily by the lead author, who was completing a medical oncology fellowship during the first year of the study and enrolled in a PhD graduate program as well as in active independent clinical practice as a medical oncologist for the subsequent portion of this study. The second and fourth authors provided methodological support in conducting and presenting the grounded theory analysis. The third author provided general research expertise and contextual expertise regarding clinical oncology practice.

### Ethical Considerations

Approval for this study was obtained through the Health Research Ethics Board of Alberta (Study ID: HREBA.CC-17-0365) before the initiation of recruitment, data collection, and data analysis.

### Recruitment

Patients were recruited using posters and invitation letters from a large outpatient cancer facility in Western Canada. Interested patient participants contacted the lead author or primary investigator who provided further details of this study, including its methods, objectives, risks, benefits, and obtained written consent. The patient participants were invited to approach any friends and family to participate in this study as secondary participants. This study was open to all patients aged 18 years or older who had received oncology care. Friends and family participants aged 18 years or older were welcome to participate. Exclusion criteria were limited to not being able to communicate in English and being aged below 18 years. Incentives for participation included being eligible to win one of four Can \$25 (US \$18.79) gift certificates.

The rationale for inclusion of both friends and family participants as well as patient participants in this study was three-fold. First, it was assumed that, besides instances where patients were receiving medications affecting their cognition or had severe neurological sequelae of their cancer, such as a debilitating brain metastasis, there would be no psychological or sociological phenomena differentiating the processes of information seeking and use for those diagnosed with cancer

from their family and friends. Therefore, the concepts and resulting theory that would emerge would likely be valid for both friends and family as well as patients. Second, it is recognized that informal caregivers are often left behind when it comes to supportive care research, including research related to information needs. Although the theory that was expected to emerge would likely be applicable to both groups, without including both patients and friends and family in the study, the validity of the theory for the group not included would likely be questioned. Finally, the contrasting perspectives of friends and family and patients were expected to provide extremely useful data for the purposes of constant comparison, ensuring that theoretical saturation occurred [23,24].

### Data Collection

After obtaining written consent, all participants completed a demographic questionnaire (Multimedia Appendix 1). A review of the electronic medical charts of patient participants was performed to facilitate the identification of details that were not readily available outside of a thorough interview in the style of detailed medical and treatment history. The data from the chart review and questionnaires were compiled into a database to assist with theoretical sampling, an iterative sampling technique to ensure theoretical coverage and heterogeneity [23,24]. For instance, patients were initially interviewed as they were recruited, resulting in a predominance of patients >50 years of age with breast and colorectal cancer being interviewed. Therefore, the database was used to identify and select participants for interviews who were primarily younger patients with less common malignancies. This was important to ensure that the emerging concepts were adequately informed by data from individuals likely to have had contrasting experiences.

All interviews were semistructured (Multimedia Appendix 2), face to face, and audio recorded. The interviews were carried out in participants' homes, apart from 3 interviews conducted over the phone. Interviewees were encouraged to stop the interview at any point if they were no longer comfortable proceeding or needed a break; in addition, they were provided with contact information for psychosocial support available through the cancer center. The interviews with participants took place separately, except for 7 interviews where the patients and their friends and family wanted to be interviewed together. Participants were interviewed once; no repeat interviews were conducted. The audio recordings from the interviews were transcribed by a professional transcriptionist as soon as possible after each interview to facilitate ongoing and iterative data analysis. The average interview length was 53 min overall, 1 hour and 4 min for interviews involving patients, and 36 min for interviews with only friends and family participants. Participants who participated in interviews were offered a 24-hour parking pass to the cancer center.

### Data Analysis

Data analysis using constant comparison was carried out in keeping with the CGT [51]. Data analysis and data collection occurred in an iterative manner, beginning once the first interview was transcribed. Coding, memoing, and theory generation were guided by comparing coded incidents and intentionally selecting participants and interview questions

likely to result in data being collected that would contrast with previously collected data, providing new insight to guide the emerging theory [51]. Data collection ceased when data saturation occurred, whereby no new data emerged from the ensuing interviews. The initial stage of coding the collected data (ie, open coding) resulted in a coding schema emerging and the identification of a core category. This facilitated the next stage of the CGT analysis (ie, selective coding), where the theory began to emerge around the coding categories (ie, variables) associated with the consequences of access to the information that participants identified as being helpful or

unhelpful. The final stage of the CGT analysis (ie, theoretical coding) involved coding to finalize theoretical links between the categories connecting the challenges faced by the participants, their experience with cancer, and information.

## Results

### Participant Characteristics

A total of 37 patients and 36 friends and family consented to be contacted for interviews, with 31 patients and 29 friends and family (Table 1) completing interviews.

**Table 1.** Interviewed participants' demographics (n=60).

Interviewed participants	Values
<b>Interviewed primary participants (n=31)</b>	
Age (years), mean (range)	60 (29-81)
<b>Gender, n (%)</b>	
Male	14 (45)
Female	17 (55)
<b>Initially curative at the time of diagnosis, n (%)</b>	
Yes	26 (84)
No	5 (16)
<b>Currently curative (at the time of interview), n (%)</b>	
Yes	19 (61)
No	12 (39)
<b>Malignancy type, n (%)</b>	
Colorectal	8 (26)
Noncolorectal gastrointestinal malignancy	5 (16)
Breast	9 (29)
Melanoma	5 (16)
Hematologic malignancies	3 (10)
Osteosarcoma	1 (3)
<b>Interviewed secondary participants (n=29)</b>	
Age (years), mean (range)	56 (27-83)
<b>Relationship with the primary participant, n (%)</b>	
Spouse	17 (59)
Child	3 (10)
Sibling	2 (7)
Parent	2 (7)
Friend	5 (17)

The sample of patients who completed the interviews was relatively balanced in terms of gender and included patients with ages ranging from their late 20s to early 80s. Importantly, there was a mix of patients who were being treated with curative intent, as well as those being treated noncuratively for de novo or recurrent metastatic disease, suggesting a wide range of cancer experiences. In terms of friends and family interviewed, ages were similar to patients, likely reflecting the high number of spouses who were included in the interviews.

### The Grounded Theory of Information Access and Use

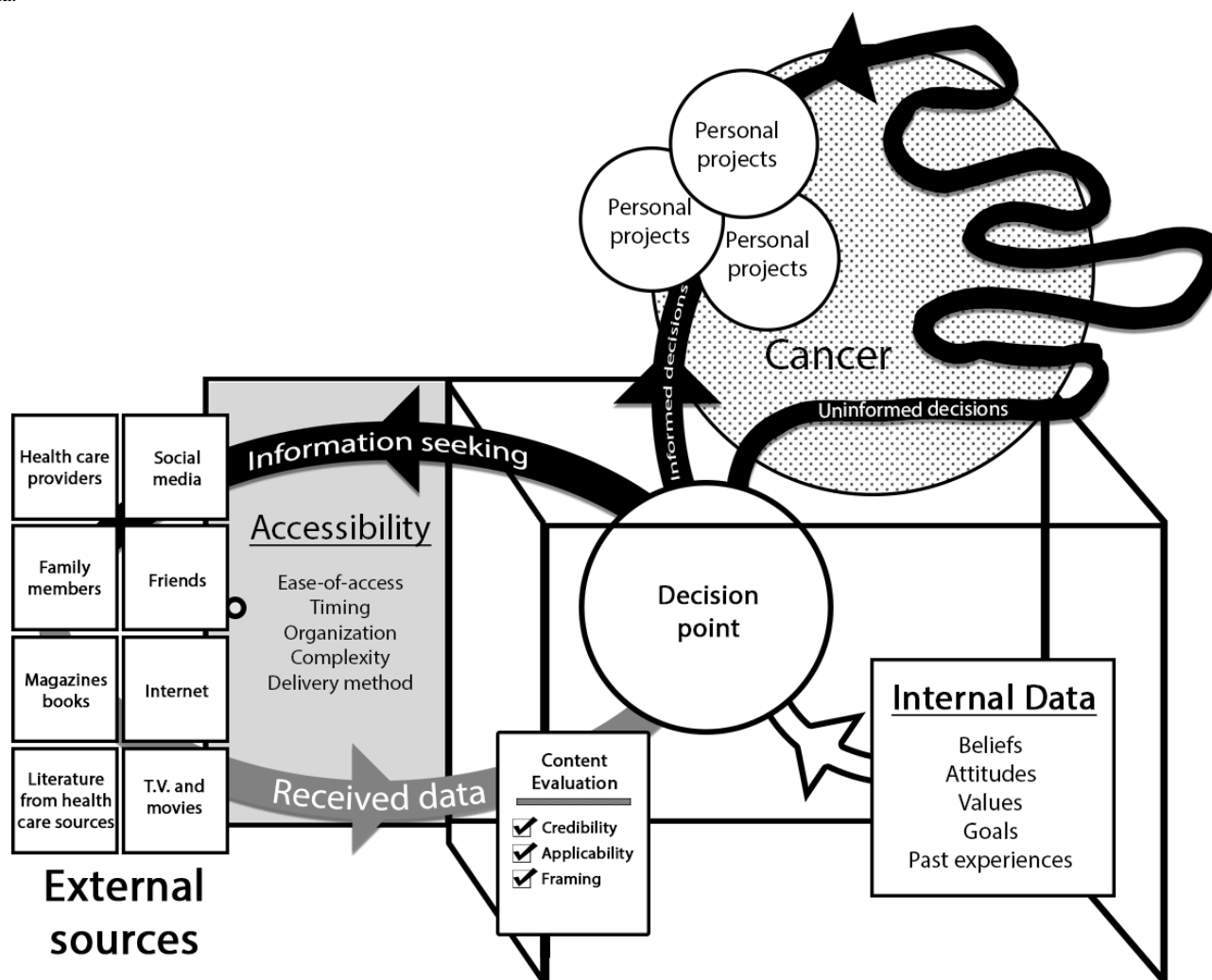
The primary finding, or core variable [51], was that the quality of cancer-related data (CRD) that patients and their friends and family received impacted their ability to plan their lives around the consequences of the malignancy diagnosis. The theory consists of 4 interrelated concepts: (1) personal projects, (2) the cancer project, (3) information as the process of receiving and

interpreting data to inform action, and (4) the quality of CRD received.

At the conclusion of this study, information was understood as the process of informing action based on the CRD that patients and their friends and family received about cancer. For those affected by cancer, CRD came from multiple sources, including health care providers, family, friends, and the internet. CRD from health care providers were the most credible and applicable; however, access to health care providers was often limited to clinical visits where the uptake of the CRD was

limited. CRD found on the internet were readily accessible and provided an opportunity for repeated access. Received CRD are interpreted in the context of internal data, including the individual's personal values, how the individual understands their life story, goals for the future, and previously obtained CRD. This process informs the individual's actions related to managing cancer or their personal projects (eg, career, raising children, being physically fit). The concepts that comprise the theory are described in the following subsections. A graphical model of the theory is presented in Figure 1.

**Figure 1.** A graphical model of the theory of information access and use. The theory suggests that a cancer diagnosis is disruptive, in part, because it decreases the ability of an individual to effectively make decisions about how to invest their time and energy into their personal projects. This is due to a resulting lack of certainty about what to expect, both in terms of the cancer itself and how the consequences of cancer, such as treatment and altered life expectancy, will affect their personal projects. Information is understood as a process involving receiving data about cancer and interpreting it to increase the certainty of the outcome of different actions. High quality data, defined as data that is accessible, credible, applicable, and positively framed, enhances decision-making support and results in improved engagement with personal projects. The dark black arrowed lines represent aspects of the information/action continuum in which energy and time is diverted between personal projects and seeking information. The indirect winding line titled “uninformed decisions” illustrates the inefficient use of an individual's finite energy and time when decisions are not informed by high-quality cancer data.



### Personal Projects: The Context of Cancer

Personal projects refer to the collection of activities that an individual invests a significant amount of energy and time in over a prolonged period of their life. The following quote demonstrates how time and energy shift from one personal

project (eg, child rearing) to another (eg, the project of being healthy) over the course of a lifetime:

*I worked and I [raised] the kids and I was very involved, and I volunteered a ton and then once they graduated it was my turn. And I started to do things for me and get myself healthy and eat healthy and all that. [Patient 1]*

The personal projects of the participants created a unique context for each individual to face the challenges of the cancer experience. Each participant had a unique group of ongoing personal projects, allotting different amounts of time and energy to each, and each was ascribed with unique meaning. Importantly, before diagnosis, these personal projects—and the activities and roles that comprised them—were what the participants invested their time and energy in.

### **Cancer: A New Disruptive Project**

The personal projects of the participants were disrupted following the malignancy diagnosis. Energy and time were diverted from preexisting personal projects to managing the consequences of the cancer diagnosis. Some participants described being able to continue with most aspects of their personal projects but were unable to plan how these would fit into the future. This was because the necessary details that would facilitate planning, such as prognosis or the time and energy commitment needed for treatment, were not made clear for weeks, or even months, after the initial diagnosis. Others described that following the diagnosis, their personal projects essentially halted:

*I was given the diagnosis, sent to the [hospital ward] and then I was in there for over I think it was two weeks or something or longer. [Patient 15]*

For friends and family, disruption to personal projects was related to the amount of support they provided to the individual who received the malignancy diagnosis. A partner of a patient described the disruptive effect of cancer as follows:

*Well obviously, it's been life altering. I guess it's – it's certainly changed what the priorities are in our short term, midterm and long-term activities. ...because like the first priority is always caring for [spouse] making sure [they] get the right [treatment]. [Friends and Family 15]*

In this example, priorities were understood to relate to the relative energy and time that the participant planned to invest in various projects. In contrast, a participant identified as a friend to a primary participant but not as a central member of the patient's support network indicated that the malignancy diagnosis was not disruptive to their personal projects stating:

*I avoid the issue of [their] illness. ...I don't make [them] sick. [Friends and Family 26]*

### **Information: Definition and Function**

At the conclusion of this study, information came to be understood as a multistage process, with steps occurring both externally and internally to the individual. The process participants described included searching for, receiving, and interpreting data related to cancer (ie, CRD) to inform action related to their personal projects and managing cancer. CRD came in many forms, including through conversations with health care providers, friends and family, web searches, and personal experiences. Participants also described actively searching for CRD, using time and energy that would otherwise be used for their personal projects, to understand the diagnosis and its consequences. For instance, one participant described

searching for CRD and using them to answer questions related to the project of raising a family following the malignancy diagnosis:

*But at the start, 'cause I was so scared, I thought "oh my God, I have these little kids that I have to raise" you know? I have a long haul ahead of me, like I got to get through this. What can I do that's going to help benefit me in the long run? [...] what things are going to benefit me health wise? That's going to... help me get through this? [Patient 2]*

This participant went on to describe finding books and other resources that contained CRD that were helpful in navigating the challenges of being a parent while dealing with the consequences of the malignancy.

In contrast to the CRD that were actively sought out, CRD were also obtained passively through sources such as TV shows, news, and casual conversations with friends and family and from clinical encounters with physicians and nurses. In addition, participants indicated that although much of the CRD that they used were obtained from external sources, they also identified internal sources of CRD. For instance, personal knowledge gained from previous cancer experiences with family and friends was identified as an important source of CRD.

After CRD were accessed, participants described using it to guide both small-scale decisions, such as the day-to-day logistical coordination of the activities related to a single personal project (eg, taking time off work to provide transportation to a patient), and large-scale decisions, such as those that would affect all of an individual's personal projects:

*[...] if I had had a better idea about what the progression, the path [forward] is going to be - that would be helpful for me. [...] And I want to plan, right now I want to plan 6 months, I want to plan a year from now. [...] Because of this situation I'm going to be leaving my job and [moving] and to the extent possible I'd like to know, this sounds terribly selfish, but there's a little bit of: how does this affect me? [Friends and Family 27]*

### **Quality CRD: Accessibility, Credibility, Applicability, and Framing**

To understand how information could be optimally provided to those affected by cancer, many of the study interviews included a focus on how CRD had been provided through the cancer center or by clinicians, whether these CRD were helpful, and discussions about how CRD could be better provided. Through this exploration, and the many contrasting examples provided by the participants, 4 themes defining the quality of CRD were identified: accessibility, credibility, applicability, and framing.

#### **Accessibility**

Accessibility refers to how CRD are made available to individuals and is characterized by *ease of access, timing, organization, complexity, and delivery method*.



### Ease of Access

Some external CRD sources can have limited accessibility, whereas others can be repeatedly and conveniently accessed. Health care providers and cancer centers are sources of CRD that have limited access. Participants described CRD available through clinic visits, support groups, and education sessions as being accessible only in certain locations and at certain times. As a result, the CRD provided were not always able to be effectively received. One participant described the experience of comparing the details that they recalled from clinician visits with a friend who had been present for the visits and had been taking notes:

*I still haven't been to a single appointment without bringing someone with me. [...] every single car ride home [when we discuss the appointment] it was like we were in two different [doctor's visits].* [Patient 30]

Resources that could be accessed repeatedly, such as printed materials or the internet, allowed participants to access and interpret CRD at their own pace. The internet was a highly accessible source of CRD for many participants. One patient participant described being confused about the prognosis of their breast cancer despite receiving prognostic CRD from a physician. The participant used Google to gather additional CRD as the physician was not readily available to provide further clarification. The participant eventually decided to accept the treatment that had been recommended after learning through the internet that their breast cancer is a “more aggressive type so that's kind of scary, but then there's the treatment for a year that's supposed to balance it out” (Patient 29).

### Timing

For CRD sources with limited access, participants needed to be able to receive them when they were available. Timing refers to issues where individuals are simply not able to receive CRD even if they are physically presented with CRD. One friend and family participant described an instance of poor timing when CRD was shared with a loved one who was in the hospital and recovering from cancer surgery:

*Well, he was on hard drugs. So I couldn't – like he is stoned, when he was in the hospital. And they are throwing a ton of information at him that I am having a hard time grasping and retaining and so he doesn't have a hope in hell of getting it.* [Friends and Family 3]

### Organization

Organization affects how efficiently individuals are able to identify what content of the provided CRD is relevant to them. One participant described the experience of receiving a printed package of CRD from the cancer center, and the subsequent investment of energy and time to identify what was important:

*[...] it was a lot of brochures and then for me it was about weeding out what was important and relevant, so I just focused on like overall what's going to happen with chemo and then just hone in on like [the patient's type of cancer]. Not that anything – like it's not like the other things aren't relevant but [I had to*

*focus on] what I could like absorb [and] what I needed to know.* [Friends and Family 22]

### Complexity

The complexity of the CRD being shared also affected participants' ability to interpret it. One participant described receiving CRD with technical medical content as “good, but you can only give so much to a laymen and they're not going to understand the rest of it like, it can only be so difficult” (Friends and Family 15). Another participant described the amount of time and energy required to navigate through complicated treatment decisions that had been offered by the medical oncologists. The participant described having multiple “family group meetings” (Friends and Family 9) in which the members of the family would sit in the patient's living room and repeatedly play the recording of the doctor's visit, trying to understand the CRD that were shared in the consultation to make decisions both regarding medical management and how to plan their lives around the data received.

### Delivery Method

The method of delivery was also important in terms of participants' ability to access CRD. Different formats of delivery, such as face-to-face discussions with clinicians, education sessions at the cancer center, and internet content, resulted in differences in terms of ease of access. Participants also expressed relative differences in their comfort in each format. Text-based CRD were universally described as helpful. However, some individuals expressed issues with retrieving internet content (“I don't do the computer” [Patient 3]) or a preference for reading things on paper as opposed to on a computer or smartphone screen.

### Credibility

Participants described receiving CRD from their health care providers, friends, family, the internet, TV, and other cancer survivors. The usefulness of these CRD was related to the credibility (ie, reliability) of the source.

CRD received from health care providers, including handouts and brochures, were generally considered credible. Oncology specialists, including physicians and nurses, were identified as being the most credible sources of CRD. They were described as being able to anticipate questions and provide answers without even being asked, capable of providing reassurance, and answering the patient's questions based on “where [the patient was] coming from” (Patient 7). General practitioners or family physicians were also considered credible sources; however, several participants indicated that they received little CRD about their cancer from their general practitioner. One participant indicated that they did not trust anything from the general practitioner stating that the general practitioner had “missed the diagnosis [of malignancy] for many years” (Patient 1). Although participants described various degrees of trust in internet sources of CRD, websites such as the Canadian Cancer Society's website were identified as highly credible.

Cancer survivors, defined here as those with a personal diagnosis or the close friend or family member of someone with a diagnosis [55], were also credible sources of CRD. Survivors provided practical, real-world knowledge about how to manage



the consequences of cancer. Participants described “comparing notes” (Patient 12) about where to source complementary products such as hand creams and how to plan for certain treatments. In addition, survivors who had been diagnosed with cancer decades earlier and were still alive existed as CRD that a malignancy diagnosis was not necessarily a death sentence:

*Since I have seen people stay over 10 years with [specific type of cancer], I'm hoping I will stay about 10 years [...]. I met a woman who told me “oh, this is my 10th year” [...]. So I believe that if some people can survive it then I will.* [Patient 10]

Participants also described CRD from sources that were not credible. They described interactions with well-meaning friends and family who provided CRD about conspiracy theories and unproven controversial treatments. These examples of CRD were described as “uncomfortable noise” (Patient 30) requiring time and energy to evaluate both its credibility and how best to manage the relationship with its source.

### Applicability

Although participants described the CRD obtained from health care providers as highly credible, the data were not always applicable. Participants described receiving general information packages about nutrition and managing side effects but finding these of limited use or even a source of potential distress. One participant who was receiving immunotherapy described receiving a list of potential side effects of treatment from the medical team providing treatment. The participant described reading through the list and feeling anxious about the potential side effects only to become frustrated when at the bottom of the list it said that immunotherapy patients should “ignore [the list of side effects] and just call the triage number” (Patient 17).

Many participants described receiving CRD from the TV and the internet. CRD from these sources presented challenges for the participants as they were a potential source of fear. One participant described being: “worried about how bad [chemotherapy] was going to be” based on “pictur[ing] it from TV and stuff. Like people just puking all the time” only to find that “nausea was hardly a problem” [Patient 31].

Personal experience provided a source of internal CRD considered to be extremely applicable. Participants, including patients and those supporting them, described that as they gained personal experience with receiving medical care, they were able to find a “rhythm” (Patient 31) as they knew what to expect. This allowed them to become increasingly able to plan activities related to their personal projects, such as their work or other relationships. However, each new challenge, such as an unfamiliar treatment, procedure, or symptom, had the potential to interrupt this rhythm, causing disruption until a new rhythm could be established.

### Framing

Whether CRD were framed in a *positive* way also affected participants' ability to use the data. *Positive framing* involves communicating information in an honest manner that (1) also highlights the best possible outcome including exceptional outliers and (2) provides options for moving forward. Even when the odds were seemingly against them, participants

stressed the importance of focusing on positive outcomes and what they, or clinicians, could do to optimize the situation:

*Maybe they're not right with me. Maybe I'm one of the 5%, because I exercise or whatever... I can pretend that maybe I'm one of the [few] that will beat this, to some degree, not beat it forever, but go a little longer than they told me.* [Patient 1]

*You don't want someone telling you, you're for sure going to have, you know, a really bad rash on your hands and feet. You want someone saying, you might have a bad rash on your feet **and this is what you do about it.*** [Patient 2]

*They've been amazing, everybody. And helpful, and encouraging. [...] They haven't been negative about [it], they've just said that there is no cure for this yet.* [Patient 8]

*I guess the negative part to me is – cause I've heard and seen people that [say] “well I have cancer so I'm going to die, I know that whether it's five years down the road, I'm going to die.” [...] The negative part is “I'm going to die” you know?* [Patient 21]

## Discussion

### Interpretation of Findings

A classic definition of information is “a difference in matter-energy which affects uncertainty in a situation where a choice exists among a set of alternatives” [29,56]. Benner [57] suggests that illness and losses such as death “can disrupt (if not shatter) one's taken-for-granted world” and that recovery comes both from “curing the body” [57] and through (re)integration of the self into “his or her particular world” [57]. The theory that emerged in this study links an individual's ability to remain integrated in a world changed by cancer, through maintenance of connection with personal projects to their ability to access helpful information in a way that does not result in further disruption to their life.

The risk benefit consideration identified in existing theories, such as the HIA [29] and the theoretical framework of HISB by Longo et al [47], identifies that an important step in the information process is deciding whether additional information should be sought (ie, cost-benefit analyses). Similarly, this study identified that the cost of seeking CRD is two-fold. First, seeking CRD was an activity that diverted time and energy from personal projects. Second, the cost of basing expectations and making decisions on CRD that are inaccurate is not negligible, as exemplified by the quotes provided in the Applicability subsection in the *Results* section. Interventions structured on the 4 components of high-quality CRD outlined in this study, including accessibility, credibility, applicability, and framing, are likely to minimize the cost and maximize the benefit of health information seeking for those living with cancer.

### Extending the Theory—A Deeper Grounding of Accessibility and Time

An important component of the information process, in addition to seeking, receiving, and ultimately acting on CRD, occurs

between when CRD are received and action occurs. Although the process of interpreting CRD to inform action was not explored explicitly in this study, insights can be gained by examining the findings of this study in conjunction with the existing literature. First, participants in this study indicated that some of their decisions regarding treatment and their personal lives were based on limited or inaccurate CRD. Second, as participants shared and reflected on their cancer journey they both reflected feeling and displayed anger, sadness, joy, and a wide range of other emotions. Both observations are congruent with the existing literature regarding the challenges individuals living with cancer face in obtaining useful information [1-4,9,11,12,14,15,38,39,42,50] and the significant role emotions play during the cancer experience [58,59]. The role of emotions in decision making has been well documented, with both theory and empirical data supporting that emotions affect decision making in different ways [60,61]. For instance, fear is associated with the interpretation of greater risk, whereas anger is associated with less perceived risk [60,61]. In addition, research supports that individuals revert to less emotional states as time passes from the inciting stressor, resulting in decision making that is less reactive, and instead guided by reasoning that is more rational, better reflecting the individual's personal values [60,62].

The insights gained from the data regarding emotions and decision making is important because they add depth to the concept of accessibility and the subconcept of timing that emerged from this theory. On the basis of the theoretical and empirical data regarding emotion and decision making [60,61], for CRD to be useful, they must be provided well in advance of decision making to increase the probability that decision making will be interpreted in a nonemotionally heightened state. Although there is limited empirical evidence to support this conclusion in the cancer context, the literature regarding patient decision aids (PDAs) is informative. PDAs consist of questionnaires or informational packages provided to patients to assist them in better understanding and engaging with medical decision making [19,63]. In a recent Cochrane Systematic Review, although not directly compared in any of the studies reviewed, PDAs provided before consultation compared with usual care appeared to have a positive impact on patient's accurate perception of risk compared with PDAs provided during consultation (risk ratio 2.25, 95% CI 1.65-3.07 vs risk ratio 1.79, 95% CI 1.28-2.52, respectively) [19]. If fear is indeed associated with inaccurate risk perception [60,61], then this trend suggests that endeavors to understand the role of early provision of high-quality CRD to patients with cancer and those supporting them, guided by the intention of reducing fear and improving decisional quality, may be fruitful.

### ***From Theory to Innovation***

The theory that emerged in this study is useful because it identifies guiding concepts for developing high-quality CRD. Providing CRD that are accessible, credible, applicable, and positively framed is predicted to minimize the cost and maximize the benefit of information seeking. On the basis of what was shared by the participants in the interviews and the resulting theory, it is expected that the internet will be the primary delivery method of any novel informational intervention

informed by this study. Although universal access to the internet is not a reality, with barriers to access existing for some groups such as those of low socioeconomic status [64], it is estimated that approximately 90% of North Americans have internet access [65], with rates of internet usage in seniors (aged 65 years and older) being over 70% in some areas [65]. As identified in this study and in the HISB literature in general [66,67], the internet circumnavigates common issues with accessibility, such as the need to travel or being available only during business hours and the requirement of appointment times to receive information [67]. In other words, it facilitates access to CRD in a way that minimizes the cost to the individual and their personal projects.

Perhaps the biggest challenge with providing highly accessible internet-based CRD is ensuring that it is adequately applicable to "assist with anticipating the sequence of events that will likely take place" [27]. It is plausible that an inversely proportional relationship exists between applicability and accessibility. For instance, in this study health care providers were identified as providing the most applicable information, yet they could only be accessed through an appointment taking place at the cancer center. The internet, on the other hand, was very accessible, but many participants identified not being sure of what information was relevant to them. Similar findings have been reported elsewhere [68,69]. Taken together with the theory that emerged, this relationship suggests that any novel online informational intervention should be integrated with and informed by local clinical practice patterns.

Given the current state of oncology practice, developing informational interventions that deliver high-quality CRD is likely possible. Contemporary clinical oncology practice relies on evidence-based, guideline-informed practice. A recent retrospective analysis of the Surveillance, Epidemiology, and End Results Program-Medicare database identified deviations from guideline recommendations in the metastatic breast cancer setting occurring only 18% of the time [70]. Similar findings in the early breast cancer population have been observed [71], supporting that, at least within the breast cancer context, care is relatively standardized in many centers. Standardization of care means that informational content can be developed that is capable of being both applicable to those in any given cancer context, providing information that helps them predict what to expect. This is because standardization likely facilitates the production of informational content that can be specific about what is going to happen regarding any given process. In contrast, when there is little standardization, specific management details, such as which clinicians will be involved, the treatments that will likely be offered, or the timing of these treatments, may quickly become inaccurate or unreliable, resulting in confusion and distress for individuals using those details to plan their lives. In addition to standardization, internet-based patient portals, which connect patients with cancer to their health care data such as consultation reports, imaging, laboratory values, and informational support, are being established at an increasing number of cancer centers [66,72-74]. This also supports that clinical integration of online informational resources delivering high-quality CRD is possible. Given that the hurdles of applicability and accessibility can be overcome, understanding

how best to meet specific content needs is an important next step on the path to improving the cancer experience.

### Clinical Implications

The theory that emerged in this study informs current clinical practice in several ways. First, it highlights that clinicians are an important source of CRD. The CRD they provide is considered to be both highly credible [68] and applicable by those affected by cancer. However, the observed limitations associated with health care providers as an information source include accessibility and framing. Clinicians are encouraged to be mindful of *overloading* patients and their informal caregivers. The theory that emerged here supports that providing CRD to patients when they are emotionally overwhelmed, physically exhausted, or impaired by medication is not effective. The concept of accessibility highlights one benefit of patients having access to recordings of their visits with health care providers [75], as this intervention allows the CRD shared by health care providers to be carefully reviewed at a time that best suits the patient and their informal caregivers. With regard to framing, it has been reported elsewhere that identifying what the clinician can do for the patient, including treatment options, is an important aspect of sharing bad news [76-78]. On the basis of the findings of this study, clinicians are also encouraged, when appropriate, to empower patients and informal caregivers following bad news discussions by helping them identify ways to help themselves. This may include assisting with realistic goal setting and identifying activities that the patient and informal caregivers can engage in that will improve their situation in a meaningful way, whether directly related to the disease outcome or not.

### Research Implications

It is anticipated that this theory will be useful in guiding the development of novel interventions by providing a framework of key considerations for maximizing the benefit of information seeking in the cancer context. However, it does not provide explicit guidance on content to be included in a novel resource or the format of that content. Although the information needs of those affected by cancer are well characterized, it is not clear from the literature what specific content and resources would be most helpful for meeting those needs. Researchers are encouraged to build on this study and engage with those affected by cancer as partners [79] to systematically identify the content and method of delivery that is most helpful to those navigating the cancer journey, both in general and in specific contexts not limited to geography, culture, age, gender, sexuality, education, and income.

### Limitations

This theory was generated by engaging with adult patients and their friends and family, without exclusion on the basis of cancer type, stage, or treatment intent. The sampling approach focused

on obtaining diverse, contrasting perspectives and experiences [51,80]. It is expected that, through the constant comparative method used in this grounded theory study, the emerging theoretical framework will be applicable across the general cancer context. However, specific demographic and cultural groups were not focused on, as this was not within the scope of this study. It is certainly plausible that the concepts, such as the components of high-quality CRD that emerged from this study will be of varying relevance in different populations. For instance, it is plausible that accessibility may be more important for persons with hectic schedules—such as young adults balancing establishing a career, growing a young family, maintaining a social schedule, and facing the challenges of a new cancer diagnosis—compared with individuals with fewer competing commitments. In addition, although the interpretation of CRD is presented here as a personal process, it has been demonstrated that some individuals may prefer to involve various members of their community, including elders, extended family, and/or spiritual leaders in decision making [81,82]. These 2 observations serve as a reminder that theory is not a substitute for engaging with the expected end users when developing interventions intended to help improve their cancer experience [83]. Finally, as this research study was conducted in the context of the cancer experience in Western Canada, it is certainly possible that how this framework applies to other areas in Canada, and the world for that matter, may differ. Researchers and clinicians are encouraged to explore how the framework presented here can be modified to best reflect the context of living with cancer in their area [23].

### Conclusions

The objective of this study is to develop a theoretical framework grounded in the cancer experience capable of guiding the development of informational resources. The framework that emerged links the quality of CRD received to the impact that the cancer diagnosis has on an individual's life. The theory comprises 4 variables: personal projects, cancer as a project that interferes with existing personal projects, information as the process of receiving and processing CRD to inform action, and CRD quality. Key features of high-quality CRD include accessibility, credibility, applicability, and framing. On the basis of this theory, the internet is foundational for delivering highly accessible information interventions. Clinicians are encouraged to consider accessibility and framing in how they provide information to those they care for. Future directions for research are expected to include engaging with those affected by cancer as partners to develop and integrate informational interventions based on this theory into clinical care. Interventions informed by this theoretical work are expected to help individuals remain effectively engaged with the personal projects in their lives following a cancer diagnosis and minimize the disruptive impact of the cancer diagnosis on patients and their informal caregivers by decreasing the cost of obtaining useful information.

### Acknowledgments

The authors would like to acknowledge the participants of this study who graciously shared their time, the details of their cancer experience, and their wisdom for the benefit of others.



## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Demographic or intake questionnaire used to guide selection of interview participants.

[DOCX File, 21 KB - [jmir\\_v22i10e20510\\_app1.docx](#)]

### Multimedia Appendix 2

Initial interview guide for semistructured interviews.

[DOCX File, 17 KB - [jmir\\_v22i10e20510\\_app2.docx](#)]

## References

1. Tariman JD, Doorenbos A, Schepp KG, Singhal S, Berry DL. Information needs priorities in patients diagnosed with cancer: a systematic review. *J Adv Pract Oncol* 2014;2014(5):115-122 [FREE Full text] [Medline: [24910808](#)]
2. Fletcher C, Flight I, Chapman J, Fennell K, Wilson C. The information needs of adult cancer survivors across the cancer continuum: a scoping review. *Patient Educ Couns* 2017 Mar;100(3):383-410. [doi: [10.1016/j.pec.2016.10.008](#)] [Medline: [27765377](#)]
3. Mistry A, Wilson S, Priestman T, Damery S, Haque M. How do the information needs of cancer patients differ at different stages of the cancer journey? A cross-sectional survey. *JRSM Short Rep* 2010 Sep 15;1(4):30 [FREE Full text] [doi: [10.1258/shorts.2010.010032](#)] [Medline: [21103122](#)]
4. Rutten LJ, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). *Patient Educ Couns* 2005 Jun;57(3):250-261. [doi: [10.1016/j.pec.2004.06.006](#)] [Medline: [15893206](#)]
5. Wang T, Molassiotis A, Chung BP, Tan J. Unmet care needs of advanced cancer patients and their informal caregivers: a systematic review. *BMC Palliat Care* 2018 Jul 23;17(1):96 [FREE Full text] [doi: [10.1186/s12904-018-0346-9](#)] [Medline: [30037346](#)]
6. Lambert SD, Harrison JD, Smith E, Bonevski B, Carey M, Laws C, et al. The unmet needs of partners and caregivers of adults diagnosed with cancer: a systematic review. *BMJ Support Palliat Care* 2012 Sep;2(3):224-230. [doi: [10.1136/bmjspcare-2012-000226](#)] [Medline: [24654195](#)]
7. Moghaddam N, Coxon H, Nabarro S, Hardy B, Cox K. Unmet care needs in people living with advanced cancer: a systematic review. *Support Care Cancer* 2016 Aug;24(8):3609-3622. [doi: [10.1007/s00520-016-3221-3](#)] [Medline: [27137214](#)]
8. Bonacchi A, Di Miceli S, Lippi D, Muraca MG, Miccinesi G. Unmet needs of Italian cancer patients in different stages of the disease and care process. *Tumori* 2018 Aug;104(4):285-291. [doi: [10.5301/tj.5000637](#)] [Medline: [28478643](#)]
9. Halkett GK, Kristjanson LJ, Lobb E, O'Driscoll C, Taylor M, Spry N. Meeting breast cancer patients' information needs during radiotherapy: what can we do to improve the information and support that is currently provided? *Eur J Cancer Care (Engl)* 2010 Jul;19(4):538-547. [doi: [10.1111/j.1365-2354.2009.01090.x](#)] [Medline: [19708930](#)]
10. Hubbeling HG, Rosenberg SM, González-Robledo MC, Cohn JG, Villarreal-Garza C, Partridge AH, et al. Psychosocial needs of young breast cancer survivors in Mexico City, Mexico. *PLoS One* 2018;13(5):e0197931 [FREE Full text] [doi: [10.1371/journal.pone.0197931](#)] [Medline: [29787612](#)]
11. McCarthy MC, McNeil R, Drew S, Orme L, Sawyer SM. Information needs of adolescent and young adult cancer patients and their parent-carers. *Support Care Cancer* 2018 May;26(5):1655-1664. [doi: [10.1007/s00520-017-3984-1](#)] [Medline: [29222598](#)]
12. Meredith C, Symonds P, Webster L, Lamont D, Pyper E, Gillis CR, et al. Information needs of cancer patients in west Scotland: cross sectional survey of patients' views. *Br Med J* 1996 Sep 21;313(7059):724-726 [FREE Full text] [doi: [10.1136/bmj.313.7059.724](#)] [Medline: [8819442](#)]
13. Preisler M, Rohrmoser A, Goerling U, Kendel F, Bär K, Riemer M, et al. Early palliative care for those who care: a qualitative exploration of cancer caregivers' information needs during hospital stays. *Eur J Cancer Care (Engl)* 2019 Mar;28(2):e12990. [doi: [10.1111/ecc.12990](#)] [Medline: [30623515](#)]
14. Vivar CG, McQueen A. Informational and emotional needs of long-term survivors of breast cancer. *J Adv Nurs* 2005 Sep;51(5):520-528. [doi: [10.1111/j.1365-2648.2005.03524.x](#)] [Medline: [16098169](#)]
15. Boons CC, Harbers L, Timmers L, de Jong J, Swart EL, Harry Hendrikse N, et al. Needs for information and reasons for (non)adherence in chronic myeloid leukaemia: be aware of social activities disturbing daily routines. *Eur J Haematol* 2018 Jul 30: - epub ahead of print. [doi: [10.1111/ejh.13155](#)] [Medline: [30058149](#)]
16. Arthurs G, Simpson J, Brown A, Kyaw O, Shyrier S, Concert CM. The effectiveness of therapeutic patient education on adherence to oral anti-cancer medicines in adult cancer patients in ambulatory care settings: a systematic review. *JBIC Database System Rev Implement Rep* 2015 Jun 12;13(5):244-292. [doi: [10.11124/jbisr-2015-2057](#)] [Medline: [26455611](#)]

17. Veroff D, Marr A, Wennberg DE. Enhanced support for shared decision making reduced costs of care for patients with preference-sensitive conditions. *Health Aff (Millwood)* 2013 Feb;32(2):285-293. [doi: [10.1377/hlthaff.2011.0941](https://doi.org/10.1377/hlthaff.2011.0941)] [Medline: [23381521](https://pubmed.ncbi.nlm.nih.gov/23381521/)]
18. Walsh T, Barr PJ, Thompson R, Ozanne E, O'Neill C, Elwyn G. Undetermined impact of patient decision support interventions on healthcare costs and savings: systematic review. *Br Med J* 2014 Jan 23;348:g188. [doi: [10.1136/bmj.g188](https://doi.org/10.1136/bmj.g188)] [Medline: [24458654](https://pubmed.ncbi.nlm.nih.gov/24458654/)]
19. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* 2017 Apr 12;4:CD001431 [FREE Full text] [doi: [10.1002/14651858.CD001431.pub5](https://doi.org/10.1002/14651858.CD001431.pub5)] [Medline: [28402085](https://pubmed.ncbi.nlm.nih.gov/28402085/)]
20. Ferrari M, Ripamonti CI, Hulbert-Williams NJ, Miccinesi G. Relationships among unmet needs, depression, and anxiety in non-advanced cancer patients. *Tumori* 2019 Apr;105(2):144-150. [doi: [10.1177/0300891618765546](https://doi.org/10.1177/0300891618765546)] [Medline: [29714666](https://pubmed.ncbi.nlm.nih.gov/29714666/)]
21. Ugalde A, Aranda S, Krishnasamy M, Ball D, Schofield P. Unmet needs and distress in people with inoperable lung cancer at the commencement of treatment. *Support Care Cancer* 2012 Feb;20(2):419-423. [doi: [10.1007/s00520-011-1296-4](https://doi.org/10.1007/s00520-011-1296-4)] [Medline: [22038481](https://pubmed.ncbi.nlm.nih.gov/22038481/)]
22. Thiessen M, Hack TF, Pitz M, Anderson M. A model of identity grounded in the acute season of survivorship. *Psychooncology* 2018 Oct;27(10):2412-2418. [doi: [10.1002/pon.4842](https://doi.org/10.1002/pon.4842)] [Medline: [30048034](https://pubmed.ncbi.nlm.nih.gov/30048034/)]
23. Glaser BG, Strauss AL. *The Discovery of Grounded Theory*. Chicago, USA: Aldine Pub Co; 1967.
24. Holton J, Walsh I. *Classic Grounded Theory: Applications With Qualitative and Quantitative Data*. Thousand Oaks, CA: Sage Publications; 2016.
25. Savolainen R. Modeling the interplay of information seeking and information sharing. *Aslib J Inf Manag* 2019 Jul 15;71(4):518-534. [doi: [10.1108/ajim-10-2018-0266](https://doi.org/10.1108/ajim-10-2018-0266)]
26. Zare-Farashbandi F, Lalazaryan A. A Review of models and theories of health information seeking behavior. *Int J Health Syst Disaster Manage* 2014;2(4):193. [doi: [10.4103/2347-9019.144371](https://doi.org/10.4103/2347-9019.144371)]
27. Lambert SD, Loisel CG. Health information seeking behavior. *Qual Health Res* 2007 Oct;17(8):1006-1019. [doi: [10.1177/1049732307305199](https://doi.org/10.1177/1049732307305199)] [Medline: [17928475](https://pubmed.ncbi.nlm.nih.gov/17928475/)]
28. Miller SM. Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. *J Pers Soc Psychol* 1987 Feb;52(2):345-353. [doi: [10.1037//0022-3514.52.2.345](https://doi.org/10.1037//0022-3514.52.2.345)] [Medline: [3559895](https://pubmed.ncbi.nlm.nih.gov/3559895/)]
29. Freimuth V, Stein J, Kean T. *Searching for Health Information: The Cancer Information Service Model*. Pennsylvania, USA: University of Pennsylvania Press; 1989.
30. Rees C, Bath P. The psychometric properties of the miller behavioural style scale with adult daughters of women with early breast cancer: a literature review and empirical study. *J Adv Nurs* 2000 Aug;32(2):366-374. [doi: [10.1046/j.1365-2648.2000.01485.x](https://doi.org/10.1046/j.1365-2648.2000.01485.x)] [Medline: [10964184](https://pubmed.ncbi.nlm.nih.gov/10964184/)]
31. Kola S, Walsh JC, Hughes BM, Howard S. Matching intra-procedural information with coping style reduces psychophysiological arousal in women undergoing colposcopy. *J Behav Med* 2013 Aug;36(4):401-412. [doi: [10.1007/s10865-012-9435-z](https://doi.org/10.1007/s10865-012-9435-z)] [Medline: [22729966](https://pubmed.ncbi.nlm.nih.gov/22729966/)]
32. Miller SM. Monitoring processing style: to see or not to see. *Eur J Pain* 2015 Apr;19(4):449-450 [FREE Full text] [doi: [10.1002/ejp.664](https://doi.org/10.1002/ejp.664)] [Medline: [25808064](https://pubmed.ncbi.nlm.nih.gov/25808064/)]
33. Biernatzki L, Kuske S, Genz J, Ritschel M, Stephan A, Bächle C, et al. Information needs in people with diabetes mellitus: a systematic review. *Syst Rev* 2018 Feb 14;7(1):27 [FREE Full text] [doi: [10.1186/s13643-018-0690-0](https://doi.org/10.1186/s13643-018-0690-0)] [Medline: [29444711](https://pubmed.ncbi.nlm.nih.gov/29444711/)]
34. Carey M, Lambert S, Smits R, Paul C, Sanson-Fisher R, Clinton-McHarg T. The unfulfilled promise: a systematic review of interventions to reduce the unmet supportive care needs of cancer patients. *Support Care Cancer* 2012 Feb;20(2):207-219 [FREE Full text] [doi: [10.1007/s00520-011-1327-1](https://doi.org/10.1007/s00520-011-1327-1)] [Medline: [22089430](https://pubmed.ncbi.nlm.nih.gov/22089430/)]
35. Clarke MA, Moore JL, Steege LM, Koopman RJ, Belden JL, Canfield SM, et al. Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: a literature review. *Health Informatics J* 2016 Dec;22(4):992-1016. [doi: [10.1177/1460458215602939](https://doi.org/10.1177/1460458215602939)] [Medline: [26377952](https://pubmed.ncbi.nlm.nih.gov/26377952/)]
36. Faury S, Koleck M, Foucaud J, M'Bailara K, Quintard B. Patient education interventions for colorectal cancer patients with stoma: a systematic review. *Patient Educ Couns* 2017 Oct;100(10):1807-1819. [doi: [10.1016/j.pec.2017.05.034](https://doi.org/10.1016/j.pec.2017.05.034)] [Medline: [28602564](https://pubmed.ncbi.nlm.nih.gov/28602564/)]
37. Howell D, Harth T, Brown J, Bennett C, Boyko S. Self-management education interventions for patients with cancer: a systematic review. *Support Care Cancer* 2017 Apr;25(4):1323-1355. [doi: [10.1007/s00520-016-3500-z](https://doi.org/10.1007/s00520-016-3500-z)] [Medline: [28058570](https://pubmed.ncbi.nlm.nih.gov/28058570/)]
38. Hyun YG, Alhashemi A, Fazlzad R, Goldberg AS, Goldstein DP, Sawka AM. A systematic review of unmet information and psychosocial support needs of adults diagnosed with thyroid cancer. *Thyroid* 2016 Sep;26(9):1239-1250. [doi: [10.1089/thy.2016.0039](https://doi.org/10.1089/thy.2016.0039)] [Medline: [27350421](https://pubmed.ncbi.nlm.nih.gov/27350421/)]
39. Kotronoulas G, Papadopoulou C, Burns-Cunningham K, Simpson M, Maguire R. A systematic review of the supportive care needs of people living with and beyond cancer of the colon and/or rectum. *Eur J Oncol Nurs* 2017 Aug;29:60-70. [doi: [10.1016/j.ejon.2017.05.004](https://doi.org/10.1016/j.ejon.2017.05.004)] [Medline: [28720267](https://pubmed.ncbi.nlm.nih.gov/28720267/)]
40. Qan'ir Y, Song L. Systematic review of technology-based interventions to improve anxiety, depression, and health-related quality of life among patients with prostate cancer. *Psychooncology* 2019 Aug;28(8):1601-1613. [doi: [10.1002/pon.5158](https://doi.org/10.1002/pon.5158)] [Medline: [31222956](https://pubmed.ncbi.nlm.nih.gov/31222956/)]



41. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. *Cogent Social Sciences* 2017 Mar 13;3(1):379-386. [doi: [10.1080/23311886.2017.1302785](https://doi.org/10.1080/23311886.2017.1302785)]
42. Saab MM, Reidy M, Hegarty J, O'Mahony M, Murphy M, Von Wagner C, et al. Men's information-seeking behavior regarding cancer risk and screening: a meta-narrative systematic review. *Psychooncology* 2018 Feb;27(2):410-419. [doi: [10.1002/pon.4506](https://doi.org/10.1002/pon.4506)] [Medline: [28728212](https://pubmed.ncbi.nlm.nih.gov/28728212/)]
43. Rutherford C, Mercieca-Bebber R, Butow P, Wu JL, King MT. Treatment decision-making in ductal carcinoma in situ: a mixed methods systematic review of women's experiences and information needs. *Patient Educ Couns* 2017 Sep;100(9):1654-1666. [doi: [10.1016/j.pec.2017.04.009](https://doi.org/10.1016/j.pec.2017.04.009)] [Medline: [28442156](https://pubmed.ncbi.nlm.nih.gov/28442156/)]
44. Heath G, Cooke R, Cameron E. A theory-based approach for developing interventions to change patient behaviours: a medication adherence example from paediatric secondary care. *Healthcare (Basel)* 2015 Dec 4;3(4):1228-1242 [FREE Full text] [doi: [10.3390/healthcare3041228](https://doi.org/10.3390/healthcare3041228)] [Medline: [27417822](https://pubmed.ncbi.nlm.nih.gov/27417822/)]
45. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new medical research council guidance. *Br Med J* 2008 Sep 29;337:a1655 [FREE Full text] [doi: [10.1136/bmj.a1655](https://doi.org/10.1136/bmj.a1655)] [Medline: [18824488](https://pubmed.ncbi.nlm.nih.gov/18824488/)]
46. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. *Health Psychol* 2010 Jan;29(1):1-8. [doi: [10.1037/a0016939](https://doi.org/10.1037/a0016939)] [Medline: [20063930](https://pubmed.ncbi.nlm.nih.gov/20063930/)]
47. Longo DR, Ge B, Radina ME, Greiner A, Williams CD, Longo GS, et al. Understanding breast-cancer patients' perceptions: health information-seeking behaviour and passive information receipt. *J Commun Healthcare* 2013 Jul 18;2(2):184-206. [doi: [10.1179/cih.2009.2.2.184](https://doi.org/10.1179/cih.2009.2.2.184)]
48. Longo DR. Understanding health information, communication, and information seeking of patients and consumers: a comprehensive and integrated model. *Health Expect* 2005 Sep;8(3):189-194 [FREE Full text] [doi: [10.1111/j.1369-7625.2005.00339.x](https://doi.org/10.1111/j.1369-7625.2005.00339.x)] [Medline: [16098149](https://pubmed.ncbi.nlm.nih.gov/16098149/)]
49. Longo DR, Patrick TB, Kruse RL. The natural history of the use of healthcare information by women with breast cancer: a conceptual model. *Proc AMIA Symp* 2001:413-417 [FREE Full text] [Medline: [11825221](https://pubmed.ncbi.nlm.nih.gov/11825221/)]
50. Galloway S, Graydon J, Harrison D, Evans-Boyd B, Palmer-Wickham S, Burlein-Hall S, et al. Informational needs of women with a recent diagnosis of breast cancer: development and initial testing of a tool. *J Adv Nurs* 1997 Jun;25(6):1175-1183. [doi: [10.1046/j.1365-2648.1997.19970251175.x](https://doi.org/10.1046/j.1365-2648.1997.19970251175.x)] [Medline: [9181414](https://pubmed.ncbi.nlm.nih.gov/9181414/)]
51. Glaser B, Strauss A. *The Discovery of Grounded Theory*. New York, USA: Aldine Transaction; 2009.
52. Suzana Mediani H. An introduction to classical grounded theory. *SOJ Nurs Health Care* 2017 Nov 6;3(3):1-5. [doi: [10.15226/2471-6529/3/3/00135](https://doi.org/10.15226/2471-6529/3/3/00135)]
53. Rieger KL. Discriminating among grounded theory approaches. *Nurs Inq* 2019 Jan;26(1):e12261 [FREE Full text] [doi: [10.1111/nin.12261](https://doi.org/10.1111/nin.12261)] [Medline: [30123965](https://pubmed.ncbi.nlm.nih.gov/30123965/)]
54. Simmons O. *Why Classic Ground Theory*. New York, USA: BrownWalker Press; 2011.
55. Twombly R. What's in a name: who is a cancer survivor? *J Natl Cancer Inst* 2004 Oct 6;96(19):1414-1415. [doi: [10.1093/jnci/96.19.1414](https://doi.org/10.1093/jnci/96.19.1414)] [Medline: [15467027](https://pubmed.ncbi.nlm.nih.gov/15467027/)]
56. Rogers E, Kincaid D. *Communication Networks: Toward a New Paradigm for Research*. New York, USA: Free Press; 1981.
57. Benner P. *The Phenomenon of Care*. In: Toombs S, editor. *Handbook of Phenomenology and Medicine*. Netherlands: Springer; 2001.
58. Jagannathan A, Juvva S. Emotions and coping of patients with head and neck cancers after diagnosis: a qualitative content analysis. *J Postgrad Med* 2016;62(3):143-149 [FREE Full text] [doi: [10.4103/0022-3859.184273](https://doi.org/10.4103/0022-3859.184273)] [Medline: [27320951](https://pubmed.ncbi.nlm.nih.gov/27320951/)]
59. Thornton LM, Levin AO, Dorfman CS, Godiwala N, Heitzmann C, Andersen BL. Emotions and social relationships for breast and gynecologic patients: a qualitative study of coping with recurrence. *Psychooncology* 2014 Apr;23(4):382-389 [FREE Full text] [doi: [10.1002/pon.3429](https://doi.org/10.1002/pon.3429)] [Medline: [24123502](https://pubmed.ncbi.nlm.nih.gov/24123502/)]
60. Lerner JS, Li Y, Valdesolo P, Kassam KS. Emotion and decision making. *Annu Rev Psychol* 2015 Jan 3;66:799-823. [doi: [10.1146/annurev-psych-010213-115043](https://doi.org/10.1146/annurev-psych-010213-115043)] [Medline: [25251484](https://pubmed.ncbi.nlm.nih.gov/25251484/)]
61. Mazzocco K, Masiero M, Carriero MC, Pravettoni G. The role of emotions in cancer patients' decision-making. *Ecancermedicalscience* 2019;13:914 [FREE Full text] [doi: [10.3332/ecancer.2019.914](https://doi.org/10.3332/ecancer.2019.914)] [Medline: [31123497](https://pubmed.ncbi.nlm.nih.gov/31123497/)]
62. McAlpine K, Lewis KB, Trevena LJ, Stacey D. What is the effectiveness of patient decision aids for cancer-related decisions? A systematic review subanalysis. *JCO Clin Cancer Inform* 2018 Dec;2:1-13. [doi: [10.1200/CCI.17.00148](https://doi.org/10.1200/CCI.17.00148)] [Medline: [30652610](https://pubmed.ncbi.nlm.nih.gov/30652610/)]
63. DrugTherapeutics Bulletin. An introduction to patient decision aids. *Br Med J* 2013 Jul 23;347:f4147. [doi: [10.1136/bmj.f4147](https://doi.org/10.1136/bmj.f4147)] [Medline: [23881944](https://pubmed.ncbi.nlm.nih.gov/23881944/)]
64. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. *J Am Med Inform Assoc* 2016 Nov;23(6):1053-1059 [FREE Full text] [doi: [10.1093/jamia/ocv204](https://doi.org/10.1093/jamia/ocv204)] [Medline: [27206459](https://pubmed.ncbi.nlm.nih.gov/27206459/)]
65. Statistics C. Canadian Internet Use Survey. Onlin 2020:e2019 [FREE Full text]

66. Ashkanani H, Asery R, Bokubar F, AlAli N, Mubarak S, Buabbas A, et al. Web-based health information seeking among students at Kuwait university: cross-sectional survey study. *JMIR Form Res* 2019 Oct 31;3(4):e14327 [[FREE Full text](#)] [doi: [10.2196/14327](#)] [Medline: [31473592](#)]
67. van der Maas M, Shi J, Elton-Marshall T, Hodgins DC, Sanchez S, Lobo DS, et al. Internet-based interventions for problem gambling: scoping review. *JMIR Ment Health* 2019 Jan 7;6(1):e65 [[FREE Full text](#)] [doi: [10.2196/mental.9419](#)] [Medline: [30617046](#)]
68. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. *J Med Internet Res* 2017 Jun 19;19(6):e218 [[FREE Full text](#)] [doi: [10.2196/jmir.7579](#)] [Medline: [28630033](#)]
69. Rolfe A, Cash-Gibson L, Car J, Sheikh A, McKinstry B. Interventions for improving patients' trust in doctors and groups of doctors. *Cochrane Database Syst Rev* 2014 Mar 4(3):CD004134 [[FREE Full text](#)] [doi: [10.1002/14651858.CD004134.pub3](#)] [Medline: [24590693](#)]
70. Williams CP, Azuero A, Kenzik KM, Pisu M, Nipp RD, Bhatia S, et al. Guideline discordance and patient cost responsibility in medicare beneficiaries with metastatic breast cancer. *J Natl Compr Canc Netw* 2019 Oct 1;17(10):1221-1228. [doi: [10.6004/jnccn.2019.7316](#)] [Medline: [31590153](#)]
71. Williams CP, Kenzik KM, Azuero A, Williams GR, Pisu M, Halilova KI, et al. Guideline discordance and patient cost responsibility in medicare beneficiaries with metastatic breast cancer. *Oncologist* 2019 Jan;24(1):31-37 [[FREE Full text](#)] [doi: [10.1634/theoncologist.2018-0076](#)] [Medline: [30120157](#)]
72. Han H, Gleason KT, Sun C, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. *JMIR Hum Factors* 2019 Dec 19;6(4):e15038 [[FREE Full text](#)] [doi: [10.2196/15038](#)] [Medline: [31855187](#)]
73. Pho K, Lu R, Gates S, Xie Y, Lee SJ, Gerber DE. Characteristics of patients using patient portals in oncology. *JAMA Oncol* 2018 Mar 1;4(3):416-418 [[FREE Full text](#)] [doi: [10.1001/jamaoncol.2017.5257](#)] [Medline: [29372242](#)]
74. Alpert JM, Morris BB, Thomson MD, Matin K, Brown RF. Implications of patient portal transparency in oncology: qualitative interview study on the experiences of patients, oncologists, and medical informaticists. *JMIR Cancer* 2018 Mar 26;4(1):e5 [[FREE Full text](#)] [doi: [10.2196/cancer.8993](#)] [Medline: [29581090](#)]
75. Hack TF, Ruether JD, Weir LM, Grenier D, Degner LF. Study protocol: addressing evidence and context to facilitate transfer and uptake of consultation recording use in oncology: a knowledge translation implementation study. *Implement Sci* 2011 Mar 14;6:20 [[FREE Full text](#)] [doi: [10.1186/1748-5908-6-20](#)] [Medline: [21401958](#)]
76. Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. SPIKES-a six-step protocol for delivering bad news: application to the patient with cancer. *Oncologist* 2000;5(4):302-311. [doi: [10.1634/theoncologist.5-4-302](#)] [Medline: [10964998](#)]
77. Girgis A, Sanson-Fisher RW. Breaking bad news: consensus guidelines for medical practitioners. *J Clin Oncol* 1995 Sep;13(9):2449-2456. [doi: [10.1200/JCO.1995.13.9.2449](#)] [Medline: [7666105](#)]
78. Rat A, Ricci L, Guillemin F, Ricatte C, Pongy M, Vieux R, et al. Development of a web-based formative self-assessment tool for physicians to practice breaking bad news (BRADNET). *JMIR Med Educ* 2018 Jul 19;4(2):e17 [[FREE Full text](#)] [doi: [10.2196/mededu.9551](#)] [Medline: [30026180](#)]
79. Mallidou AA, Frisch N, Doyle-Waters MM, MacLeod ML, Ward J, Atherton P. Patient-oriented research competencies in health (PORCH) for patients, healthcare providers, decision-makers and researchers: protocol of a scoping review. *Syst Rev* 2018 Jul 19;7(1):101 [[FREE Full text](#)] [doi: [10.1186/s13643-018-0762-1](#)] [Medline: [30025543](#)]
80. Birks M, Mills J. Grounded Theory: A Practical Guide. Thousand Oaks, CA: Sage Publications; 2010.
81. Dolan H, Alden DL, Friend JM, Lee PY, Lee YK, Ng CJ, et al. Culture, self, and medical decision making in Australia and China: a structural model analysis. *MDM Policy Pract* 2019;4(2):2381468319871018. [doi: [10.1177/2381468319871018](#)] [Medline: [31565670](#)]
82. Hawley ST, Morris AM. Cultural challenges to engaging patients in shared decision making. *Patient Educ Couns* 2017 Jan;100(1):18-24 [[FREE Full text](#)] [doi: [10.1016/j.pec.2016.07.008](#)] [Medline: [27461943](#)]
83. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. *Health Aff (Millwood)* 2013 Feb;32(2):223-231. [doi: [10.1377/hlthaff.2012.1133](#)] [Medline: [23381514](#)]

## Abbreviations

**CGT:** classic grounded theory

**CRD:** cancer-related data

**HIA:** health information acquisition

**HISB:** health information-seeking behavior

**MBSS:** Miller Behavioral Style Scale

**PDA:** patient decision aid

*Edited by G Eysenbach; submitted 01.06.20; peer-reviewed by S Atanasova, C Gabarre; comments to author 04.07.20; revised version received 29.08.20; accepted 17.09.20; published 29.10.20.*

*Please cite as:*

Thiessen M, Sinclair S, Tang PA, Raffin Bouchal S

*Information Access and Use by Patients With Cancer and Their Friends and Family: Development of a Grounded Theory*

*J Med Internet Res* 2020;22(10):e20510

URL: <http://www.jmir.org/2020/10/e20510/>

doi: [10.2196/20510](https://doi.org/10.2196/20510)

PMID: [33118940](https://pubmed.ncbi.nlm.nih.gov/33118940/)

©Maclean Thiessen, Shane Sinclair, Patricia A Tang, Shelley Raffin Bouchal. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# A Web Application About Herd Immunity Using Personalized Avatars: Development Study

Hina Hakim<sup>1</sup>, BDS, MSc; Julie A Bettinger<sup>2</sup>, PhD, MPH; Christine T Chambers<sup>3</sup>, PhD; S Michelle Driedger<sup>4</sup>, PhD; Eve Dubé<sup>5</sup>, PhD; Teresa Gavaruzzi<sup>6</sup>, PhD; Anik M C Giguere<sup>1</sup>, PhD; Éric Kavanagh<sup>7</sup>, MA; Julie Leask<sup>8</sup>, Dip Health Sci, MPH, PhD; Shannon E MacDonald<sup>9</sup>, RN, PhD; Rita Orji<sup>10</sup>, PhD; Elizabeth Parent<sup>1</sup>, MA; Jean-Sébastien Paquette<sup>1</sup>, MSc, CFMC, MD; Jacynthe Roberge<sup>7</sup>, MA; Beate Sander<sup>11</sup>, RN, MBA, MEcDev, PhD; Aaron M Scherer<sup>12</sup>, PhD; Martin Tremblay-Breault<sup>1</sup>, BSc, MSc; Kumanan Wilson<sup>13</sup>, MSc, FRCPC, MD; Daniel Reinhartz<sup>14</sup>, PhD, MD; Holly O Witteman<sup>1</sup>, PhD

<sup>1</sup>Department of Family and Emergency Medicine, Laval University, Quebec City, QC, Canada

<sup>2</sup>Vaccine Evaluation Center, BC Children's Hospital, University of British Columbia, Vancouver, BC, Canada

<sup>3</sup>Department of Psychology and Neuroscience and Pediatrics, Dalhousie University, Halifax, NS, Canada

<sup>4</sup>Department of Community Health Sciences, University of Manitoba, Winnipeg, MB, Canada, Winnipeg, MB, Canada

<sup>5</sup>Institut national de santé publique du Québec, Institut national de santé publique du Québec, Quebec City, QC, Canada

<sup>6</sup>Department of Developmental Psychology and Socialization, University of Padova, Italy, Padova, Italy

<sup>7</sup>École de design, Édifice La Fabrique, Laval University, Quebec City, QC, Canada

<sup>8</sup>Faculty of Medicine and Health, Susan Wakil School of Nursing and Midwifery, University of Sydney, Sydney, Australia

<sup>9</sup>Faculty of Nursing, University of Alberta, Edmonton, AB, Canada

<sup>10</sup>Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada

<sup>11</sup>University Health Network, Toronto General Hospital, Eaton Building, Toronto, ON, Canada

<sup>12</sup>Department of Internal Medicine, University of Iowa, Iowa, IA, United States

<sup>13</sup>Department of Medicine, Bruyere Research Institute and Ottawa Hospital Research Institute, University of Ottawa, Ottawa, ON, Canada

<sup>14</sup>Department of Social and Preventive Medicine, Laval University, Quebec City, QC, Canada

## Corresponding Author:

Holly O Witteman, PhD

Department of Family and Emergency Medicine

Laval University

1050 avenue de la Médecine, Pavillon Ferdinand-Vandry 2881

Université Laval

Quebec City, QC, G1V 0A6

Canada

Phone: 1 418 656 2131 ext 3981

Email: [holly.witteman@fmed.ulaval.ca](mailto:holly.witteman@fmed.ulaval.ca)

## Abstract

**Background:** *Herd immunity* or *community immunity* refers to the reduced risk of infection among susceptible individuals in a population through the presence and proximity of immune individuals. Recent studies suggest that improving the understanding of community immunity may increase intentions to get vaccinated.

**Objective:** This study aims to design a web application about community immunity and optimize it based on users' cognitive and emotional responses.

**Methods:** Our multidisciplinary team developed a web application about community immunity to communicate epidemiological evidence in a personalized way. In our application, people build their own community by creating an avatar representing themselves and 8 other avatars representing people around them, for example, their family or coworkers. The application integrates these avatars in a 2-min visualization showing how different parameters (eg, vaccine coverage, and contact within communities) influence community immunity. We predefined communication goals, created prototype visualizations, and tested four iterative versions of our visualization in a university-based human-computer interaction laboratory and community-based settings (a cafeteria, two shopping malls, and a public library). Data included psychophysiological measures (eye tracking, galvanic skin

response, facial emotion recognition, and electroencephalogram) to assess participants' cognitive and affective responses to the visualization and verbal feedback to assess their interpretations of the visualization's content and messaging.

**Results:** Among 110 participants across all four cycles, 68 (61.8%) were women and 38 (34.5%) were men (4/110, 3.6%; not reported), with a mean age of 38 (SD 17) years. More than half (65/110, 59.0%) of participants reported having a university-level education. Iterative changes across the cycles included adding the ability for users to create their own avatars, specific signals about who was represented by the different avatars, using color and movement to indicate protection or lack of protection from infectious disease, and changes to terminology to ensure clarity for people with varying educational backgrounds. Overall, we observed 3 generalizable findings. First, visualization does indeed appear to be a promising medium for conveying what community immunity is and how it works. Second, by involving multiple users in an iterative design process, it is possible to create a short and simple visualization that clearly conveys a complex topic. Finally, evaluating users' emotional responses during the design process, in addition to their cognitive responses, offers insights that help inform the final design of an intervention.

**Conclusions:** Visualization with personalized avatars may help people understand their individual roles in population health. Our app showed promise as a method of communicating the relationship between individual behavior and community health. The next steps will include assessing the effects of the application on risk perception, knowledge, and vaccination intentions in a randomized controlled trial. This study offers a potential road map for designing health communication materials for complex topics such as community immunity.

(*J Med Internet Res* 2020;22(10):e20113) doi:[10.2196/20113](https://doi.org/10.2196/20113)

## KEYWORDS

community immunity; herd immunity; vaccination; vaccine hesitancy; avatar; web application

## Introduction

### Background

*Herd immunity* or *community immunity* refers to the reduced risk of transmission of infection among susceptible individuals in a population through the presence and proximity of immune individuals. Community immunity (the term we use throughout this paper) works by breaking the chain of transmission and decreasing the probability of contact with an infectious agent, thereby preventing the spread of infectious agents in susceptible populations [1,2]. High vaccination coverage is generally needed to achieve this protection at the population level [3]. Decisions not to vaccinate affect population-level vaccine coverage and can result in outbreaks of vaccine-preventable diseases by pushing the vaccine coverage rate below the community immunity threshold [4-6].

Although some research suggests that people's immunization decisions are primarily influenced by perceived benefits and harm at the individual level rather than those at the community level [7], other studies have suggested that improving the understanding of community immunity may lead to an increase in the intention to be vaccinated [8-10].

Community immunity is a challenging concept to convey. It depends on multiple factors, namely, vaccine effectiveness and coverage, whether or not susceptible individuals form clusters, timing of vaccine administration (ie, delayed vaccination results in longer periods of susceptibility and therefore increased likelihood of infection), and the presence or absence of serotype replacement [11]. It is also affected by historical rates of vaccination coverage where there are potential immunity gaps among people in specific age groups (eg, adolescents and young adults for MMR [measles, mumps, and rubella] vaccines). Possibly because the interplay between all these variables is complicated, people demonstrate an uneven understanding of

the connection between individual-level vaccination behavior and community-level risk and benefits [12].

A systematic review identified visualization as a promising avenue for communicating the complex concept of community immunity [13]. By visualization, we mean the visual presentation of data or information. Visualization is a powerful communication mechanism because it enables people to rapidly understand complex information [14]. In this paper, we use the term *visualization* to refer to a brief narrated animation about community immunity. We use the term *application* when referring to a complete web-based application, combining the visualization with an interactive section in which people make their own avatars.

### Objectives

In this study, we seek to iteratively develop an application about community immunity that would be understood by people with varying levels of education and to assess and optimize people's cognitive and emotional responses to the application. Our focus included emotions because emotions influence people's decisions [15-17], including health decisions [18-20], and specifically vaccination decisions [21,22].

Our study aims to determine (1) whether and how people attend to different visual elements to explain the concept of community immunity (what is community immunity and how it works), (2) whether these elements are understandable, and (3) whether people understand how community immunity safeguards people, especially vulnerable populations who cannot be vaccinated or who may not respond to vaccines owing to their age or suppressed immune system.

## Methods

### Ethics Approval and Consent to Participate

This project was approved by 'Comité d'éthique de la recherche en sciences de la santé' ethics committee of Laval University



(approval no: 2017-137 R-2/15-07-2019). All participants provided written informed consent.

### Concept Map

Before designing the first prototype, our multidisciplinary team began by developing a concept map [23] of what the prototype should convey ([Multimedia Appendix 1](#)). Concept maps are defined as tools for organizing and representing knowledge [24] or a graphical representation of different concepts and the relationship between those concepts [25]. Our concept map was used to organize the underlying content presented in the visualization within three major themes: (1) community, (2) infection, and (3) vaccines. We expanded and refined the components of each theme throughout the study in response to participants' feedback. The theme *community* included content about how a community is made up of individuals, including vulnerable people living among other individuals. The theme *infections* included content about how different pathogens cause different infections and spread at different rates. The theme *vaccines* included content about how effective vaccines may or may not be, how some vaccine effectiveness may wane over time, and how different diseases require different vaccine coverage to prevent the spread of infection and to create community immunity.

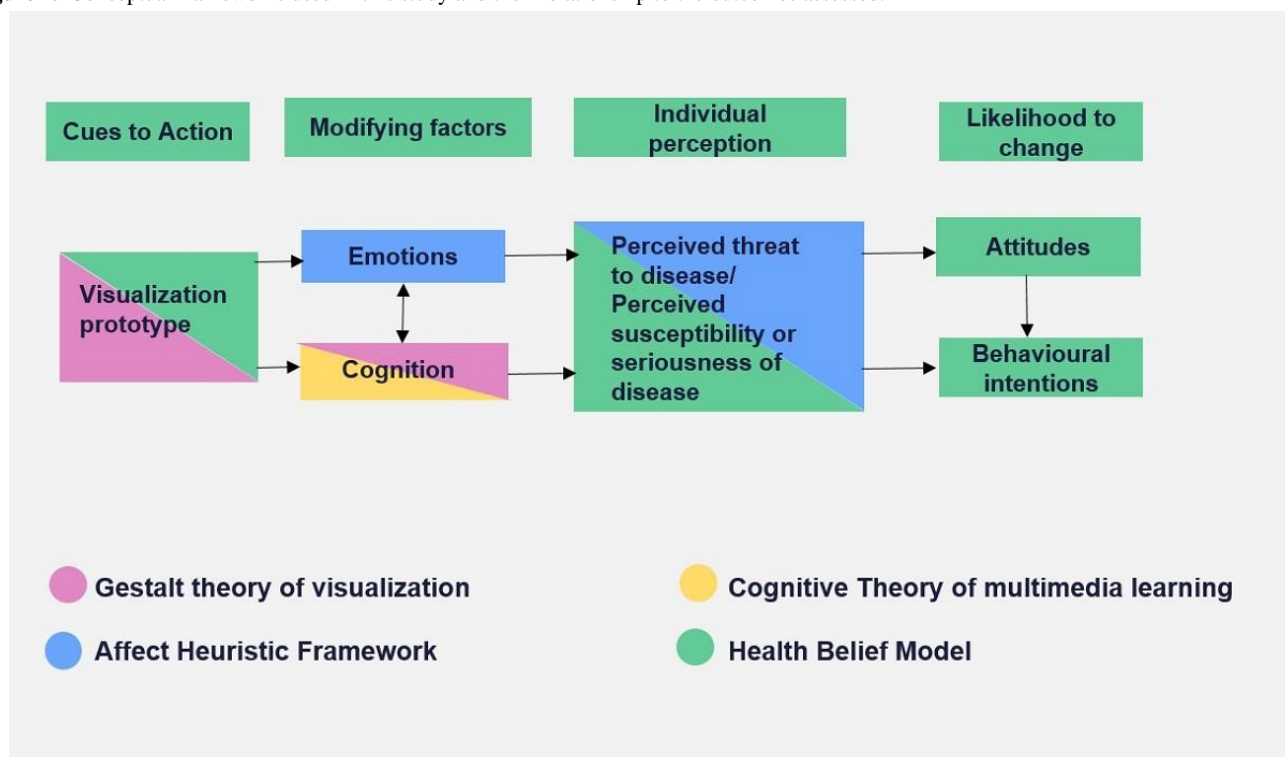
### Overall Approach

We developed our prototype application according to the concept map and predefined our communication goals for each element of the prototype. Each element of the prototype was linked to what it was intended to convey in the concept map, and what cognitive and/or affective (emotional) responses we aimed to evoke among participants. Across multiple iterative cycles ([Multimedia Appendices 2-5](#)), we then measured participants' responses to assess the extent to which each element of our application met its associated communication goals. In each cycle, we further sought to understand participants' needs, strengths, and limitations; observe how they attended to visual elements and colors; and identify potential improvements that could be made to the application.

### Framework

To design our application and interpret people's responses, we developed an integrated framework, as shown in [Figure 1](#), combining four existing frameworks or models: (1) the *Health Belief Model* [26], (2) *Gestalt visual principles* [27], (3) the *Cognitive Theory of Multimedia Learning* [28,29], and (4) *Affect Heuristic* [30,31].

**Figure 1.** Conceptual frameworks used in this study and their relationship to the outcomes assessed.



We selected the *Health Belief Model* [26] as the most likely framework to help us understand potential health behavior as a result of exposure to our intervention. This model has been developed and used to assess behavioral changes among people. However, this model hypothesizes that the intention or likelihood of an individual to take action stems from individual perception, and there is less detail regarding how such perceptions are shaped by different cues to action. We augmented this model to better understand the antecedents of

perception by using *Gestalt visual principles* to inform the design of our visualization. Gestalt visual principles emphasize that the whole cannot be determined by simply knowing the individual pieces but emerges through how the pieces are combined or structured. These principles can be used to understand how the structure, configuration, or layout of elements in a visualization influence how people perceive the visualization. For example, the figure-ground principle describes how humans perceive objects or figures according to the contrast

between elements and their backgrounds, and the proximity principle describes how images or figures located near each other are considered as a part of the same group, whereas objects apart are perceived as separate. Gestalt visual principles can thus help predict the effects of spacing, timing, and configuration when presenting information visually [32]. The *Cognitive Theory of Multimedia Learning* describes how people learn via two channels—auditory and visual—and use both together to build mental representations from words (audio) and images (visual) [28,29]. Finally, the *Affect Heuristic* provides an explicit framework for how the experiential system influences decisions via affect and emotions. The experiential system encodes reality in images, metaphors, and narratives, to which people have affective responses [31]. The Affect Heuristic helps structure analyses of emotions in response to the visualization.

Our guiding methodological framework was that of a user-centered design [33] in which potential users are consulted early and often, with their responses to prototype versions serving to help guide iterative improvements of the intervention or tool.

### Study Participants and Setting

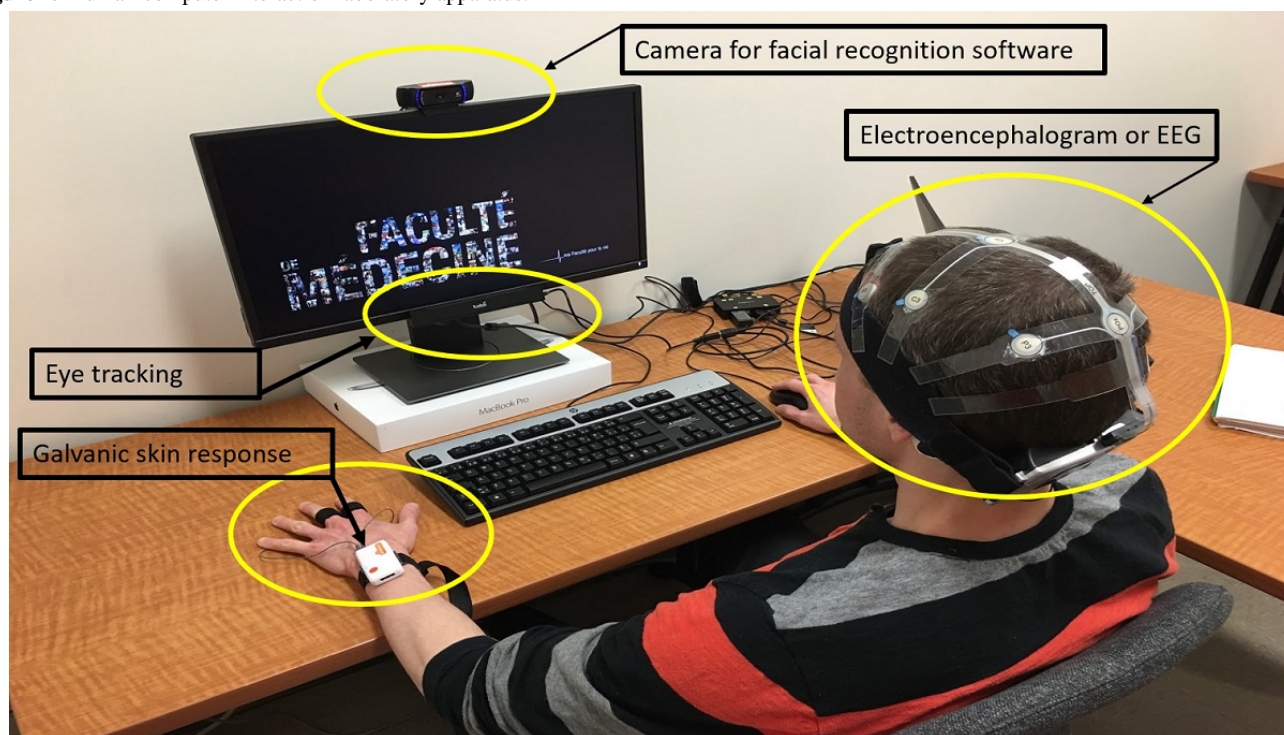
Across all four study cycles, we recruited participants who were aged 18 years or older, had either no vision problems or corrected vision problems (eg, using eyeglasses or contact lenses), and were able to provide written informed consent, read and understand French or English, and use computers. In cycles 1, 3, and 4, we recruited participants to come to our university-based human-computer interaction laboratory by sending invitations to a university-wide listserv directed at all students, staff, and others. In cycle 2, we recruited participants in person by approaching them at a university-based cafeteria. In cycle 3, in addition to the listserv recruitment, we also recruited participants in person at a public library and two shopping malls located in areas of the city whose postal codes are associated with more diverse educational backgrounds. An incentive of either Can \$10 (US \$7.46; cycles 1, 2, 4) or Can \$20 (US \$14.92; cycle 3) was offered for their time and any transportation costs incurred. In cycle 3, we offered a larger incentive because, after viewing our visualization, participants subsequently interacted with materials developed for other studies, meaning that the individual sessions were of a longer duration.

### Psychophysiological Measurement

Design cycles 1, 3, and 4 used four psychophysiological data collection methods: eye tracking, galvanic skin response, electroencephalogram (EEG), and facial emotion recognition. We used eye tracking to determine what people were looking at and to measure participants' visual attention [34]. We used galvanic skin response to determine when participants experienced peaks in emotional arousal [35]. Such peaks indicate instances of strong emotions. We expected the visualization to elicit strong emotions when, for example, something alarming happened, such as a vulnerable person getting infected with a contagious disease. We used facial emotion recognition software to assess emotional valence (ie, whether emotions were positive or negative) [36]. We expected participants' emotions to be positive when the visualization depicted positive things happening, for example, community immunity being achieved and protecting community members, and to be negative when the visualization depicted negative things happening, for example, an infection spreading in the community. We used EEG to assess participants' cognitive workload and engagement while looking at the information provided in the visualization [37]. We aimed for participants to experience higher engagement when interacting with the visualization without exceeding a cognitive load threshold above which they might be less likely to process new information.

### Apparatus and Procedures

As shown in Figure 2, participants sat in a stationary chair in front of a desk with a mobile eye tracker (Tobii X2-30) and a webcam mounted on the computer monitor, a keyboard, a mouse, and computer speakers. A member of the research team explained each participant-worn device while placing it. These participant-worn devices were a portable galvanic skin response apparatus (Shimmer Sensing Shimmer3 GSR+) worn on the participant's nondominant hand and an EEG (Advanced Brain Monitoring B-Alert X-Series) fitted on the participant's head, using a gel on the electrodes. We followed standard procedures for each device's calibration [38-41]. Data streams for all devices were synchronized and saved using the iMotions Attention Tool version 7 (cycle 1) or version 8 (cycles 3 and 4) [42].

**Figure 2.** Human-computer interaction laboratory apparatus.

### Verbal Feedback

We complemented psychophysiological data on participants' nonverbal reactions with brief verbal feedback. Using semistructured interview questions ([Multimedia Appendices 6-8](#)), we asked participants to summarize in their own words what they saw in the visualization, what message it aimed to convey, and anything they found confusing or unclear. They

were also asked questions about how to improve the visualization or personalized avatar building. If their explanation about the visualization indicated that they may have missed some visual elements, we probed for more specific information on how to improve those visual elements. We recorded responses using an audio recorder and took notes. [Table 1](#) shows the summarized study design.

**Table 1.** Summarized study design.

Cycles	Study setting	Sample size	Method for data collection
First cycle	University-based human-computer interaction laboratory via university-wide listserv (email)	n=8	Psychophysiological measurement and verbal feedback
Second cycle	University-based cafeteria (by approaching them)	n=11	Verbal feedback
Third cycle	<ul style="list-style-type: none"> <li>University sample: university-based human-computer interaction laboratory via university-wide listserv (email)</li> <li>Community sample: a public library and two shopping malls (by approaching them)</li> </ul>	<ul style="list-style-type: none"> <li>University sample: n=49</li> <li>Community sample: n=34</li> </ul>	Psychophysiological measurement and verbal feedback
Fourth cycle	University-based human-computer interaction laboratory via university-wide listserv (email)	n=8	Psychophysiological measurement (eye-tracking only) and verbal feedback

### Analysis

Our analytical aim was to assess whether the application achieved its communication goals. To analyze *psychophysiological measurements*, we examined participants' reactions to each element according to its associated communication goal. We first identified the periods of each element in the visualization according to the voice-over timing. We assessed whether the participant was visually attending to each element by defining an area of interest for each element (eg, a rectangular region around a symbol) and examining whether the participant had any eye fixations of 200 ms or more

in that area of interest. Fixations are described in the literature as lasting from 100 to 500 ms [43,44], 150 to 600 ms [45], or as low as 100 ms but *typically 200 to 600 ms* [46]. We selected 200 ms to maximize the likelihood of detecting fixations among people viewing a rapidly moving visualization while avoiding contaminating our data with shorter pauses in eye movement that might not indicate the person extracting any visual information. During the times when the element was present, we then also examined galvanic skin response, facial emotion, and EEG data as predefined for each communication goal. To analyze the galvanic skin response, we used an algorithm to detect peaks in arousal [47]. Previous literature suggests that



this algorithm performs well in detecting such peaks [48,49]. To account for known lags in galvanic skin response (ie, the fact that skin response lags behind experience of heightened arousal by 3 to 5 s [50]), we inspected data for peaks in arousal during the defined time for each communication goal and for an additional 5 s afterward. The existence of such peaks would indicate that the participant experienced a heightened emotion of some kind while that element was displayed. For instance, in the first cycle, some participants showed a peak in arousal when the visualization showed vulnerable people getting infected. To analyze facial emotions, we used the facial recognition software FACET (Emotient) within the iMotions Attention Tool [51]. This software uses algorithms to translate the movement of facial features, such as eyes, eye corners, brows, mouth corners, and nose tip, into classifications of emotional valence. Recent work suggests that this automated facial-expression analysis software performs well for detecting emotional states [52,53]. We inspected the aggregated data for the number of occurrences across all respondents, and for any positive, negative, or neutral emotional valence elicited by the visualization. To analyze the EEG data, we used algorithms to estimate participants' cognitive workload and engagement [39]. Cognitive workload indicates the extent to which working memory is being used. Engagement indicates a participant's attentiveness while watching the visualization. Previous studies have validated these algorithms for measuring cognitive workload and engagement [54-56]. Cognitive workload is reported on a continuous scale from 0 to 1, with 0 to 0.4 classified as boredom, 0.4 to 0.7 as optimal workload, and 0.7 and above as information overload. Engagement levels are also reported on a continuous scale from 0 to 1, with 0 to 0.1 classified as sleepiness and drowsiness, 0.3 as distraction, 0.6 as low engagement, and 0.6 to 1 as high engagement. A summary score was computed by averaging values for each communication goal across all participants. For cycles with fewer than 10 participants, we examined emotional valence and EEG data at the individual level only. For cycles with 10 or more participants, to summarize data while continuing to weigh data from each participant equally, we calculated the mean valence, cognitive load, and engagement for each participant for each element, and then computed summary statistics and indices of dispersion across all participants. When these mean values were normally distributed across participants, our summary statistic was a global sample mean and our index of dispersion was a sample SD. When these mean values were skewed across participants, our summary statistic was a global sample median and our index of dispersion was an IQR. In addition to analyses by area of interest, we also inspected the heat maps of full screens. Heat maps are visual representations of data showing the relative intensity of participants' visual attention to see where participants are looking at the most.

To analyze *verbal feedback*, two independent analysts (HH and EP) examined the responses independently and assessed the extent to which responses aligned with communication goals for each cycle by deductively comparing participant responses to our detailed concept maps. Any disagreement was resolved through discussion with the senior author (HW). We noted anything that failed to align with communication goals or was confusing to participants to guide changes for the next cycle.

After collecting data for each cycle, the first author (HH) compiled and reviewed data with coauthors (EP and MTB), summarized problems, and drafted recommendations. These recommendations were then discussed with the senior author (HW) and, when necessary, the larger team (remaining authors) to determine changes for the next cycle.

## Iterative Cycles

### First Cycle

Our multidisciplinary team developed the first version of a visualization based on epidemiological evidence that we had organized in the concept map. We prespecified communication goals for different visual design elements (ie, what we wanted to convey with each element of the visualization and how we expected people to respond). We used four devices (Figure 2) and brief verbal feedback (audio-recorded) to assess participants' interpretations and reactions to the content of the visualization. After viewing, the participants described the visualization in their own words. They were also asked the following questions: What do they think this visualization wants to convey? Is there anything in the visualization that they find unclear or confusing?

### Second Cycle

We developed a revised version of the visualization based on participants' feedback in the first cycle. We predefined our communication goals for the second cycle (Multimedia Appendix 9) and refined the concept map by adding how different diseases spread differently (pertussis, measles, and influenza as test case) and that different diseases require a different number of vaccine doses (eg, a single dose, multiple, booster, or annual doses). The visualization showed how different parameters (eg, vaccine coverage and intracommunity contact) can influence community immunity. We audio-recorded a brief verbal feedback.

### Verbal Feedback

In this cycle, we only used audio-recorded verbal feedback (no psychophysiological measurements were used in this cycle [Table 1]) to assess participants' interpretations of the visual content and their suggestions to improve it. We chose this method to increase the richness of verbal responses for each visual element. We asked participants to describe their understanding of the visualization, how vaccines work to protect people from diseases, what it means to be immune, and if there was anything confusing or unclear in the visualization. We showed images from the visualization to participants and asked specific questions (eg, what do the icons of the older woman and the baby represent in this visualization? What do the images of viruses causing different diseases represent?). We also asked participants about different terms used to explain community immunity, that is, herd immunity, community immunity, and community protection and which term they prefer.

### Third Cycle

We developed a third version of a visualization based on participants' feedback in the second cycle. We used the same techniques as in the human-computer interaction laboratory described earlier, along with verbal feedback. The third cycle was tested in two different settings: a university and different

locations in a community setting (two shopping malls and a public library). We predefined the communication goals for the third cycle (see [Multimedia Appendix 10](#) for university sample and [Multimedia Appendix 11](#) for community sample). We asked participants to describe, in their own words, the visualization shown to them. We included a larger number of participants in this cycle, as our visualization was closer to launch and we wanted to make sure that it was easily understood and that people grasped the concept of community immunity. We also wanted to test if our communication goals were achieved among people with varied levels of education.

#### **Fourth Cycle**

By the fourth cycle, the content of our visualization had achieved nearly all predefined communication goals. However, one major issue remained. Up to this cycle, we had used generic avatars in our visualization. On the basis of data from previous cycles, we were concerned about the extent to which people could identify with the generic avatars presented in the visualization. Therefore, we developed an additional piece in which people were asked to build their own communities by making personalized avatars (their own, 2 vulnerable people in their community, and 6 avatars of people around them who could be family members or coworkers). We added this feature so that people could better identify with the avatars that were subsequently integrated into our application to explain community immunity. We asked participants to provide critical feedback on the process of creating their own avatars and

building their own communities. In this cycle, we focused on three questions related to the new features: (1) Was an onboarding tutorial describing how to build avatars a useful addition? (2) Was it easy to build the avatars? and (3) Was the length of the avatar-building process reasonable? We further asked what participants thought of the avatar-building options, including the accessories and color palettes for skin tone and hair color. Participants also described the application in their own words. We only used the eye-tracking device in this cycle to assess visual attention.

## **Results**

### **Study Participants**

A total of 110 eligible participants across the four cycles (cycle 1 [n=8], cycle 2 [n=11], cycle 3 [n=83], and cycle 4 [n=8]) participated in the study ([Table 2](#)). Overall, 61.8% (68/110) of the participants were women and 34.5% (38/110) were men; 3.6% (4/110) did not report their gender. The mean age was 38 years (SD 17). Furthermore, 96.3% (106/110) of the participants spoke and understood French, 29.0% (32/110) spoke and understood English, whereas 3.6% (4/110) did not report the language spoken. More than half of the participants (65/110, 59.0%) had a university-level education. Most participants (85/110, 77.2%) reported no physical disability, 16.3% (18/110) reported some form of disability, and 2.7% (3/110) preferred not to answer. Across the 110 participants, 3 (2.7%) did not complete the sociodemographic questionnaire.



**Table 2.** Sociodemographics of each cycle.

Demographic characteristic	First cycle (n=8)	Second cycle (n=11)	Third cycle (university sample; n=49)	Third cycle (community sample; n=34)	Fourth cycle (n=8)	Across all cycles (N=110)
<b>Self-identified gender , n (%)</b>						
Female	3 (38)	7 (64)	34 (69)	16 (47)	8 (100)	68 (62)
Male	2 (25)	4 (36)	15 (31)	17 (50)	0 (0)	38 (35)
Not reported	3 (38)	0 (0)	0 (0)	1 (3)	0 (0)	4 (4)
Age (years), mean (SD)	28 (8)	24 (7)	37 (13)	52 (15)	26 (8)	38 (17)
<b>Language, n (%)</b>						
French	5 (63)	11 (100)	48 (98)	34 (100)	8 (100)	106 (96)
English	5 (63)	10 (91)	14 (29)	1 (3)	2 (25)	32 (29)
Not reported	3 (38)	0 (0)	1 (3)	0 (0)	0 (0)	4 (4)
<b>Physical disability, n (%)</b>						
Yes	0 (0)	0 (0)	9 (18)	9 (26)	0 (0)	18 (16)
No	5 (63)	11 (100)	40 (82)	21 (62)	8 (100)	85 (77)
Not reported	3 (38)	0 (0)	0 (0)	1 (3)	0 (0)	4 (4)
Prefer not to answer	0 (0)	0 (0)	0 (0)	3 (9)	0 (0)	3 (3)
<b>Education level, n (%)</b>						
Some elementary School	0 (0)	0 (0)	0 (0)	4 (12)	0 (0)	4 (4)
High school diploma	0 (0)	0 (0)	2 (4)	9 (26)	1 (13)	12 (11)
College or polytechnic school certificate or diploma (CÉGEP <sup>a</sup> , AEC, DEC)	1 (13)	4 (36)	8 (16)	6 (18)	3 (38)	22 (20)
University graduate degree (bachelor's)	1 (13)	2 (18)	14 (29)	9 (26)	1 (13)	27 (25)
University graduate degree (master's)	3 (38)	5 (45)	20 (41)	1 (3)	3 (38)	32 (29)
University graduate degree (doctorate)	0 (0)	0 (0)	5 (10)	1 (3)	0 (0)	6 (5)
Do not know	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Prefer not to answer	0 (0)	0 (0)	0 (0)	3 (9)	0 (0)	3 (3)
Not reported	3 (38)	0 (0)	0 (0)	1 (3)	0 (0)	4 (4)

<sup>a</sup>Quebec educational level requiring 2 years of study after completion of grade 11. CÉGEP students are typically 17 to 19 years old, and students typically must complete CÉGEP to be admitted to university.

## First Cycle

### Findings From This Cycle

We obtained psychophysiological data from 6 of 8 participants and qualitative verbal feedback from 8 of 8 participants. There were missing psychophysiological data for 2 participants because of technical issues with the devices. Specifically, we had problems initializing the EEG.

As described in Table 3, the design elements of the visualization achieved their communication goals to varying degrees. All participants (8/8) reported that people in clusters of hexagons represented members of the community. Most participants (7/8) reported that a yellow background indicated vulnerable people,

and 6 of 6 participants responded psychophysiological in desired ways, that is, peaks in arousal, and high engagement when a vulnerable person became infected. All participants (8/8) reported that red connecting lines represented the spread of infection. Half of the participants (3/6) did not visually attend to the appearance of a thick blue band indicating community immunity upon its appearance. When questioned about its meaning, most participants (6/8) reported that the blue band around vulnerable people meant community immunity, whereas 2 of 8 participants interpreted it as some sort of linkage between the older woman and the baby. All participants (8/8) explicitly mentioned in their explanation that when enough people were vaccinated, this created a protective barrier of community immunity to prevent the spread of infection. Overall, all






participants had a neutral (6/6) facial expression when community immunity was explained in the visualization.



### ***Changes for the Next Cycle***






A number of aspects of the first version of the visualization needed improvement. First, only a few participants (1/6) visually attended to the appearance of the central avatar, and only half of the participants (4/8) reported that the central avatar represents them. Second, most people (5/8) did not understand that the avatars around them could be others they see often but who are not members of their immediate family, for example, coworkers. To address these two issues, for the next cycle, we presented the center avatar, immediate family members, colleagues, and other regular contacts in the same visual frame by zooming in and out. Third, participants showed either low engagement (2/6) or drowsy or unengaged (4/6) when an infection first entered the community. To address this, rather than having the infection simply appear, we used a red line to

visually represent the entry and spread of infection in the community. Fourth, participants (8/8) suggested that the visualization came across as a simple promotion of immunization rather than explaining how community immunity works. Although these concepts are interrelated in the sense that community immunity requires sufficient numbers of people to be immunized, our goal was to explain community immunity. To address this issue, we increased the focus on community immunity in the narration of our visualization. In discussing this latter change among our team, we identified a need to test the terms herd immunity, community protection, and community immunity by asking participants in the next cycle about their reactions to each of the three terms. In our team discussions, we also identified the need for new visual elements about different viruses (using measles, pertussis, and influenza as examples) to explain in greater detail why different diseases require vaccine doses and schedules.

**Table 3.** The communication goals set for the first cycle of visualization.

S. no.	Design element or a concept	Message design elements intended to convey in the visualization (desired interpretation and/or reaction)	What users reported when viewing these design elements (verbal feedback; n=8)	How users reacted to these design elements (psychophysiology; n=6)
1.		The avatar represents the participant.	Of 8 participants, 4 reported that the avatar represents them. The other 4 participants interpreted it as representing a person, but not them.	Of 6 participants, 1 visually attended to the appearance of the avatar. Overall valence was positive across the 6 participants.
2.		The hexagonal shape represents a unit.	Of 8 participants, 2 reported that each hexagonal shape was a separate unit. The other 6 participants interpreted it as an unspecified symbol or a honeycomb.	N/A <sup>a</sup> (no psychophysiology data specific to this visual element).
3.		A person in a hexagonal shape around the central avatar represents the participant's regular contacts (family members, friends, neighbors, or colleagues).	Of 8 participants, 3 reported that a person in hexagonal shape was a member of their community; 5 participants interpreted it as their family member.	N/A (no psychophysiology data specific to this visual element).
4.		<ul style="list-style-type: none"> <li>Icon of an older woman and a baby represents vulnerable people or those with fragile immune systems (eg, patients with cancer).</li> <li>High arousal and visual attention were expected when vulnerable people appeared in the visualization.</li> </ul>	All participants (8/8) reported that an older woman and a baby in the visualization represent vulnerable people.	Of 6 participants, 4 visually attended when vulnerable people appeared in the visualization. Of 6 participants, 3 showed a peak in arousal when vulnerable people appeared.
5.	Yellow color behind <i>baby</i> and <i>an older woman</i>	Yellow color signals vulnerable people.	Of 8 participants, 7 reported that yellow color signals vulnerable people; 1 participant did not pay attention to the yellow color in the visualization.	N/A (no psychophysiology data specific to this visual element).
6.		Red color signals diseased or infected; blue color signals vaccinated or protected; gray color signals susceptible to disease or infection	<ul style="list-style-type: none"> <li>All participants (8/8) reported that the red color in the visualization represents disease, infection, or danger.</li> <li>All participants (8/8) reported that the blue color in the visualization signals being safe from diseases or vaccinated.</li> <li>Of 8 participants, 6 reported that gray color signals being susceptible to disease/infection or not vaccinated; 2 interpreted gray color as people who can be vulnerable.</li> </ul>	N/A (no psychophysiology data specific to this visual element).

S. no.	Design element or a concept	Message design elements intended to convey in the visualization (desired interpretation and/or reaction)	What users reported when viewing these design elements (verbal feedback; n=8)	How users reacted to these design elements (psychophysiology; n=6)
7.	When infection first enters the community.	High arousal, engagement, and visual attention were expected when the visualization shows when the infection first enters the community.	No comments recorded.	<ul style="list-style-type: none"> <li>Of 6 participants, 3 visually attended when infection first entered the community.</li> <li>Of 6 participants, 2 showed a peak in arousal when infection first entered the community.</li> <li>No participants (0/6) were most likely to be in a high-engagement state when the infection first entered the community; 2 of 6 participants were most likely to be in a low-engagement state; 4 of 6 participants were most likely to be in a drowsy (unengaged) state.</li> </ul>
8.	When the central avatar gets infected.	High arousal, engagement, and visual attention were expected when the visualization shows the central avatar representing the participant getting infected.	No comments recorded.	<ul style="list-style-type: none"> <li>Of 6 participants, 1 visually attended when the avatar got infected.</li> <li>Of 6 participants, 4 showed peaks in arousal when the avatar got infected.</li> <li>Of 6 participants, 4 were most likely to be in a high-engagement state when the avatar got infected.</li> </ul>
9.		<ul style="list-style-type: none"> <li>Red connecting lines represent the spread of infection.</li> <li>High arousal, engagement, and visual attention was expected when the visualization showed red connecting lines indicating the spread of infection.</li> </ul>	All participants (8/8) reported that red connecting lines indicate the spread of infection.	<ul style="list-style-type: none"> <li>Of 6 participants, 1 visually attended to red connecting lines.</li> <li>Of 6 participants, 1 showed peak in arousal when red connecting lines appeared.</li> <li>All participants (6/6) were most likely to be in a high-engagement state when red connecting lines appeared.</li> </ul>
10.	When the vulnerable people get infected.	High arousal, engagement, and visual attention were expected when the vulnerable people got infected.	No comments recorded.	<ul style="list-style-type: none"> <li>Of 6 participants, 3 visually attended when vulnerable people got infected.</li> <li>All participants (6/6) showed a peak in arousal and a negative valence when vulnerable people got infected.</li> <li>All participants (6/6) were most likely to be in the state of high engagement when vulnerable people got infected.</li> </ul>
11.	When community immunity was explained	<ul style="list-style-type: none"> <li>Participants' explanations include the concept of community immunity.</li> <li>High arousal, visual attention, and positive valence was expected when the visualization demonstrated the concept of community immunity.</li> </ul>	All participants' (8/8) explanations include the concept of community immunity, that is what it is and how it works.	<ul style="list-style-type: none"> <li>Of 6 participants, 4 visually attended when community immunity was explained.</li> <li>Of 6 participants, 4 showed peak in arousal when community immunity was explained.</li> <li>Overall facial expression was neutral across the 6 participants.</li> </ul>
12.				

S. no.	Design element or a concept	Message design elements intended to convey in the visualization (desired interpretation and/or reaction)	What users reported when viewing these design elements (verbal feedback; n=8)	How users reacted to these design elements (psychophysiology; n=6)
		<ul style="list-style-type: none"> <li>Thick blue band around vulnerable people indicates community immunity.</li> <li>High engagement and visual attention was expected when the thick blue band appeared around vulnerable people.</li> </ul>	Of 88 participants, 6 reported that the thick blue band around vulnerable people represents community immunity, which protects them from getting infected.	<ul style="list-style-type: none"> <li>Of 6 participants, 3 visually attended when the blue line appeared around vulnerable people.</li> <li>All participants (6/6) were most likely to be in a high-engagement state when the blue line appeared around vulnerable people.</li> </ul>
13.		<ul style="list-style-type: none"> <li>Blue lines spreading out from vaccinated people indicate the community immunity.</li> <li>High engagement was expected when blue lines appeared indicating the community immunity.</li> </ul>	All participants (8/8) reported that blue lines spreading out from vaccinated people show the protective barrier that is community immunity.	All participants (6/6) were most likely to be in a high-engagement state when blue lines appeared indicating the community immunity.
14.		The cluster of hexagons represent different communities.	All participants (8/8) reported that clusters of hexagons represent different communities.	N/A (no psychophysiology data specific to this visual element).
15.		The avatar in the cluster of hexagons represents members of the community.	All participants (8/8) reported that the avatar in the clusters of hexagons represents members of the community.	N/A (no psychophysiology data specific to this visual element).
16.		The gray outline around the cluster of hexagons indicates a group or members of the same community.	Of 8 participants, 6 reported that the gray outline indicates the group or members of the same community.	N/A (no psychophysiology data specific to this visual element).
17.		<ul style="list-style-type: none"> <li>The orange outline showed the participant's community.</li> <li>High engagement was expected when an orange outline appeared around their community.</li> </ul>	Of 8 participants, 7 interpreted the orange outline as their community.	<ul style="list-style-type: none"> <li>Of 6 participants, 3 visually attended when the orange outline appeared around their community.</li> <li>All participants (6/6) were most likely to be in a high-engagement state when the orange outline appeared around their community.</li> </ul>

<sup>a</sup>N/A: not applicable.

## Second Cycle

### Findings From This Cycle

The second version of the visualization achieved most of its communication goals. [Multimedia Appendix 9](#) provides details analogous to those provided in [Table 3](#) for the first cycle. All participants (11/11) reported that the people in the hexagon represent members of their community or people with whom they were in daily contact, the *older woman* and the *baby* in the visualization represented vulnerable members of the community, and the hexagons represented individuals. Most participants (7/11) reported that the visualization communicated that vaccines are not perfect, and nearly all reported that some vaccines require multiple doses or booster shots to work (10/11). All participants' (11/11) responses showed that they understood the use of colors to signal vulnerability and infection such as a yellow background indicating vulnerable people, and that red

color showed propagation of the disease. All participants (11/11) reported that community immunity safeguards vulnerable people, that is, when sufficient number of people around them were vaccinated, whereas lower vaccine coverage puts communities and the people within them, especially vulnerable populations, at risk of becoming sick. Participants indicated that the term *community immunity* best conveyed the concept compared with terms *herd immunity* (which implies herds of animals) or *community protection* (which participants indicated evoked images of protection via firearms.) Few (2/11) participants reported that the color blue indicated immunity, and none (0/11) showed understanding that the color gray indicated susceptibility to infection. Some participants reported that diseases differ (3/11) and spread at different rates (3/11). Few participants (2/11) reported vaccine-induced immunity, whereas none reported the concept of natural immunity (0/11). Some participants (3/11) reported the role of vaccine



effectiveness in creating community immunity, whereas others did not.

### *Changes for the Next Cycle*

Aspects of the visualization that needed to be improved included conveying that the color blue means being vaccinated or immune, the color gray means being susceptible, and focusing attention on the fact that different diseases spread at different rates. In addition, the visualization did not yet help participants understand the role of vaccine effectiveness in community immunity or distinguish between natural immunity and vaccine-induced immunity. Participants further suggested that the visualization was too long and provided too much information to retain. In the third cycle, we kept the colors blue and gray but explained their meaning in the narration. We removed the images representing different viruses but kept the narration explaining how different infections spread at different rates, illustrating it with infection spread. We further added depictions of different vaccine coverage for different diseases to show how community immunity prevents the spread of infection. We removed images illustrating natural and vaccine-induced immunity and different vaccine doses, and instead wove this information into the narration illustrated by a single image of immunity. We shortened the visualization for the next cycle to about 2 min and used the term community immunity in the narration.

### **Third Cycle**

#### *Findings From This Cycle*

The third cycle mostly achieved its communication goals (see [Multimedia Appendix 10](#) for university sample and [Multimedia Appendix 11](#) for community sample). A total of 83 participants (university sample:  $n=48$ ; community sample:  $n=34$ ) participated in our third cycle. Most participants (51/83, 61%) reported that the *older woman* and *the baby* represented vulnerable people or those with fragile immune systems (eg, patients with cancer). Most participants' (60/83, 72%) verbal feedback summarizing the visualization included the point that vaccines prevent the spread of infection. Most participants (42/83, 50%) reported that community immunity safeguards everyone, some participants (34/83, 41%) reported that the thick blue band around an *older woman* and *the baby* demonstrated community immunity protecting vulnerable population and that an individual's decision to get vaccinated or not has an impact on other people in their community (36/83, 43%). Most participants visually attended to all communication goals in a desired way; for example, nearly all participants (80/83, 96%) visually attended when the contagious disease spread to vulnerable

people, when vaccines wane over time (74/83, 89%), and when community immunity safeguards everyone (81/83, 97%). Overall, across all 83 participants, people were likely in a state of high engagement and optimal workload during the explanation of how community immunity safeguards everyone.

### *Changes for the Next Cycle*

Results from this cycle suggested that participants mostly understood the information presented in the visualization. However, some wording was unclear, so we made changes to the script to clarify it. For example, in the portion of the animation explaining how vaccines' effectiveness wanes over time, we changed the script from, *They don't work every time, and can wane over time*, to, *their protection can fade over time*. We also changed the order of some design elements to better align with how people appeared to understand the information during testing. For example, rather than first presenting how different diseases spread at different rates and then explaining community immunity, we changed these to present community immunity first, facilitating the explanation of why some diseases need more people to be vaccinated to create community immunity. Most importantly, until this cycle, we used generic avatars in the visualization. However, the generic avatars continued to be difficult for participants to interpret. To personalize the avatars so that people could better identify with them, we added a functionality so that people could build their own communities by making an avatar for themselves, 2 avatars for vulnerable people in their community, and 6 avatars of other people they see regularly, such as family members or coworkers. These personalized avatars were then integrated into our visualization to help participants better understand and respond emotionally to the idea of family members, friends, or other close contacts being vulnerable and at risk of infection.

### **Fourth Cycle**

We tested the last version of the application with 8 participants. All participants (8/8) reported what community immunity was and how it worked. Participants found the tutorial on how to create avatars confusing and preferred to make avatars by reading instructions. All participants (8/8) were able to easily create avatars by following step-by-step instructions, without a tutorial. All participants liked the color palettes for skin tone and hair colors. A participant suggested adding different accessories with options such as caps, hats, and a hijab to include more culturally diverse and realistic avatars.

The key findings of all cycles and major changes implemented are summarized in [Table 4](#).

**Table 4.** Key findings of all cycles.

Cycles	Key communication goals achieved	Key communication goals not achieved	Summary of psychophysiological data (where applicable)	Summary of how issues were addressed in the next cycle
First cycle	<ul style="list-style-type: none"> <li>Nearly all participants reported that the color yellow represents vulnerable people.</li> <li>Most participants reported that the blue band around vulnerable people meant protection.</li> <li>All participants reported that the avatars in the cluster of hexagons represent members of the community.</li> </ul>	<ul style="list-style-type: none"> <li>Most participants did not understand that the central avatar represents them.</li> <li>Some participants did not understand that the avatars around them could include nonfamily contacts, for example, coworkers.</li> <li>All participants understood the purpose of visualization as promoting immunization rather than explaining the concept of community immunity.</li> </ul>	<ul style="list-style-type: none"> <li>Most participants visually attended to the appearance of vulnerable people.</li> <li>Most participants had peaks in arousal and showed high engagement when avatars got infected.</li> <li>All participants showed high engagement when red lines showed the spread of infection.</li> <li>Most participants visually attended, and all participants had peaks in arousal and showed high engagement when vulnerable people got infected.</li> <li>Most participants visually attended and had peaks in arousal when community immunity was explained.</li> <li>All participants showed high engagement when blue lines around people spreading out from vaccinated people showed community immunity.</li> </ul>	<ul style="list-style-type: none"> <li>We presented the center avatar, immediate family members, colleagues, and communities in the same visual frame by zooming in and out.</li> <li>We removed the term immunization in the narration script and focused more on community immunity.</li> <li>We decided to test the terms herd immunity, community protection, and community immunity by asking participants which term they relate to and prefer.</li> <li>We added a question to be asked in the next cycle about the shape of hexagons presented in the visualization.</li> <li>We added a new design element in the next cycle explaining how different viruses (eg, measles, pertussis, influenza) spread at different rates and require different vaccine schedules.</li> </ul>
Second cycle	<ul style="list-style-type: none"> <li>All participants reported that yellow signaled vulnerability.</li> <li>All participants reported that red signaled infection.</li> <li>All participants reported that hexagons represent people.</li> <li>Nearly all participants reported community immunity safeguards vulnerable people when sufficient people around them are vaccinated.</li> <li>All participants preferred the term community immunity over herd immunity or community protection.</li> </ul>	<ul style="list-style-type: none"> <li>Few participants reported that blue meant vaccinated or immune.</li> <li>No participants reported that gray meant susceptible.</li> <li>Most participants did not report that different diseases spread at different rates.</li> <li>Most participants did not report the role of vaccine effectiveness in community immunity.</li> <li>Few participants reported vaccine-induced immunity.</li> <li>None of the participants reported accurate understanding of natural immunity.</li> <li>All participants suggested that the visualization should be shorter.</li> </ul>	<ul style="list-style-type: none"> <li>N/A<sup>a</sup></li> </ul>	<ul style="list-style-type: none"> <li>We removed images of viruses but retained in the narration explanation of how different infections spread at different rates.</li> <li>We added images about different levels of vaccine coverage to achieve community immunity for different diseases.</li> <li>We removed images about different vaccine doses and natural and vaccine-induced immunity.</li> <li>We shortened the visualization to about 2 min and used the term community immunity.</li> </ul>

Cycles	Key communication goals achieved	Key communication goals not achieved	Summary of psychophysiological data (where applicable)	Summary of how issues were addressed in the next cycle
Third cycle	<ul style="list-style-type: none"> <li>• Most participants reported that the older woman and baby avatars represent vulnerable people or those with fragile immune systems.</li> <li>• Most participants reported that vaccines prevent the spread of infection.</li> <li>• Most participants reported that community immunity safeguards everyone.</li> <li>• Some participants reported that the thick blue band around an older woman and the baby shows community immunity protecting vulnerable populations.</li> <li>• Some participants reported that their decision to get vaccinated or not has an impact on other people in their community.</li> </ul>	<ul style="list-style-type: none"> <li>• Nearly all participants found it difficult to identify with the generic avatars.</li> </ul>	<ul style="list-style-type: none"> <li>• Some participants had peaks in arousal when the avatar first appeared.</li> <li>• Most participants had peaks in arousal when the vulnerable population was explained in the visualization.</li> <li>• Some participants had peaks in arousal when the infection first entered the community in the visualization.</li> <li>• Most participants visually attended when the community immunity was shown, along with how it safeguards everyone.</li> <li>• Overall, participants showed high engagement and an optimal workload throughout the visualization.</li> </ul>	<ul style="list-style-type: none"> <li>• We added a functionality for people to build their own avatars and their own communities.</li> </ul>
Fourth cycle	<ul style="list-style-type: none"> <li>• All participants reported an accurate understanding of what community immunity is and how it works.</li> <li>• All participants reported that community immunity safeguards vulnerable people and everyone in the community.</li> <li>• All participants reported that some infections spread faster and need enough people to get vaccinated to prevent the spread of infections.</li> <li>• All participants found it easy to create avatars by following instructions without a tutorial.</li> <li>• All participants liked the palettes for skin and hair colors.</li> </ul>	<ul style="list-style-type: none"> <li>• For all participants, the avatar creation tutorial was confusing. They preferred to make avatars just by reading the step-by-step instructions.</li> <li>• A participant suggested adding additional accessories such as caps, hats, hijab, and other head and hair coverings.</li> </ul>	<ul style="list-style-type: none"> <li>• Nearly all participants visually attended to the avatar creation elements, including written instructions.</li> </ul>	<ul style="list-style-type: none"> <li>• Head and hair covering options (caps, hats, hijab, turban) were added.</li> </ul>

<sup>a</sup>N/A: not applicable.

In [Table 4](#), we summarize quantitative findings by referring to all participants when 100% of participants exhibited this; nearly all: 80% to 99%; most: 50% to 79%; some: 25% to 49%; few: 1% to 24%; no participants: 0%.

## Discussion

### Principal Findings and Comparison With Previous Literature

Considering our study as a whole, we observed three principal findings. First, visualization does indeed appear to be a promising medium for conveying what community immunity is and how it works. Our project builds on the limited previous literature on visualization to convey the concept of community immunity. On the basis of our systematic review [13], Betsch et al [10] are the only team to have developed and evaluated such an interactive visualization. Their visualization increased vaccination intentions and demonstrated the promise of this

medium for conveying the concept of community immunity. We built on this by adding personalization to increase the user's identification with the avatars, a voice-over to increase learning, especially among people with lower literacy, and a focus on the protection of vulnerable members of a community. In addition to previous research by Betsch et al [10,57], other studies have also pointed to the potential advantages of using visualization and videos to convey the concept of community immunity [58,59].

Second, our study shows that by involving users in iterative cycles during the design process, it is possible to create a relatively short and simple visualization that conveys a mathematically complex topic. This aligns with previous literature suggesting that visualizations can support people in understanding complex concepts [60-62], and users' involvement in the design process can facilitate an understanding of the information [63].

Finally, our study shows that considering emotion during the design process can help inform the final design of the intervention. Emotions play an important role in health decision making [17], especially when deciding on behalf of loved ones [64], as is the case when deciding about vaccinating one's children [65]. To the best of our knowledge, no previous study has considered emotions in developing a tool to explain community immunity [13]. In keeping with the Affect Heuristic within our framework, our study explicitly considered emotion, as expressed in verbal feedback and measured with psychophysiological data. According to Peters et al [66], affect has four possible functions in health communication and decision making. Affect can directly influence decisions according to a person's subjective sense of the *goodness* or *badness* of options; it can function "as a spotlight" to direct a person's attention toward information, which, in turn, shapes their judgments and decisions; it can motivate information processing and behavior; and it can help people trade-off between concepts that are difficult to compare directly. Because our application is designed primarily to convey a complex concept to inform decisions, we focused on affect's function as a spotlight and adapted our application to better provoke emotional reactions to key information, such as the vulnerability of some community members. Attending to data about participants' emotions throughout the design process therefore helped us carefully adapt our application to the way people perceive and use information to make health-related or other decisions.

### Limitations

This study has four main limitations. First, study participants were primarily French-speaking people in Quebec City, Canada, predominantly women, and many had a relatively high level of education. Our recruitment materials for different cycles mentioned that the study was about vaccination or health, which may have contributed to the over-representation of women in our university-based samples. Women seek health services more frequently than men [67] and are the overwhelming majority

of participants in studies on childhood vaccine decision making [68]. To address this, in our largest cycle (third cycle), we expanded our recruitment strategy to include community-based settings. By deliberately recruiting a large subsample from a population that was more likely to include men and more likely to include people with lower levels of education, we were better able to ensure that the final design would be appropriate for a broad audience. However, despite our best efforts to diversify our study sample, our results may not be generalizable to other populations. Second, our application currently requires that users be able to visually perceive presentations on a screen. Further work will ensure accessibility for people who are blind or visually impaired. Third, building avatars and launching an application requires a certain level of computer literacy, meaning that the application will not necessarily serve people who are uncomfortable using or unable to use computers. Finally, the studies included people who were specifically recruited to participate in a study. It remains to be seen whether people are willing to view a 2-min visualization of community immunity outside of a study setting.

### Conclusions

Our application shows promise as a method of conveying the concept of community immunity to a broad range of members of the general public. This study has practical implications regarding how to design health communication materials about complex topics, such as community immunity, and other concepts that combine individual and population benefits and harms, such as antibiotic resistance, health resource allocation, and interventions during epidemics. Applications with personalized avatars may be more effective than abstract visual representations or text-based explanations to help people understand their personal role in population health. Further research could evaluate the specific effects of personalization. Our future work will test our application in a web-based randomized controlled trial to assess its effects on risk perception, knowledge, and vaccination intentions.

### Acknowledgments

The authors gratefully acknowledge the contributions of Mehdi Aouami and Rosalie Baker for designing the first version of their prototype and Professor Frédéric Lepinay (Design School of Laval University), Dr Ruth Ndjaboue (epidemiologist and postdoctoral fellow at Laval University), Dr Bryna Warshawsky (Public Health Physician at Public Health Ontario), and other colleagues for their constructive comments on an earlier version of prototypes to improve our design. We thank Donovan Aziaka, Magniol Noubi, and Nadège Touko Toudjeu for programming the final version of the application and making it ready to launch. This study was funded by the Canadian Institutes of Health Research (grant number FDN-148426, 2016-2021, principal investigator: Witteman). The CIHR had no role in determining the study design, the plans for data collection or analysis, the decision to publish, or the preparation of this manuscript. During this study, HW received salary support from Fonds de Recherche du Québec—Santé (Research Scholar Junior 1, then Junior 2 Career Development Awards) and from a Canada Research Chair (Tier 2) in Human-Centred Digital Health.

### Authors' Contributions

HH and HW contributed to the design of the study. HH, EP, and MB contributed to data collection. HH, EP, MB, DA, MN, NT, and HW helped in designing and programming the application. HH and HW drafted the first version of the article. DA, JB, CC, ED, MD, TG, AG, EK, SM, MN, RO, EP, JP, JR, BS, AS, NT, MB, KW, JL, and DR critically revised the article and approved the final version for submission for publication. HH and HW had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## Conflicts of Interest

KW is the CEO and founder of CANImmunize, a digital immunization platform.

### Multimedia Appendix 1

Concept map.

[[XLSX File \(Microsoft Excel File\), 70 KB](#) - [jmir\\_v22i10e20113\\_app1.xlsx](#) ]

### Multimedia Appendix 2

Script for 1st cycle.

[[DOCX File , 765 KB](#) - [jmir\\_v22i10e20113\\_app2.docx](#) ]

### Multimedia Appendix 3

Script for 2nd cycle.

[[DOCX File , 25908 KB](#) - [jmir\\_v22i10e20113\\_app3.docx](#) ]

### Multimedia Appendix 4

Script for 3rd cycle.

[[DOCX File , 3954 KB](#) - [jmir\\_v22i10e20113\\_app4.docx](#) ]

### Multimedia Appendix 5

Script for 4th or last cycle.

[[DOCX File , 3958 KB](#) - [jmir\\_v22i10e20113\\_app5.docx](#) ]

### Multimedia Appendix 6

First cycle verbal feedback.

[[XLSX File \(Microsoft Excel File\), 23 KB](#) - [jmir\\_v22i10e20113\\_app6.xlsx](#) ]

### Multimedia Appendix 7

Second cycle verbal feedback.

[[XLSX File \(Microsoft Excel File\), 773 KB](#) - [jmir\\_v22i10e20113\\_app7.xlsx](#) ]

### Multimedia Appendix 8

Third cycle verbal feedback.

[[XLSX File \(Microsoft Excel File\), 72 KB](#) - [jmir\\_v22i10e20113\\_app8.xlsx](#) ]

### Multimedia Appendix 9

The communication goals set for the second iterative cycle of visualization.

[[DOCX File , 1901 KB](#) - [jmir\\_v22i10e20113\\_app9.docx](#) ]

### Multimedia Appendix 10

The communication goals set for the third iterative cycle of visualization (university sample).

[[DOCX File , 3240 KB](#) - [jmir\\_v22i10e20113\\_app10.docx](#) ]

### Multimedia Appendix 11

The potential communication goals set for the third iterative cycle of visualization (community sample).

[[DOCX File , 3238 KB](#) - [jmir\\_v22i10e20113\\_app11.docx](#) ]

## References

1. Rashid H, Khandaker G, Booy R. Vaccination and herd immunity: what more do we know? *Curr Opin Infect Dis* 2012 Jun;25(3):243-249. [doi: [10.1097/QCO.0b013e328352f727](#)] [Medline: [22561998](#)]
2. Uchida M, Kaneko M, Hidaka Y, Yamamoto H, Honda T, Takeuchi S, et al. High vaccination coverage is associated with low epidemic level of seasonal influenza in elementary schools: an observational study in Matsumoto City, Japan. *BMC Infect Dis* 2018 Mar 13;18(1):128 [FREE Full text] [doi: [10.1186/s12879-018-3025-9](#)] [Medline: [29534676](#)]
3. Fine P, Eames K, Heymann D. 'Herd immunity': a rough guide. *Clin Infect Dis* 2011 Apr 1;52(7):911-916. [doi: [10.1093/cid/cir007](#)] [Medline: [21427399](#)]



4. Leslie TF, Delamater PL, Yang YT. It could have been much worse: the Minnesota measles outbreak of 2017. *Vaccine* 2018 Mar 27;36(14):1808-1810 [FREE Full text] [doi: [10.1016/j.vaccine.2018.02.086](https://doi.org/10.1016/j.vaccine.2018.02.086)] [Medline: [29496348](https://pubmed.ncbi.nlm.nih.gov/29496348/)]
5. Darnal J, Tsheten, Peldon S, Dorji T, Dorji T, Dorji G, et al. Measles outbreak among nomadic population with low herd immunity in an eastern district of Bhutan, 2016. *OSIR J* 2018;11(1):14-21.
6. Majumder MS, Cohn EL, Mekaru SR, Huston JE, Brownstein JS. Substandard vaccination compliance and the 2015 measles outbreak. *JAMA Pediatr* 2015 May;169(5):494-495 [FREE Full text] [doi: [10.1001/jamapediatrics.2015.0384](https://doi.org/10.1001/jamapediatrics.2015.0384)] [Medline: [25774618](https://pubmed.ncbi.nlm.nih.gov/25774618/)]
7. Quadri-Sheriff M, Hendrix KS, Downs SM, Sturm LA, Zimet GD, Finnell SM. The role of herd immunity in parents' decision to vaccinate children: a systematic review. *Pediatrics* 2012 Sep;130(3):522-530. [doi: [10.1542/peds.2012-0140](https://doi.org/10.1542/peds.2012-0140)] [Medline: [22926181](https://pubmed.ncbi.nlm.nih.gov/22926181/)]
8. Logan J, Nederhoff D, Koch B, Griffith B, Wolfson J, Awan FA, et al. 'What have you HEARD about the HERD?' Does education about local influenza vaccination coverage and herd immunity affect willingness to vaccinate? *Vaccine* 2018 Jun 27;36(28):4118-4125 [FREE Full text] [doi: [10.1016/j.vaccine.2018.05.037](https://doi.org/10.1016/j.vaccine.2018.05.037)] [Medline: [29789242](https://pubmed.ncbi.nlm.nih.gov/29789242/)]
9. Arnesen S, Bærøe K, Cappelen C, Carlsen B. Could information about herd immunity help us achieve herd immunity? Evidence from a population representative survey experiment. *Scand J Public Health* 2018 Dec;46(8):854-858. [doi: [10.1177/1403494818770298](https://doi.org/10.1177/1403494818770298)] [Medline: [29741455](https://pubmed.ncbi.nlm.nih.gov/29741455/)]
10. Betsch C, Böhm R, Korn L, Holtmann C. On the benefits of explaining herd immunity in vaccine advocacy. *Nature Human Behaviour* 2017 Mar 6;1(3):0056. [doi: [10.1038/s41562-017-0056](https://doi.org/10.1038/s41562-017-0056)]
11. Lefebvre CD, Terlinden A, Standaert B. Dissecting the indirect effects caused by vaccines into the basic elements. *Hum Vaccin Immunother* 2015;11(9):2142-2157 [FREE Full text] [doi: [10.1080/21645515.2015.1052196](https://doi.org/10.1080/21645515.2015.1052196)] [Medline: [26186100](https://pubmed.ncbi.nlm.nih.gov/26186100/)]
12. Downs JS, de Bruin WB, Fischhoff B. Parents' vaccination comprehension and decisions. *Vaccine* 2008 Mar 17;26(12):1595-1607. [doi: [10.1016/j.vaccine.2008.01.011](https://doi.org/10.1016/j.vaccine.2008.01.011)] [Medline: [18295940](https://pubmed.ncbi.nlm.nih.gov/18295940/)]
13. Hakim H, Provencher T, Chambers CT, Driedger SM, Dube E, Gavaruzzi T, et al. Interventions to help people understand community immunity: a systematic review. *Vaccine* 2019 Jan 7;37(2):235-247. [doi: [10.1016/j.vaccine.2018.11.016](https://doi.org/10.1016/j.vaccine.2018.11.016)] [Medline: [30528593](https://pubmed.ncbi.nlm.nih.gov/30528593/)]
14. Michalos M, Tselenti P, Nalmpantis SL. Visualization techniques for large datasets. *Journal of Engineering Science and Technology Review* 2012 Mar;5(1):72-76. [doi: [10.25103/jestr.051.13](https://doi.org/10.25103/jestr.051.13)]
15. Bechara A, Damasio A. The somatic marker hypothesis: a neural theory of economic decision. *Games Econ Behav* 2005 Aug;52(2):336-372. [doi: [10.1016/j.geb.2004.06.010](https://doi.org/10.1016/j.geb.2004.06.010)]
16. Keltner D, Oatley K, Jenkins J. *Understanding Emotions*. United States: Wiley Global Education; 2013.
17. So J, Achar C, Han D, Agrawal N, Duhachek A, Maheswaran D. The psychology of appraisal: specific emotions and decision-making. *J Consum Psychol* 2015 Jul;25(3):359-371. [doi: [10.1016/j.jcps.2015.04.003](https://doi.org/10.1016/j.jcps.2015.04.003)]
18. Revenson TA, Pranikoff JR. A contextual approach to treatment decision making among breast cancer survivors. *Health Psychol* 2005 Jul;24(4S):S93-S98. [doi: [10.1037/0278-6133.24.4.S93](https://doi.org/10.1037/0278-6133.24.4.S93)] [Medline: [16045426](https://pubmed.ncbi.nlm.nih.gov/16045426/)]
19. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Risky feelings: why a 6% risk of cancer does not always feel like 6%. *Patient Educ Couns* 2010 Dec(81 Suppl):S87-S93 [FREE Full text] [doi: [10.1016/j.pec.2010.07.041](https://doi.org/10.1016/j.pec.2010.07.041)] [Medline: [20739135](https://pubmed.ncbi.nlm.nih.gov/20739135/)]
20. Ellis EM, Klein WM, Orehek E, Ferrer RA. Effects of emotion on medical decisions involving tradeoffs. *Med Decis Making* 2018 Nov;38(8):1027-1039 [FREE Full text] [doi: [10.1177/0272989X18806493](https://doi.org/10.1177/0272989X18806493)] [Medline: [30343624](https://pubmed.ncbi.nlm.nih.gov/30343624/)]
21. Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger J. Vaccine hesitancy: an overview. *Hum Vaccin Immunother* 2013 Aug;9(8):1763-1773 [FREE Full text] [doi: [10.4161/hv.24657](https://doi.org/10.4161/hv.24657)] [Medline: [23584253](https://pubmed.ncbi.nlm.nih.gov/23584253/)]
22. Tsuda A, Muis KR. The anti-vaccination debate: a cross-cultural exploration of emotions and epistemic cognition. *ASSRJ* 2018 Sep 30;5(9):- [doi: [10.14738/assrj.59.5215](https://doi.org/10.14738/assrj.59.5215)]
23. Chen N, Kinshuk P, Wei C, Chen H. Mining e-Learning domain concept map from academic articles. *Comput Educ* 2008 Apr 01;50(3):1009-1021. [doi: [10.1016/j.compedu.2006.10.001](https://doi.org/10.1016/j.compedu.2006.10.001)]
24. Novak JD. How do we learn our lesson? *Sci Teach* 1993;60(3):50-55.
25. Plotnick E. *Concept Mapping: a Graphical System for Understanding the Relationship Between Concepts*. ERIC - Education Resources Information Center. 1997. URL: <https://eric.ed.gov/?id=ED407938> [accessed 2020-09-25]
26. Rosenstock I. Historical origins of the health belief model. *Health Educ Monogr* 1974;2(4):328.
27. Wertheimer M. A brief introduction to gestalt, identifying key theories and principles. *Psychologische Forschung* 1923;4:301-350.
28. Mayer RE. *Cognitive Theory of Multimedia Learning*. Cambridge, England: Cambridge University Press; 1997.
29. Sorden SD. *The Cognitive Theory of Multimedia Learning*. North Carolina, USA: Charlotte: Information Age Publishing; 2012.
30. Slovic P, Finucane ML, Peters E, MacGregor DG. The affect heuristic. *European journal of operational research* 2007 Mar 16;177(3):1333-1352. [doi: [10.1016/j.ejor.2005.04.006](https://doi.org/10.1016/j.ejor.2005.04.006)]
31. Slovic P, Peters E, Finucane ML, Macgregor DG. Affect, risk, and decision making. *Health psychology* 2005 Jul;24(4S):S35-S40. [doi: [10.1037/0278-6133.24.4.S35](https://doi.org/10.1037/0278-6133.24.4.S35)] [Medline: [16045417](https://pubmed.ncbi.nlm.nih.gov/16045417/)]
32. Graham L. Gestalt theory in interactive media design. *J Hum Soc Sci* 2008;2(1):1-12 [FREE Full text] [doi: [10.1007/978-0-387-74378-3\\_2](https://doi.org/10.1007/978-0-387-74378-3_2)]

33. Witteman HO, Dansokho SC, Colquhoun H, Coulter A, Dugas M, Fagerlin A, et al. User-centered design and the development of patient decision aids: protocol for a systematic review. *Syst Rev* 2015 Jan 26;4:11 [FREE Full text] [doi: [10.1186/2046-4053-4-11](https://doi.org/10.1186/2046-4053-4-11)] [Medline: [25623074](https://pubmed.ncbi.nlm.nih.gov/25623074/)]
34. Kessler S, Guenther L. Eyes on the frame Explaining people's online searching behavior in response to TV consumption. *Internet Research* 2017;27(2):303-320 [FREE Full text] [doi: [10.1108/intr-01-2016-0015](https://doi.org/10.1108/intr-01-2016-0015)]
35. Nepal O, Jha RK, Bhattarai A, Khadka P, Kapoor BK. Galvanic skin response as a simple physiology lab teaching tool-an alternative indicator of sympathetic arousal. *Kathmandu Univ Med J (KUMJ)* 2018;16(62):156-160. [Medline: [30636757](https://pubmed.ncbi.nlm.nih.gov/30636757/)]
36. Fish CS, Griffin AL. Emotional Responses to Climate Change Map Framing Using Facial Emotion Recognition Technology. In: Conference Publications of the International Cartographic Association. 2019 Presented at: ICA'19; September 24-26, 2019; Tokyo, Japan. [doi: [10.5194/ica-abs-1-83-2019](https://doi.org/10.5194/ica-abs-1-83-2019)]
37. Hu W, Booth J, Reid T. The Relationship Between Design Outcomes and Mental States During Ideation. *J Mech Design* 2017 Mar 21;139(5):051101-051116. [doi: [10.1115/1.4036131](https://doi.org/10.1115/1.4036131)]
38. Guide. iMotion. 2016. URL: <https://imotions.com/guides/eye-tracking/>
39. PocketGuide. iMotion. 2016. URL: <https://imotions.com/guides/> [accessed 2020-09-25]
40. Guide GSR. iMotion. 2015. URL: <https://imotions.com/guides/> [accessed 2020-09-25]
41. Matsumoto D, Ekman P. Facial Expression Analysis. iMotion. 2008. URL: <https://imotions.com/guides/> [accessed 2020-09-25]
42. Press and News. iMotions: Unpack Human Behavior. URL: <https://imotions.com/press-release/imotions-launches-attention-tool-50-the-complete-biometric-research-platform/> [accessed 2018-07-19]
43. Bojko A. Eye Tracking the User Experience: A Practical Guide to Research. United States: Rosenfeld Media; 2013.
44. Mele M, Millar D, Rijnders C. Explicit and Implicit Measures in Video Quality Assessment. In: 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. 2019 Presented at: VISIGRAPP'19; February 25-27, 2019; Prague, Czech Republic. [doi: [10.5220/0007396100380049](https://doi.org/10.5220/0007396100380049)]
45. Duchowski A. Eye Tracking Methodology: Theory and Practice. Second Edition. New York, USA: Springer Science & Business Media; 2007.
46. Holland J. Eye tracking: biometric evaluations of instructional materials for improved learning. *Int J Educ Pedagog Sci* 2019;13(7):1001-1008.
47. Shimmer Sensing. 2013. URL: <http://www.shimmersensing.com/support/wireless-sensor-networks-documentation/> [accessed 2020-09-08]
48. Akash K, Hu W, Jain N, Reid T. A classification model for sensing human trust in machines using EEG and GSR. *ACM Trans Interact Intell Syst* 2018 Nov 22;8(4):1-20. [doi: [10.1145/3132743](https://doi.org/10.1145/3132743)]
49. Muller L, Bernin A, Kamenz A, Ghose S, von Luck K, Grecos C, et al. Emotional Journey for an Emotion Provoking Cycling Exergame. 2017 Presented at: ISCM'17; November 23-24, 2017; Port Louis, Mauritius. [doi: [10.1109/iscmi.2017.8279607](https://doi.org/10.1109/iscmi.2017.8279607)]
50. Terkildsen T, Makransky G. Measuring presence in video games: an investigation of the potential use of physiological measures as indicators of presence. *International journal of human-computer studies* 2019;126:64-80. [doi: [10.1016/j.jchcs.2019.02.006](https://doi.org/10.1016/j.jchcs.2019.02.006)]
51. Emotient: Use Facial Expressions to Understand Emotional Reactions. iMotions. 2016. URL: <https://imotions.com/> [accessed 2020-09-08]
52. Stöckli S, Schulte-Mecklenbeck M, Borer S, Samson A. A practical guide for automated facial emotion classification 1. A Handbook of Process Tracing Methods 2019:198-216. [doi: [10.4324/9781315160559-14](https://doi.org/10.4324/9781315160559-14)]
53. González-Rodríguez M, Díaz-Fernández M, Pacheco Gómez C. Facial-expression recognition: an emergent approach to the measurement of tourist satisfaction through emotions. *Telematics and Informatics* 2020 Aug;51:101404. [doi: [10.1016/j.tele.2020.101404](https://doi.org/10.1016/j.tele.2020.101404)]
54. Berka C, Levendowski DJ, Cvetinovic MM, Petrovic MM, Davis G, Lumicao MN, et al. Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *Int J Hum-Comput Int* 2004 Jun;17(2):151-170. [doi: [10.1207/s15327590ijhc1702\\_3](https://doi.org/10.1207/s15327590ijhc1702_3)]
55. Berka C, Levendowski DJ, Lumicao MN, Yau A, Davis G, Zivkovic VT, et al. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat Space Environ Med* 2007 May;78(5 Suppl):B231-B244. [Medline: [17547324](https://pubmed.ncbi.nlm.nih.gov/17547324/)]
56. Ortiz de Guinea A, Titah R, Léger P. Measure for measure: a two study multi-trait multi-method investigation of construct validity in IS research. *Comput Hum Behav* 2013 May;29(3):833-844. [doi: [10.1016/j.chb.2012.12.009](https://doi.org/10.1016/j.chb.2012.12.009)]
57. Betsch C, Böhm R, Korn L. Inviting free-riders or appealing to prosocial behavior? game-theoretical reflections on communicating herd immunity in vaccine advocacy. *Health Psychol* 2013 Sep;32(9):978-985. [doi: [10.1037/a0031590](https://doi.org/10.1037/a0031590)] [Medline: [24001248](https://pubmed.ncbi.nlm.nih.gov/24001248/)]
58. van Alstyne JA, Nowak GJ, Aikin AL. What is 'confidence' and what could affect it?: A qualitative study of mothers who are hesitant about vaccines. *Vaccine* 2018 Oct 22;36(44):6464-6472. [doi: [10.1016/j.vaccine.2017.09.007](https://doi.org/10.1016/j.vaccine.2017.09.007)] [Medline: [28899629](https://pubmed.ncbi.nlm.nih.gov/28899629/)]

59. Nowak GJ, Evans NJ, Wojdowski BW, Ahn SJG, Len-Rios ME, Carera K, et al. Using immersive virtual reality to improve the beliefs and intentions of influenza vaccine avoidant 18-to-49-year-olds: considerations, effects, and lessons learned. *Vaccine* 2020 Jan 29;38(5):1225-1233. [doi: [10.1016/j.vaccine.2019.11.009](https://doi.org/10.1016/j.vaccine.2019.11.009)] [Medline: [31806533](https://pubmed.ncbi.nlm.nih.gov/31806533/)]
60. Rieber LP, Tzeng S, Tribble K. Discovery learning, representation, and explanation within a computer-based simulation: finding the right mix. *Learn Instr* 2004 Jun;14(3):307-323. [doi: [10.1016/j.learninstruc.2004.06.008](https://doi.org/10.1016/j.learninstruc.2004.06.008)]
61. Luzón JM, Letón E. Use of animated text to improve the learning of basic mathematics. *Comput Educ* 2015 Oct;88:119-128. [doi: [10.1016/j.compedu.2015.04.016](https://doi.org/10.1016/j.compedu.2015.04.016)]
62. Hasana SN, Alifiani A. Multimedia development using visual basic for application (VBA) to improve students' learning motivation in studying mathematics of economics. *Int J Math Educ* 2019 Apr 30;2(1):34. [doi: [10.31002/ijome.v2i1.1230](https://doi.org/10.31002/ijome.v2i1.1230)]
63. Breakey VR, Warias AV, Ignas DM, White M, Blanchette VS, Stinson JN. The value of usability testing for internet-based adolescent self-management interventions: 'managing hemophilia online'. *BMC Med Inform Decis Mak* 2013 Oct 4;13:113 [FREE Full text] [doi: [10.1186/1472-6947-13-113](https://doi.org/10.1186/1472-6947-13-113)] [Medline: [24094082](https://pubmed.ncbi.nlm.nih.gov/24094082/)]
64. Jonas E, Schulz-Hardt S, Frey D. Giving advice or making decisions in someone else's place: the influence of impression, defense, and accuracy motivation on the search for new information. *Pers Soc Psychol Bull* 2005 Jul;31(7):977-990. [doi: [10.1177/0146167204274095](https://doi.org/10.1177/0146167204274095)] [Medline: [15951368](https://pubmed.ncbi.nlm.nih.gov/15951368/)]
65. Zikmund-Fisher B, Sarr B, Fagerlin A, Ubel P. A matter of perspective: choosing for others differs from choosing for yourself in making treatment decisions. *J Gen Intern Med* 2006;21(6):618-622. [doi: [10.1111/j.1525-1497.2006.00410.x](https://doi.org/10.1111/j.1525-1497.2006.00410.x)]
66. Peters E, Lipkus I, Diefenbach M. The functions of affect in health communications and in the construction of health preferences. *J Commun* 2006;56:S140-S162. [doi: [10.1111/j.1460-2466.2006.00287.x](https://doi.org/10.1111/j.1460-2466.2006.00287.x)]
67. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. *J Fam Pract* 2000 Feb;49(2):147-152. [Medline: [10718692](https://pubmed.ncbi.nlm.nih.gov/10718692/)]
68. Brown KF, Kroll JS, Hudson MJ, Ramsay M, Green J, Long SJ, et al. Factors underlying parental decisions about combination childhood vaccinations including MMR: a systematic review. *Vaccine* 2010 Jun 11;28(26):4235-4248. [doi: [10.1016/j.vaccine.2010.04.052](https://doi.org/10.1016/j.vaccine.2010.04.052)] [Medline: [20438879](https://pubmed.ncbi.nlm.nih.gov/20438879/)]

## Abbreviations

**EEG:** electroencephalogram

*Edited by G Eysenbach; submitted 11.05.20; peer-reviewed by P Wegier, J Wijesundara; comments to author 21.06.20; revised version received 03.07.20; accepted 26.07.20; published 30.10.20.*

### *Please cite as:*

Hakim H, Bettinger JA, Chambers CT, Driedger SM, Dubé E, Gavaruzzi T, Giguere AMC, Kavanagh É, Leask J, MacDonald SE, Orji R, Parent E, Paquette JS, Roberge J, Sander B, Scherer AM, Tremblay-Breault M, Wilson K, Reinharz D, Witteman HO  
*A Web Application About Herd Immunity Using Personalized Avatars: Development Study*  
*J Med Internet Res* 2020;22(10):e20113  
URL: <https://www.jmir.org/2020/10/e20113>  
doi: [10.2196/20113](https://doi.org/10.2196/20113)  
PMID: [33124994](https://pubmed.ncbi.nlm.nih.gov/33124994/)

©Hina Hakim, Julie A Bettinger, Christine T Chambers, S Michelle Driedger, Eve Dubé, Teresa Gavaruzzi, Anik M C Giguere, Éric Kavanagh, Julie Leask, Shannon E MacDonald, Rita Orji, Elizabeth Parent, Jean-Sébastien Paquette, Jacynthe Roberge, Beate Sander, Aaron M Scherer, Martin Tremblay-Breault, Kumanan Wilson, Daniel Reinharz, Holly O Witteman. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Patients' Adoption of Electronic Personal Health Records in England: Secondary Data Analysis

Alaa Abd-Alrazaq<sup>1</sup>, PhD; Ali Abdallah Alalwan<sup>2</sup>, PhD; Brian McMillan<sup>3</sup>, PhD; Bridgette M Bewick<sup>4</sup>, PhD; Mowafa Househ<sup>1</sup>, PhD; Alaa T AL-Zyadat<sup>2</sup>, MSc

<sup>1</sup>Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

<sup>2</sup>Amman University College for Banking and Financial Sciences, Al-Balqa Applied University, Amman, Jordan

<sup>3</sup>Centre for Primary Care and Health Services Research, University of Manchester, Manchester, United Kingdom

<sup>4</sup>Leeds Institute of Health Sciences, School of Medicine, University of Leeds, Leeds, United Kingdom

**Corresponding Author:**

Alaa Abd-Alrazaq, PhD

Division of Information and Computing Technology

College of Science and Engineering

Hamad Bin Khalifa University

LAS Building, Education City

Doha,

Qatar

Phone: 974 55708549

Email: [alaa\\_alzoubi88@yahoo.com](mailto:alaa_alzoubi88@yahoo.com)

## Abstract

**Background:** In England, almost all general practices (GPs) have implemented GP online services such as electronic personal health records (ePHRs) that allow people to schedule appointments, request repeat prescriptions, and access parts of their medical records. The overall adoption rate of GP online services has been low, reaching just 28% in October 2019. In a previous study, Abd-Alrazaq et al adopted a model to assess the factors that influence patients' use of GP online services in England. According to the previous literature, the predictive power of the Abd-Alrazaq model could be improved by proposing new associations between the existing variables in the model.

**Objective:** This study aims to improve the predictive power of the Abd-Alrazaq model by proposing new relationships between the existing variables in the model.

**Methods:** The Abd-Alrazaq model was amended by proposing new direct, mediating, moderating, and moderated mediating effects. The amended model was examined using data from a previous study, which were collected by a cross-sectional survey of a convenience sample of 4 GPs in West Yorkshire, England. Structural equation modeling was used to examine the theoretical model and hypotheses.

**Results:** The new model accounted for 53% of the variance in performance expectancy (PE), 76% of the variance in behavioral intention (BI), and 49% of the variance in use behavior (UB). In addition to the significant associations found in the previous study, this study found that social influence (SI) and facilitating conditions (FCs) are associated with PE directly and BI indirectly through PE. The association between BI and UB was stronger for younger women with higher levels of education, income, and internet access. The indirect effects of effort expectancy (EE), perceived privacy and security (PPS), and SI on BI were statistically stronger for women without internet access, patients with internet access, and patients without internet access, respectively. The indirect effect of PPS on BI was stronger for patients with college education or diploma than for those with secondary school education and lower, whereas the indirect effect of EE on BI was stronger for patients with secondary school education or lower than for those with college education or a diploma.

**Conclusions:** The predictive power of the Abd-Alrazaq model improved by virtue of new significant associations that were not examined before in the context of ePHRs. Further studies are required to validate the new model in different contexts and to improve its predictive power by proposing new variables. The influential factors found in this study should be considered to improve patients' use of ePHRs.

(*J Med Internet Res* 2020;22(10):e17499) doi:[10.2196/17499](https://doi.org/10.2196/17499)



**KEYWORDS**

health records, personal; patient portal; medical informatics

## Introduction

### Background

An electronic personal health record (ePHR) has been defined by the Markle Foundation as “an electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure and confidential environment” [1]. Several services can also be provided by more advanced ePHRs, such as requesting repeat prescriptions, booking appointments, viewing test results, and messaging providers [2-4]. ePHRs have the potential to empower patients [5,6], improve patient self-management and medication adherence [7,8], enhance the rapport and communication between patients and health care providers [9,10], ease access to health services [11,12], avoid duplicated tests and medical images [9,11], and decrease adverse events [4,9,11,13].

In England, almost all general practices (GPs) have implemented GP online services, that is, ePHRs that allow people to schedule appointments, request repeat prescriptions, and access coded information in their medical records, such as demographics, medications, allergies, and test results [14]. The number of providers offering GP online services is growing [15].

### Research Problems and Aims

Despite the aforementioned potential benefits of ePHRs, the overall adoption rate of GP online services has been low, reaching just 28% in October 2019 [16]. To improve the adoption and implementation of ePHRs, it is important to identify the factors that influence individuals' use of the system [17-23]. A recent systematic review of 97 studies found that more than 150 factors could affect patients' acceptance and adoption of ePHRs [24]. Unfortunately, none of these studies were carried out in the United Kingdom and included a number of shortcomings, namely, few studies were theory based [21,25-28], many focused on factors affecting patients' intention to use ePHRs instead of actual use [29-32], many assessed factors affecting self-reported use rather than actual use [28,33-36], almost all examined independent and dependent variables at one point in time using the same data collection instrument and were therefore at risk of common method bias [26,33,37], and almost all the studies did not differentiate between factors affecting initial use and continuing use of ePHRs. Therefore, Abd-Alrazaq et al [38] conducted a cross-sectional survey to assess the factors that influence patients' use of ePHRs in England. The study identified several significant factors (performance expectancy [PE], effort expectancy [EE], perceived privacy and security [PPS], behavioral intention [BI], and some moderators), which were able to predict 48% of the variance in use behavior (UB). On the basis of previous research, we propose an amended model

that we expect will predict UB more accurately. This study aims to improve the predictive power of the Abd-Alrazaq model by proposing new relationships between the variables existing in the model.

### Theoretical Foundation

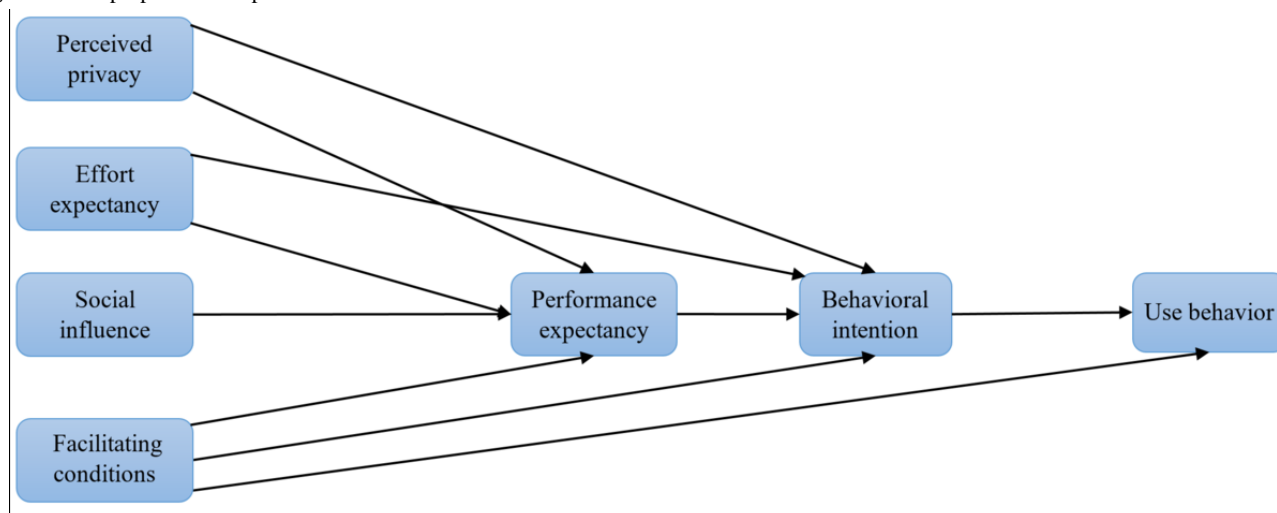
In a previous study [38], the Unified Theory of Acceptance and Use of Technology (UTAUT) [39] was selected from 12 models as the theoretical foundation. The selection process was based on 6 predefined criteria: 4 criteria related to the goodness of the theory (ie, logical consistency, explanatory power, falsifiability, and parsimony), and 2 criteria related to the applicability of the theory on the phenomena of interest (ie, population and type of behavior). [Multimedia Appendix 1](#) [27,29,37,39-65] elaborates on the selection process of the appropriate theory. Abd-Alrazaq et al [38] adapted UTAUT to the context of ePHRs by removing experience and voluntariness and adding PPS, education, income, and internet access to the model ([Multimedia Appendix 2](#)). Their justifications for these adaptations are provided in [Multimedia Appendix 3](#).

Given that the study did not find a significant association between social influence (SI) and BI, the authors recommended that researchers examine other associations of SI [38]. Several studies have found that SI positively affects PE [66-68]. In other words, individuals who perceive that using technology is recommended by those important to them are more likely to perceive that the technology is useful. Therefore, this study proposes that SI directly affects PE and indirectly affects BI through PE.

The 2019 Abd-Alrazaq model [38] could be criticized for failing to hypothesize that facilitating conditions (FCs) are associated with BI. This argument is in line with the findings of Venkatesh et al [69], who proposed this relationship in the extended Unified Theory of Acceptance and Use of Technology (UTAUT2) framework, which is suitable for the consumer context [69]. Several studies have found that FCs are also associated with PE [70-72]. Accordingly, this study proposed that FCs directly affect both PE and BI and indirectly affect BI through PE.

The study by Abd-Alrazaq et al [38] also highlighted the need to assess the effect of moderators on indirect relationships (ie, moderated mediation) in the context of ePHRs. To address this recommendation, this study hypothesized that all mediating effects are moderated by sex, education, income, and internet access. To the best of our knowledge, no studies have assessed the moderating effects of these variables on the relationship between BI and UB. Thus, we explored the moderating effect of age, sex, education, income, and internet access on the relationship between BI and UB. Our proposed model and hypotheses are presented in [Figure 1](#) and [Table 1](#), respectively. The conceptual definitions of the constructs in the proposed model are presented in [Multimedia Appendix 4](#).



**Figure 1.** The proposed conceptual model.

**Table 1.** The proposed research hypotheses.

Hypothesis #	Hypothesis
H1	PPS <sup>a</sup> positively affects PE <sup>b</sup> .
H2	PPS positively affects BI <sup>c</sup> .
H3	PPS indirectly and positively affects BI through PE.
H4	The positive relationship between PPS and PE is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a higher level of education and lower income and with internet access.
H5	The positive relationship between PPS and BI is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a higher level of education and lower income and with internet access.
H6	The indirect effect of PPS on BI is moderated by sex, education, income, and internet access, such that the influence is stronger for women with a higher level of education and lower income and with internet access.
H7	EE <sup>d</sup> positively affects PE.
H8	EE positively affects BI.
H9	EE indirectly and positively affects BI through PE.
H10	The positive relationship between EE and PE is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a lower level of education and income and without internet access.
H11	The positive relationship between EE and BI is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a lower level of education and income and without internet access.
H12	The indirect effect of EE on BI is moderated by sex, education, income, and internet access, such that the influence is stronger for women with a lower level of education and income and without internet access.
H13	SI <sup>e</sup> positively affects PE.
H14	SI indirectly and positively affects BI through PE.
H15	The positive relationship between SI and PE is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a lower level of education and income and with internet access.
H16	The indirect effect of SI on BI is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a lower level of education and income and with internet access.
H17	FCs <sup>f</sup> positively affect PE.
H18	FCs positively affect BI.
H19	FCs positively affect UB <sup>g</sup> .
H20	FCs indirectly and positively affect BI through PE.
H21	The positive relationship between FCs and PE is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a lower level of education and income and without internet access.
H22	The positive relationship between FCs and BI is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a lower level of education and income and without internet access.
H23	The positive relationship between FCs and UB is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a lower level of education and income and without internet access.
H24	The indirect effect of FCs on BI is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a lower level of education and income and without internet access.
H25	The indirect effect of FCs on UB is moderated by age, sex, education, income, and internet access, such that the influence is stronger for older women with a lower level of education and income and without internet access.
H26	PE positively affects BI.
H27	The positive relationship between PE and BI is moderated by age and sex, such that the influence is stronger for younger men with a lower level of education, higher income, and internet access
H28	BI positively affects UB.
H29	The positive relationship between BI and UB is moderated by age, sex, education, income, and internet access, such that the influence is stronger for younger women with a higher level of education and income and with internet access.

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.

<sup>d</sup>EE: effort expectancy.

<sup>e</sup>SI: social influence.

<sup>f</sup>FC: facilitating condition.

<sup>g</sup>UB: use behavior.

## Methods

### Study Design and Setting

This study used secondary data analysis from data collected by Abd-Alrazaq et al [38] using a cross-sectional survey of 4 West Yorkshire GPs (Multimedia Appendix 5). Health Research Authority approval was granted before starting data collection (REC reference: 17/SC/0323).

### Measurement

All variables except UB were measured using self-administered questionnaires. The questionnaires were composed of 29 questions adopted from previous research (Multimedia Appendix 6). A panel of experts evaluated the face validity and content validity of the questions, and based on their suggestions, the questionnaire was amended and sent via email to 37 patients for pilot testing. The questionnaire was subsequently amended slightly because of the issues reported by patients (Multimedia Appendix 7). UB was measured objectively using system logs by extracting data on the number of times that each participant logged into the system during the 6 months after completing the questionnaire. One open-ended question was added to the questionnaire to obtain qualitative data that enabled the exploration of additional factors. The qualitative data were analyzed using thematic analysis, and the results have been reported elsewhere.

### Recruitment

A convenience sampling approach was used to recruit patients. Patients were eligible to participate if they were living in England, were registered at 1 of the 4 GP practices, were aged 18 years or older, and had not used GP online services before (nonusers). The questionnaire was delivered to eligible participants visiting 1 of the 4 GP practices during the study period. Data on participants' use of GP online services were extracted from the system logs after 6 months of completing the questionnaire.

### Statistical Analysis

Before assessing the proposed model, it is a prerequisite to check normality [73,74], linearity [73], multicollinearity [73], and common method bias [75,76]. Univariate normality was examined by assessing skewness and kurtosis [73,77]. This study checked the linearity between each proposed relationship using scatterplot graphs [73] and the curve estimation procedure [78]. Multicollinearity was assessed in this study using tolerance, which refers to the proportion of the variability of one predictor that is unexplained by other predictors [73,77]. We checked the common method bias using Harman single-factor test [75]. All the aforementioned analyses were carried out using SPSS v.22 (IBM).

The theoretical model and hypotheses were examined using structural equation modeling (SEM). In SEM, models consist

of 2 elements: a measurement model in which the relationships between observed variables and latent variables are examined, and a structural model in which the relationships proposed among the latent variables are assessed [73,79]. Although the measurement model in this study is identical to the original study [38], it was reassessed just for the sake of completeness. The measurement model was examined in terms of 3 aspects: model fit, construct reliability, and construct validity [73,77]. The structural model was then assessed for model fit, predictive power, and strength of relationships [77,79,80]. The strength of relationships was tested using different approaches depending on the type of the proposed effect. To be more precise, path coefficients were checked to examine direct effects [81]. Mediating effects were examined by assessing the indirect effects of using bootstrapping. The moderating effect for the metric moderator (ie, age) was examined using the interaction effect method [73,82]. The moderating effects for the nonmetric moderator (sex) were tested using multigroup SEM [73,74,82]. Moderated mediating effects were assessed using multigroup SEM for indirect effects. All analyses were conducted using the Analysis of Moment Structures v.24 (IBM) software.

## Results

### Participants' Characteristics

The response rate was 78.0% (624/800). As shown in Multimedia Appendix 8, the mean age of participants was 44.2 (SD 1.89) years. Most participants were White (498/624, 79.8%) and had internet access (528/624, 84.6%). About half of the sample (284/624, 45.5%) had an income level of less than US \$25,000 per year. The most prominent education levels among respondents were bachelor's degrees (174/624, 27.9%), college or diploma (165/624, 26.4%), and secondary school (147/624, 23.6%). There were no significant differences between participants and the target population in terms of age, sex, and ethnicity ( $P=.21$ ,  $P=.06$ , and  $P=.64$ , respectively; Multimedia Appendix 8). Thus, the risk of nonresponse bias is minimal.

### Normality, Linearity, Multicollinearity, and Common Method Bias

Histograms presented in Multimedia Appendix 9 show no severe skewness and kurtosis for all items. This finding was confirmed by the absolute values of skewness and kurtosis, which were considerably less than the cutoff points of 3 and 10, respectively [77] (Multimedia Appendix 10).

According to the scatterplots shown in Multimedia Appendix 11, there was an indication of possible nonlinearity for only 2 relationships: the effect of BI and FCs on UB. However, the results of the curve estimation procedure showed that the F values for all proposed relationships in the linear model were significant and higher than the F values of the proposed relationships in the 10 nonlinear models, indicating that all proposed relationships between variables are linear (Multimedia Appendix 12).

As shown in [Multimedia Appendix 13](#), all values of tolerance are within the predetermined cutoff point ( $\geq 0.2$ ) [83], indicating that there is no serious multicollinearity between independent variables.

With regard to the common method bias, 5 factors emerged from the Harman single-factor test; a single factor was able to explain less than half of the variance (47.3%; [Multimedia Appendix 14](#)). This means that there are no concerns regarding the presence of the common method bias in this study.

## Measurement Model

### Model Fit

Nine indices were used to assess the absolute model fit (chi-square/df, goodness-of-fit index, adjusted goodness-of-fit index, root mean square error of approximation, p of Close Fit, and standardized root mean square residual) and incremental fit (normed-fit index, comparative fit index, and Tucker-Lewis index) [73,77]. Given that the measurement model in this study is identical to the modified measurement model in the original study, the results of the fit indices were the same between the 2 studies and were within their suggested levels ([Multimedia Appendix 15](#)). This indicates that the measurement model adequately fits the collected data.

### Construct Reliability

Three measures were used to assess the construct reliability: Cronbach alpha ( $\alpha$ ), composite reliability (CR), and average variance extracted (AVE). Yielded values of  $\alpha$ , CR, and AVE for each construct were within their recommended values of  $\geq .70$ ,  $\geq .70$ , and  $\geq .50$ , respectively ([Multimedia Appendix 16](#)) [73,77]. This means that the measurement items are consistent and reproducible in measuring what it is assumed to measure.

## Construct Validity

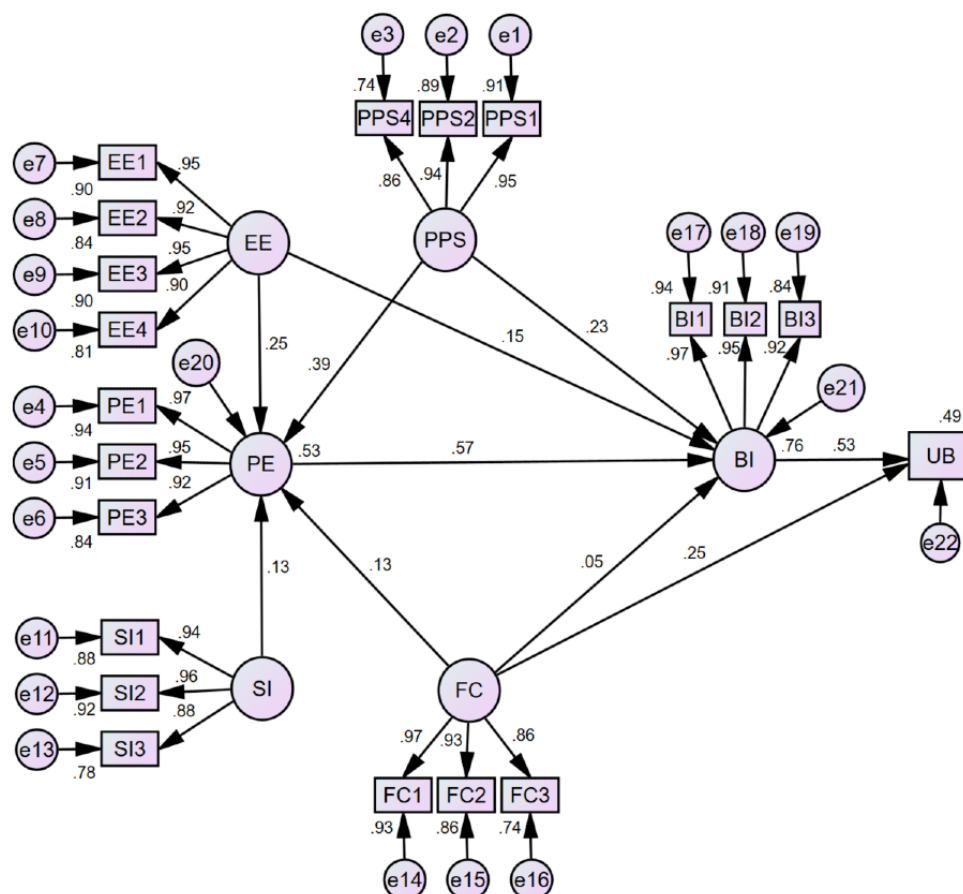
Two components of construct validity were examined in this study: convergent validity and discriminant validity [73,77]. The convergent validity was examined by checking factor loadings and the AVE [73]. As shown in [Multimedia Appendix 17](#), the values of factor loading and AVE for all items considerably exceeded the thresholds of .70 and .50, respectively [73]. These results indicate that each item relates strongly to the latent variable that it is assumed to measure.

Discriminant validity was assessed by checking intercorrelation coefficients, comparing the square root of AVE with the intercorrelation coefficients, and comparing loadings and cross-loadings [73,77,81]. [Multimedia Appendix 18](#) shows that the intercorrelation coefficients (off-diagonal values) are located within acceptable ranges ( $< .85$ ) [84]. Furthermore, each value of the square root of AVE for a construct (values on the diagonal) is higher than all intercorrelation coefficients between that construct and each other construct ([Multimedia Appendix 18](#)). As shown in [Multimedia Appendix 19](#), the loading of each item on its construct was higher than the cross-loadings in rows and columns. The results of the three measures indicate that items of each construct are not related to the other constructs that it is not supposed to measure; therefore, the measurement model has acceptable discriminant validity.

## Structural Model

### Model Fit and Predictive Power

The indices that were used to assess the fit of the measurement model were used again to assess the fit of the structural model. As shown in [Multimedia Appendix 20](#), all fit indices were within the recommended values, indicating that the structural model adequately fits the collected data. The model was able to predict about 0.53 of the variance in PE, 0.76 of the variance in BI, and 0.49 of the variance in UB ([Figure 2](#)).

**Figure 2.** Structural model estimates.

### Strengths of Relationships

#### Direct Effects

As seen in Figure 2 and detailed in Table 2, all proposed direct effects were statistically significant, except for the effect of FCs on BI ( $\beta=.05$ ,  $P=.08$ ). Specifically, PPS was significantly associated with PE ( $\beta=.39$ ) and BI ( $\beta=.23$ ). The paths from EE

to PE and BI were statistically significant ( $\beta=.25$  and  $\beta=.15$ , respectively). There was a statistically significant relationship between SI and PE ( $\beta=.13$ ). FCs were significantly associated with PE ( $\beta=.13$ ) and UB ( $\beta=.25$ ). The relationship between PE and BI was statistically significant ( $\beta=.57$ ). BI and UB were significantly associated ( $\beta=.76$ ). To sum up, the following hypotheses were supported: H1, H2, H7, H8, H13, H17, H19, H26, and H28 (Table 2).



**Table 2.** Results of the direct effects.

Hypothesis #	Path	Standardized estimate ( $\beta$ )	95% CI	P value
H1	PPS <sup>a</sup> →PE <sup>b</sup>	.39	0.32 to 0.46	<.001
H2	PPS→BI <sup>c</sup>	.23	0.17 to 0.29	<.001
H7	EE <sup>d</sup> →PE	.25	0.18 to 0.32	<.001
H8	EE→BI	.15	0.10 to 0.21	<.001
H13	SI <sup>e</sup> →PE	.13	0.04 to 0.22	<.001
H17	FCs <sup>f</sup> →PE	.13	0.06 to 0.20	<.001
H18	FCs→BI	.05	−0.003 to 0.10	.08
H19	FCs→UB <sup>g</sup>	.25	0.20 to 0.30	<.001
H26	PE→BI	.57	0.51 to 0.64	<.001
H28	BI→UB	.53	0.47 to 0.58	<.001

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.<sup>g</sup>UB: use behavior.**Mediating Effects**

With regard to the mediating effects, results showed that PE significantly mediated the effect of PPS, EE, SI, and FCs on BI

( $\beta=.22$ ,  $\beta=.14$ ,  $\beta=.09$ , and  $\beta=.07$ , respectively; Table 3). Accordingly, H3, H9, H14, and H20 were supported in this study.

**Table 3.** Results of the mediating effects.

Hypothesis #	Indirect effect	Standardized estimate ( $\beta$ )	95% CI	P value
H3	PPS <sup>a</sup> →PE <sup>b</sup> →BI <sup>c</sup>	.22	0.18-0.28	<.001
H9	EE <sup>d</sup> →PE→BI	.15	0.10-0.19	<.001
H14	SI <sup>e</sup> →PE→BI	.09	0.04-0.14	<.001
H20	FCs <sup>f</sup> →PE→BI	.07	0.03-0.11	.002

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.**Moderating Effects**

With respect to the moderating effects, the effect of EE and FCs on BI statistically increased with increasing age ( $P=.03$ ,  $P<.001$ ,

respectively; Table 4). In contrast, the effect of PE on BI and the effect of BI on UB statistically decreased with increasing age ( $P<.001$ , for both moderating effects).

**Table 4.** Results of the moderating effect of age.

Hypothesis #	Interaction effect	Standardized estimate ( $\beta$ )	<i>P</i> value
H4	PPS <sup>a</sup> ×age→PE <sup>b</sup>	.18	.66
H5	PPS×age→BI <sup>c</sup>	-.02	.25
H10	EE <sup>d</sup> ×age→PE	.14	.22
H11	EE×age→BI	.05	.03
H15	SI <sup>e</sup> ×age→PE	.03	.45
H21	FCs <sup>f</sup> ×age→PE	.21	.30
H22	FCs×age→BI	.03	.10
H23	FCs×age→UB <sup>g</sup>	.16	<.001
H27	PE×age→BI	-.10	<.001
H29	BI×age→UB	-.21	<.001

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.<sup>g</sup>UB: use behavior.

Concerning the moderating effects of sex, the association between PE and BI was statistically stronger for men than for women ( $\beta=.59$  vs  $\beta=.50$ ,  $P=.004$ ; Table 5). The path from BI to UB was statistically stronger for women than for men ( $\beta=.53$  vs  $\beta=.03$ ,  $P=.001$ ).

**Table 5.** Results of the moderating effect of sex.

Hypothesis #	Hypothesized path	Men		Women		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H4	PPS <sup>a</sup> →PE <sup>b</sup>	.41	<.001	.32	<.001	.81
H5	PPS→BI <sup>c</sup>	.25	<.001	.21	<.001	.39
H10	EE <sup>d</sup> →PE	.22	<.001	.26	<.001	.12
H11	EE→BI	.17	<.01	.17	<.01	.19
H15	SI <sup>e</sup> →PE	.068	.30	.17	<.001	.09
H21	FCs <sup>f</sup> →PE	.08	.23	.16	<.01	.14
H22	FCs→BI	.05	.35	.04	.33	.86
H23	FCs→UB <sup>g</sup>	.34	<.001	.24	<.001	.32
H27	PE→BI	.59	<.001	.50	<.001	.004
H29	BI→UB	.29	<.001	.53	<.001	.001

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.<sup>g</sup>UB: use behavior.

In relation to the moderating effect of education (Tables 6-8), the association between EE and PE was statistically stronger for the “secondary school or lower” group than for the “bachelor or higher” group ( $\beta=.31$  vs  $\beta=.01$ ,  $P=.049$ ). The association between EE and BI was statistically weaker for the “bachelor or higher” group than for the “secondary school or lower” group ( $\beta=-.08$  vs  $\beta=.13$ ,  $P=.04$ ) and for the college group ( $\beta=-.08$  vs  $\beta=.12$ ,  $P=.02$ ; Tables 6-8). The path from FCs to UB was statistically stronger for the “secondary school or lower” group

than for the college group ( $\beta=.38$  vs  $\beta=.29$ ,  $P=.003$ ) and the “bachelor or higher” group ( $\beta=.38$  vs  $\beta=.21$ ,  $P=.03$ ). The relationship between BI and UB was statistically stronger for the “bachelor or higher” group than for the “secondary school or lower” group ( $\beta=.48$  vs  $\beta=.14$ ,  $P<.001$ ) and the college group ( $\beta=.48$  vs  $\beta=.39$ ,  $P=.003$ ). The relationship between BI and UB was statistically stronger for the college group than for the “secondary school or lower” group ( $\beta=.39$  vs  $\beta=.14$ ,  $P<.001$ ).

**Table 6.** Results of the moderating effect of education level (secondary school vs college).

Hypothesis #	Hypothesized path	Secondary school or lower		College or diploma		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H4	PPS <sup>a</sup> →PE <sup>b</sup>	.22	<.001	.25	<.001	.41
H5	PPS→BI <sup>c</sup>	.15	.01	.28	<.001	.18
H10	EE <sup>d</sup> →PE	.31	<.001	.14	.045	.93
H11	EE→BI	.13	.02	.12	.008	.43
H15	SI <sup>e</sup> →PE	.10	.13	.14	.047	.42
H21	FCs <sup>f</sup> →PE	.20	.004	.07	.31	.56
H22	FCs→BI	.05	.40	.02	.64	.82
H23	FCs→UB <sup>g</sup>	.38	<.001	.29	<.001	.003
H27	PE→BI	.55	<.001	.62	<.001	.53
H29	BI→UB	.14	.04	.39	<.001	.001

<sup>a</sup>PPS: perceived privacy and security.

<sup>b</sup>PE: performance expectancy.

<sup>c</sup>BI: behavioral intention.

<sup>d</sup>EE: effort expectancy.

<sup>e</sup>SI: social influence.

<sup>f</sup>FC: facilitating condition.

<sup>g</sup>UB: use behavior.

**Table 7.** Results of the moderating effect of education level (secondary school vs bachelor or higher).

Hypothesis #	Hypothesized path	Secondary school or lower		Bachelor or higher		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H4	PPS <sup>a</sup> →PE <sup>b</sup>	.22	<.001	.18	.03	.98
H5	PPS→BI <sup>c</sup>	.15	.01	.28	<.001	.07
H10	EE <sup>d</sup> →PE	.31	<.001	.01	.89	.049
H11	EE→BI	.13	.02	-.08	.20	.04
H15	SI <sup>e</sup> →PE	.10	.13	.17	.02	.37
H21	FCs <sup>f</sup> →PE	.20	.004	.18	.02	.27
H22	FCs→BI	.05	.40	.004	.95	.82
H23	FCs→UB <sup>g</sup>	.38	<.001	.21	.002	.03
H27	PE→BI	.55	<.001	.59	<.001	.24
H29	BI→UB	.14	.04	.48	<.001	.001

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.<sup>g</sup>UB: use behavior.**Table 8.** Results of the moderating effect of education level (college vs bachelor or higher).

Hypothesis #	Hypothesized path	College or diploma		Bachelor or higher		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H4	PPS <sup>a</sup> →PE <sup>b</sup>	.25	<.001	.18	.03	.12
H5	PPS→BI <sup>c</sup>	.28	<.001	.28	<.001	.19
H10	EE <sup>d</sup> →PE	.14	.045	.01	.89	.17
H11	EE→BI	.12	.008	-.08	.20	.02
H15	SI <sup>e</sup> →PE	.14	.047	.17	.02	>.99
H21	FCs <sup>f</sup> →PE	.07	.31	.18	.02	.17
H22	FCs→BI	.02	.64	.004	.95	.91
H23	FCs→UB <sup>g</sup>	.29	<.001	.21	.002	.26
H27	PE→BI	.62	<.001	.59	<.001	.06
H29	BI→UB	.39	<.001	.48	<.001	.003

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.<sup>g</sup>UB: use behavior.

As shown in [Tables 9-11](#), the association between FCs and UB was statistically stronger for patients with low-income than for

patients with moderate income ( $\beta=.42$  vs  $\beta=.23$ ,  $P=.04$ ) and higher income ( $\beta=.42$  vs  $\beta=.07$ ,  $P=.03$ ). The path between FCs

and UB was statistically stronger for patients with moderate income and those with high income ( $\beta=.23$  vs  $\beta=.07$ ,  $P=.003$ ). The relationship between BI and UB was statistically stronger for patients with high income than for those with low income ( $\beta=.61$  vs  $\beta=.43$ ,  $P=.008$ ) and middle income ( $\beta=.61$  vs  $\beta=.41$ ,  $P=.03$ ).

**Table 9.** Results of the moderating effect of income (low income vs middle income).

Hypothesis #	Hypothesized path	Low income <sup>a</sup>		Middle income <sup>b</sup>		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H4	PPS <sup>c</sup> →PE <sup>d</sup>	.38	<.001	.40	<.001	.71
H5	PPS→BI <sup>e</sup>	.24	<.001	.27	<.001	.91
H10	EE <sup>f</sup> →PE	.18	<.001	.32	<.001	.07
H11	EE→BI	.14	<.001	.21	<.001	.41
H15	SI <sup>g</sup> →PE	.14	.006	.13	.07	.98
H21	FC <sup>h</sup> →PE	.22	<.001	.06	.40	.43
H22	FCs→BI	.08	.09	.06	.31	.96
H23	FCs→UB <sup>i</sup>	.42	<.001	.23	<.001	.04
H27	PE→BI	.53	<.001	.52	<.001	.40
H29	BI→UB	.43	<.001	.41	<.001	.87

<sup>a</sup>Low income: <US \$25,000.

<sup>b</sup>Medium income: US \$25,000-US \$50,999.

<sup>c</sup>PPS: perceived privacy and security.

<sup>d</sup>PE: performance expectancy.

<sup>e</sup>BI: behavioral intention.

<sup>f</sup>EE: effort expectancy.

<sup>g</sup>SI: social influence.

<sup>h</sup>FC: facilitating condition.

<sup>i</sup>UB: use behavior.



**Table 10.** Results of the moderating effect of income (low income vs high income).

Hypothesis #	Hypothesized path	Low income <sup>a</sup>		High income <sup>b</sup>		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H4	PPS <sup>c</sup> →PE <sup>d</sup>	.38	<.001	.39	<.001	.92
H5	PPS→BI <sup>e</sup>	.24	<.001	.23	<.001	.81
H10	EE <sup>f</sup> →PE	.18	<.001	.24	.01	.60
H11	EE→BI	.14	<.001	.09	.13	.53
H15	SI <sup>g</sup> →PE	.14	.006	.19	.054	.45
H21	FCs <sup>h</sup> →PE	.22	<.001	-.11	.29	.06
H22	FCs→BI	.08	.09	.03	.63	.96
H23	FCs→UB <sup>i</sup>	.42	<.001	.07	.40	.03
H27	PE→BI	.53	<.001	.66	<.001	.12
H29	BI→UB	.43	<.001	.61	<.001	.008

<sup>a</sup>Low income: <US \$25,000.<sup>b</sup>High income: ≥US \$51,000.<sup>c</sup>PPS: perceived privacy and security.<sup>d</sup>PE: performance expectancy.<sup>e</sup>BI: behavioral intention.<sup>f</sup>EE: effort expectancy.<sup>g</sup>SI: social influence.<sup>h</sup>FC: facilitating condition.<sup>i</sup>UB: use behavior.

**Table 11.** Results of the moderating effect of income (middle income vs high income).

Hypothesis #	Hypothesized path	Middle income <sup>a</sup>		High income <sup>b</sup>		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H4	PPS <sup>c</sup> →PE <sup>d</sup>	.40	<.001	.39	<.001	.83
H5	PPS→BI <sup>e</sup>	.27	<.001	.23	<.001	.75
H10	EE <sup>f</sup> →PE	.32	<.001	.24	.01	.22
H11	EE→BI	.21	<.001	.09	.13	.21
H15	SI <sup>g</sup> →PE	.13	.07	.19	.054	.51
H21	FCs <sup>h</sup> →PE	.06	.40	-.11	.29	.18
H22	FCs→BI	.06	.31	.03	.63	.96
H23	FCs→UB <sup>i</sup>	.23	<.001	.07	.40	.003
H27	PE→BI	.52	<.001	.66	<.001	.07
H29	BI→UB	.41	<.001	.61	<.001	.03

<sup>a</sup>Medium income: US \$25,000-US \$50,999.

<sup>b</sup>High income: ≥US \$51,000.

<sup>c</sup>PPS: perceived privacy and security.

<sup>d</sup>PE: performance expectancy.

<sup>e</sup>BI: behavioral intention.

<sup>f</sup>EE: effort expectancy.

<sup>g</sup>SI: social influence.

<sup>h</sup>FC: facilitating condition.

<sup>i</sup>UB: use behavior.

With respect to the moderating effect of internet access (Table 12), the paths EE→BI and FCs→UB were statistically stronger for patients without internet access than for those with internet access ( $P=.03$ ,  $P<.001$ , respectively). In contrast, the paths PE→BI and BI→UB were statistically stronger for patients

with internet access than for those without internet access ( $P=.005$  and  $P=.002$ , respectively). According to the results of all moderating effects, the following hypotheses were partially supported: H10, H11, H23, H27, and H29.

**Table 12.** Results of the moderating effect of internet access.

Hypothesis #	Hypothesized path	Internet access		No internet access		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H4	PPS <sup>a</sup> →PE <sup>b</sup>	.40	<.001	.35	<.001	.68
H5	PPS→BI <sup>c</sup>	.21	<.001	.30	.002	.93
H10	EE <sup>d</sup> →PE	.25	<.001	.24	.02	.36
H11	EE→BI	.10	<.001	.27	.007	.03
H15	SI <sup>e</sup> →PE	.14	<.001	.07	.51	.26
H21	FCs <sup>f</sup> →PE	.12	.006	.22	.06	.89
H22	FCs→BI	.04	.17	.07	.46	.89
H23	FCs→UB <sup>g</sup>	.20	.03	.34	<.001	<.001
H27	PE→BI	.59	<.001	.39	<.001	.005
H29	BI→UB	.51	<.001	.31	.004	.002

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FCs: facilitating conditions.<sup>g</sup>UB: use behavior.

### Moderated Mediating Effects

With regard to the proposed moderated mediations, the indirect effects of EE and SI on BI were statistically stronger for women than for men ( $P=.03$  and  $P=.01$ , respectively; Table 13). The indirect effect of PPS on BI was stronger for patients with college or diploma compared with those with secondary school and lower (Tables 14-16). In contrast, the indirect effect of EE on BI was stronger for patients with secondary school or lower

than for those with college or diploma (Tables 14-16). There was no moderating effect of income on all indirect effects (Tables 16-19). As shown in Table 20, the indirect effect of PPS on BI is statistically stronger for patients with internet access ( $P<.001$ ). The indirect effect of EE on BI was statistically stronger for patients without internet access ( $P=.03$ ). Accordingly, the following hypotheses were partially supported: H6, H12, and H16.

**Table 13.** Results of the moderating effect of sex on indirect paths.

Hypothesis #	Hypothesized path	Men		Women		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H6	PPS <sup>a</sup> →PE <sup>b</sup> →BI <sup>c</sup>	.21	<.001	.19	<.001	.24
H12	EE <sup>d</sup> →PE→BI	.11	<.001	.16	<.001	.03
H16	SI <sup>e</sup> →PE→BI	.03	.18	.10	<.001	.01
H24	FCs <sup>f</sup> →PE→BI	.04	.23	.10	.004	.06

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.

**Table 14.** Results of the moderating effect of education on indirect paths (school vs college).

Hypothesis #	Hypothesized path	Secondary school or lower		College or diploma		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H6	PPS <sup>a</sup> →PE <sup>b</sup> →BI <sup>c</sup>	.12	<.001	.30	.002	.007
H12	EE <sup>d</sup> →PE→BI	.17	<.001	.01	.90	.045
H16	SI <sup>e</sup> →PE→BI	.06	.11	.08	.02	.45
H24	FCs <sup>f</sup> →PE→BI	.11	.007	.04	.31	.49

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.**Table 15.** Results of the moderating effect of education on indirect paths (school vs bachelor).

Hypothesis #	Hypothesized path	Secondary school or lower		Bachelor or higher		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H6	PPS <sup>a</sup> →PE <sup>b</sup> →BI <sup>c</sup>	.12	<.001	.11	.04	.75
H12	EE <sup>d</sup> →PE→BI	.17	<.001	.08	.09	.81
H16	SI <sup>e</sup> →PE→BI	.06	.11	.10	.03	.27
H24	FCs <sup>f</sup> →PE→BI	.11	.007	.10	.05	.26

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.**Table 16.** Results of the moderating effect of education on indirect paths (college vs bachelor).

Hypothesis #	Hypothesized path	College or diploma		Bachelor or higher		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H6	PPS <sup>a</sup> →PE <sup>b</sup> →BI <sup>c</sup>	.30	.002	.11	.04	.16
H12	EE <sup>d</sup> →PE→BI	.01	.90	.08	.09	.16
H16	SI <sup>e</sup> →PE→BI	.08	.02	.10	.03	.59
H24	FCs <sup>f</sup> →PE→BI	.04	.31	.10	.05	.14

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.

**Table 17.** Results of the moderating effect of income on indirect paths (low income vs middle income).

Hypothesis #	Hypothesized path	Low income		Middle income		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H6	PPS <sup>a</sup> →PE <sup>b</sup> →BI <sup>c</sup>	.20	<.001	.21	<.001	.84
H12	EE <sup>d</sup> →PE→BI	.10	<.001	.19	.002	.13
H16	SI <sup>e</sup> →PE→BI	.07	.01	.07	.09	.90
H24	FCs <sup>f</sup> →PE→BI	.12	.002	.03	.37	.32

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.**Table 18.** Results of the moderating effect of income on indirect paths (low income vs high income).

Hypothesis #	Hypothesized path	Low income		High income		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H6	PPS <sup>a</sup> →PE <sup>b</sup> →BI <sup>c</sup>	.20	<.001	.27	.002	.56
H12	EE <sup>d</sup> →PE→BI	.10	<.001	.16	.03	.37
H16	SI <sup>e</sup> →PE→BI	.07	.01	.13	.04	.28
H24	FCs <sup>f</sup> →PE→BI	.12	.002	-.07	.22	.06

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.**Table 19.** Results of the moderating effect of income on indirect paths (middle income vs high income).

Hypothesis #	Hypothesized path	Middle income		High income		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H6	PPS <sup>a</sup> →PE <sup>b</sup> →BI <sup>c</sup>	.21	<.001	.27	.002	.45
H12	EE <sup>d</sup> →PE→BI	.19	.002	.16	.03	.84
H16	SI <sup>e</sup> →PE→BI	.07	.09	.13	.04	.30
H24	FCs <sup>f</sup> →PE→BI	.03	.37	-.07	.22	.17

<sup>a</sup>PPS: perceived privacy and security.<sup>b</sup>PE: performance expectancy.<sup>c</sup>BI: behavioral intention.<sup>d</sup>EE: effort expectancy.<sup>e</sup>SI: social influence.<sup>f</sup>FC: facilitating condition.



**Table 20.** Results of the moderating effect of internet access on indirect paths.

Hypothesis #	Hypothesized path	Internet access		No internet access		<i>P</i> value for chi-square difference test
		Standardized estimate ( $\beta$ )	<i>P</i> value	Standardized estimate ( $\beta$ )	<i>P</i> value	
H6	PPS <sup>a</sup> →PE <sup>b</sup> →BI <sup>c</sup>	.25	<.001	.02	.72	.001
H12	EE <sup>d</sup> →PE→BI	.10	.02	.15	<.001	.03
H16	SI <sup>e</sup> →PE→BI	.08	<.001	.03	.42	.06
H24	FCs <sup>f</sup> →PE→BI	.07	.004	.09	.04	.54

<sup>a</sup>PPS: perceived privacy and security.

<sup>b</sup>PE: performance expectancy.

<sup>c</sup>BI: behavioral intention.

<sup>d</sup>EE: effort expectancy.

<sup>e</sup>SI: social influence.

<sup>f</sup>FC: facilitating condition.

## Discussion

### Principal Findings

This study aimed to improve the predictive power of a model proposed by Abd-Alrazaq et al [38] by proposing and examining new relationships between the variables existing in that model. The predictive power of the new model was slightly higher than that of the Abd-Alrazaq model for PE (53% vs 51%) and UB (49% vs 48%), but it was exactly the same in both models for BI (76%).

With regard to the direct effects, there was no considerable difference between the new model and the Abd-Alrazaq model for the following paths PE→BI (0.57 vs 0.57), EE→BI (0.15 vs 0.16), PPS→BI (0.23 vs 0.24), FCs→UB (0.25 vs 0.25), and BI→UB (0.53 vs 0.53). Compared with the Abd-Alrazaq model, the current model showed a considerable decrease in the effect of EE (0.25 vs 0.34) and PPS (0.39 vs 0.49) on PE; however, both paths were still significant in the current model. This decrease resulted from proposing 2 new predictors for PE (ie, SI and FCs) in the current model, which were significant. The only direct path that was nonsignificant was FCs→BI in the current model. This finding is in line with the findings of a study conducted by Tavares and Oliveira, who did not find a significant association between FCs and BI to use ePHRs [40]. Venkatsh et al [39] attributed this nonsignificant path to the fact that this effect disappears when a model includes both PE and EE.

Compared with the Abd-Alrazaq model, the current model showed a decrease in the indirect associations between BI and each of EE (0.15 vs 0.20) and PPS (0.22 vs 0.28) through PE. However, both indirect effects are still significant in the current model. This decrease resulted from proposing 2 new predictors for PE (ie, SI and FCs) in the current model. Two new indirect paths were found SI→PE→BI and FCs→PE→BI. This means that patients who perceive that important others believe they should use GP online services are more likely to perceive it as a useful system; therefore, they are more likely to intend to use it. Furthermore, patients who believe that an organizational and technical infrastructure exists to support the use of GP online

services are more likely to perceive it as a useful system and are therefore more likely to intend to use it.

All proposed moderating effects that are common between the new model and the Abd-Alrazaq model were comparable between both models. In addition to the significant moderating effects found in the Abd-Alrazaq model, this study found that the association between BI and UB is significantly moderated by age, sex, education, income, and internet access and that the association between EE and PE is moderated by education. Specifically, the association between BI and UB is stronger for younger women with higher levels of education, income, and internet access, and the association between EE and PE is stronger for patients with lower levels of education.

With regard to the moderated mediations proposed in the new model, this study found that the indirect effect of EE on BI through PE was statistically stronger for women without internet access. The indirect effect of PPS on BI was stronger for patients with college education or diploma compared with those with secondary school education and lower, whereas the indirect effect of EE on BI was stronger for patients with secondary school or lower than for those with college education or diplomas. Furthermore, the indirect effect of SI on BI through PE was stronger for patients without internet access. Last but not least, the indirect effect of PPS on BI through PE was statistically stronger for patients with internet access.

### Theoretical and Practical Contributions

This study is one of the very few theory-based studies conducted to identify the factors that affect patients' use of ePHRs or patient portals [21,25-27]. The predictive power of the new model (49%) is higher than that of the previous models proposed in our previous study (48%) and other studies conducted in the context of ePHRs: Tavares and Oliveira (26.8%) [40] and Hsieh (42.7%) [85]. Moreover, the predictive power of the new model is higher than that of the original UTAUT model (48%) [39]. Accordingly, this study contributes to the literature by providing the most predictive model to explain the adoption of ePHRs to date.

To the best of our knowledge, this is the first study in the context of ePHRs that examined the direct effect of SI and FCs on PE, their indirect effects on BI through PE, the moderation effects on the association between BI and UB, and the moderated effects on indirect relationships. This extends our understanding of the complex associations between the factors that affect the adoption of ePHRs.

In addition to the practical contributions reported in the previous study [38], this study provides some contributions based on the newly proposed relationships. People who are important to patients (eg, family members, friends, physicians, and caregivers) can play an important role in enhancing their perceived usefulness of the system and their intention to use it. The influence of these important people is more evident on women than on men. Interventions aimed at increasing the uptake of online access could harness the influence of such individuals to encourage patients to use such services. For example, GPs could prompt the recurrent users of the system to become ePHR champions and speak to their friends or family members about their experiences. GPs could also train practice staff to offer these services to their patients routinely in their communications, and campaigns aimed at increasing ePHR uptake could use social influencing techniques, such as celebrity endorsements. As FCs are directly associated with perceived usefulness of ePHRs, and this, in turn affects BIs, steps should be taken to improve the degree to which patients believe that an organizational and technical infrastructure exists to support their use. For example, the National Health Service app has demonstrated an efficient infrastructure supporting the patient registration process, which enables patients to sign up to access their records online without needing to visit their GP surgery [15]. Instead of registration requiring patients to show evidence of their identity to practice staff, they can instead register by uploading a photograph of identifying documentation and taking a short selfie video on their mobile device. Other potential approaches to targeting FCs include the provision of online educational materials, 24/7 technical support, or drop-in training sessions at GP practices.

### Research Limitations

The proposed model was examined using data collected from 4 GP practices that have implemented the same system (SystemOnline); therefore, our findings may not be generalizable to other systems (eg, Patient Access and i-Patient). Nonetheless, the findings may still be applicable to other ePHRs because all participants were nonusers, and these systems offer the same services to the patients. Consequently, participants would be unlikely to have different perceptions of the different systems.

This study focused on the factors that influence the initial use of ePHRs, given that the system is new in England and has a low adoption rate. Thus, the generalizability of the findings in the context of the continuing use of ePHRs is limited. Given that the study used secondary data, it was not possible to assess the effect of new factors, such as those recommended by Abd-Alrazaq et al [38].

Sampling bias may be a concern in this study owing to the convenience sampling technique used to recruit the participants [37,86]. This study showed that there was no statistically significant difference between the participants and nonparticipants in terms of age, sex, and ethnicity. Accordingly, our findings may be generalizable to GPs, similar to the 4 GPs in this study.

### Recommendations for Future Research

The applicability of the proposed model to other contexts should be examined in further studies. Specifically, researchers may assess the applicability of the model to other providers of GP online services (eg, Patient Access), specific platforms (eg, mobiles, tablets, and computers), other settings (eg, hospitals), and other cities or countries.

Further studies are required to validate the new significant associations proposed in this study, such as SI→PE, FCs→PE, and SI→PE→BI. In addition, future studies should endeavor to improve the predictive power of the current model by adding other factors such as awareness of the system, health status, perceived severity, patient satisfaction, and patient activation level.

It is well known that the eventual success of information technology depends on continued use more than initial use [33,87-89]. There is a lack of studies that have assessed factors affecting the continuing use of ePHRs or even consumer health information technologies (CHITs). Therefore, we prompt researchers to develop and examine a theoretical model that explains the variables affecting the continuing use of ePHRs and CHITs.

This study did not assess series mediations, such as the indirect effect of EE on UB through PE and BI (ie, EE→PE→BI→UB). Furthermore, to the best of our knowledge, such effects have not been examined in previous studies in the context of ePHRs and CHITs. This highlights a need to assess such effects.

### Conclusions

This study slightly improved the predictive power of the Abd-Alrazaq model. More importantly, the improved model showed new significant relationships that were not examined before in the ePHR context, such as the direct effect of SI and FCs on PE, their indirect effects on BI through PE, moderation effect of age, sex, educational level, income, and internet access on the association between BI and UB, and the moderating effects on some indirect relationships. These findings extend our understanding of the complex associations between factors affecting the adoption of ePHRs. The predictive power of 49% indicates that there are other, as yet unidentified, factors that affect the use of ePHRs. Further studies are required to validate the new model in different contexts and to improve its predictive power by proposing new factors. Interventions could focus on the role of significant others (eg, health care professionals, friends, and family members) in influencing web access usage, for example, by discussing the potential benefits of such services with patients.

## Acknowledgments

The authors would like to thank Dr Peter Gardner and Dr Tracey Farragher for their supervision during the study and Dr Hamish Fraser for his help in the initial stages of the study, especially for his contribution to the development of the study design and setting up links with practices.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Selection of theory.

[\[DOCX File , 70 KB - jmir\\_v22i10e17499\\_app1.docx \]](#)

### Multimedia Appendix 2

The previous model.

[\[PNG File , 195 KB - jmir\\_v22i10e17499\\_app2.png \]](#)

### Multimedia Appendix 3

Justification of adaptation of UTAUT.

[\[DOCX File , 73 KB - jmir\\_v22i10e17499\\_app3.docx \]](#)

### Multimedia Appendix 4

Conceptual definitions of constructs.

[\[DOCX File , 15 KB - jmir\\_v22i10e17499\\_app4.docx \]](#)

### Multimedia Appendix 5

Characteristics of the four general practices.

[\[DOCX File , 17 KB - jmir\\_v22i10e17499\\_app5.docx \]](#)

### Multimedia Appendix 6

Measures of constructs.

[\[DOCX File , 18 KB - jmir\\_v22i10e17499\\_app6.docx \]](#)

### Multimedia Appendix 7

Questionnaire.

[\[DOCX File , 70 KB - jmir\\_v22i10e17499\\_app7.docx \]](#)

### Multimedia Appendix 8

Participants' characteristics.

[\[DOCX File , 20 KB - jmir\\_v22i10e17499\\_app8.docx \]](#)

### Multimedia Appendix 9

Histograms.

[\[DOCX File , 711 KB - jmir\\_v22i10e17499\\_app9.docx \]](#)

### Multimedia Appendix 10

Values of skewness and kurtosis.

[\[DOCX File , 14 KB - jmir\\_v22i10e17499\\_app10.docx \]](#)

### Multimedia Appendix 11

Scatterplot graphs.

[\[DOCX File , 590 KB - jmir\\_v22i10e17499\\_app11.docx \]](#)

### Multimedia Appendix 12

Curve estimation procedure.

[\[DOCX File , 15 KB - jmir\\_v22i10e17499\\_app12.docx \]](#)

## Multimedia Appendix 13

Tolerance values.

[\[DOCX File, 13 KB - jmir\\_v22i10e17499\\_app13.docx\]](#)

## Multimedia Appendix 14

Harman's single-factor test.

[\[DOCX File, 14 KB - jmir\\_v22i10e17499\\_app14.docx\]](#)

## Multimedia Appendix 15

Results of fit indices of the measurement model.

[\[DOCX File, 14 KB - jmir\\_v22i10e17499\\_app15.docx\]](#)

## Multimedia Appendix 16

Results of construct reliability.

[\[DOCX File, 13 KB - jmir\\_v22i10e17499\\_app16.docx\]](#)

## Multimedia Appendix 17

Results of convergent validity.

[\[DOCX File, 14 KB - jmir\\_v22i10e17499\\_app17.docx\]](#)

## Multimedia Appendix 18

Intercorrelation coefficients and squared roots of AVE.

[\[DOCX File, 15 KB - jmir\\_v22i10e17499\\_app18.docx\]](#)

## Multimedia Appendix 19

Item loadings and cross-loadings.

[\[DOCX File, 21 KB - jmir\\_v22i10e17499\\_app19.docx\]](#)

## Multimedia Appendix 20

Results of fit indices of the structural model.

[\[DOCX File, 15 KB - jmir\\_v22i10e17499\\_app20.docx\]](#)**References**

1. The Personal Health Working Group: Final Report. Markle: Advancing America's Future. 2003. URL: [https://www.markle.org/sites/default/files/final\\_phwg\\_report1.pdf](https://www.markle.org/sites/default/files/final_phwg_report1.pdf) [accessed 2020-09-04]
2. Pai HH, Lau F, Barnett J, Jones S. Meeting the health information needs of prostate cancer patients using personal health records. *Curr Oncol* 2013 Dec;20(6):e561-e569 [[FREE Full text](#)] [doi: [10.3747/co.20.1584](https://doi.org/10.3747/co.20.1584)] [Medline: [24311957](https://pubmed.ncbi.nlm.nih.gov/24311957/)]
3. Tulu B, Trapp AC, Strong DM, Johnson SA, Hoque M, Trudel J, et al. An analysis of patient portal utilization: what can we learn about online patient behavior by examining portal click data? *Health Syst* 2017 Dec 19;5(1):66-79. [doi: [10.1057/hs.2015.5](https://doi.org/10.1057/hs.2015.5)]
4. McMillan B, Eastham R, Brown B, Fitton R, Dickinson D. Primary care patient records in the United Kingdom: past, present, and future research priorities. *J Med Internet Res* 2018 Dec 19;20(12):e11293 [[FREE Full text](#)] [doi: [10.2196/11293](https://doi.org/10.2196/11293)] [Medline: [30567695](https://pubmed.ncbi.nlm.nih.gov/30567695/)]
5. Laranjo L, Rodolfo I, Pereira AM, de Sá AB. Characteristics of innovators adopting a national personal health record in Portugal: cross-sectional study. *JMIR Med Inform* 2017 Oct 11;5(4):e37 [[FREE Full text](#)] [doi: [10.2196/medinform.7887](https://doi.org/10.2196/medinform.7887)] [Medline: [29021125](https://pubmed.ncbi.nlm.nih.gov/29021125/)]
6. Ose D, Kunz A, Pohlmann S, Hofmann H, Qreini M, Krisam J, et al. A personal electronic health record: study protocol of a feasibility study on implementation in a real-world health care setting. *JMIR Res Protoc* 2017 Mar 2;6(3):e33 [[FREE Full text](#)] [doi: [10.2196/resprot.6314](https://doi.org/10.2196/resprot.6314)] [Medline: [28254735](https://pubmed.ncbi.nlm.nih.gov/28254735/)]
7. Rice R. Healthcare leaders' lived experiences regarding the implementation of electronic personal health records. Walden University. 2014. URL: [https://www.researchgate.net/publication/340135954\\_Patients'\\_Adoption\\_of\\_Electronic\\_Personal\\_Health\\_Records\\_in\\_England\\_Secondary\\_data\\_analysis](https://www.researchgate.net/publication/340135954_Patients'_Adoption_of_Electronic_Personal_Health_Records_in_England_Secondary_data_analysis) [accessed 2020-09-11]
8. Vermeir P, Degroote S, Vandijck D, van Tiggelen H, Peleman R, Verhaeghe R, et al. The patient perspective on the effects of medical record accessibility: a systematic review. *Acta Clin Belg* 2017 Jun;72(3):186-194. [doi: [10.1080/17843286.2016.1275375](https://doi.org/10.1080/17843286.2016.1275375)] [Medline: [28056665](https://pubmed.ncbi.nlm.nih.gov/28056665/)]

9. Alyami M, Song Y. Removing Barriers in Using Personal Health Record Systems. In: 15th International Conference on Computer and Information Science. 2016 Presented at: ICIS'16; June 26-29, 2016; Okayama, Japan. [doi: [10.1109/icis.2016.7550810](https://doi.org/10.1109/icis.2016.7550810)]
10. Ochoa IA, Kitayama K, Uijtdehaage S, Vermillion M, Eaton M, Carpio F, et al. Patient and provider perspectives on the potential value and use of a bilingual online patient portal in a Spanish-speaking safety-net population. *J Am Med Inform Assoc* 2017 Nov 1;24(6):1160-1164. [doi: [10.1093/jamia/ocx040](https://doi.org/10.1093/jamia/ocx040)] [Medline: [28460130](https://pubmed.ncbi.nlm.nih.gov/28460130/)]
11. Morton A. Examining Acceptance of an Integrated Personal Health Record (PHR). University of Maryland. 2012. URL: <https://dl.acm.org/doi/book/10.5555/2395584> [accessed 2020-09-11]
12. Pagliari C, Detmer D. Electronic Personal Health Records: Emergence and Implications for the UK. Nuffield Trust. 2007. URL: <https://www.nuffieldtrust.org.uk/files/2017-01/electronic-personal-health-records-uk-web-final.pdf> [accessed 2020-09-11]
13. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. *Br Med J* 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: [10.1136/bmj.39279.482963.AD](https://doi.org/10.1136/bmj.39279.482963.AD)] [Medline: [17703042](https://pubmed.ncbi.nlm.nih.gov/17703042/)]
14. Online Health and Prescription Services. United Kingdom National Health Service. 2016. URL: <https://www.nhs.uk/using-the-nhs/nhs-services/gps/gp-online-services/> [accessed 2020-09-04]
15. GP Online Services. United Kingdom National Health Service. 2016. URL: <https://www.nhs.uk/using-the-nhs/about-the-nhs/how-to-access-your-health-records/> [accessed 2020-09-04]
16. Patient Online Management Information (POMI). NHS Digital. 2018. URL: <https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/pomi> [accessed 2020-09-04]
17. Fung V, Ortiz E, Huang J, Fireman B, Miller R, Selby JV, et al. Early experiences with e-health services (1999-2002): promise, reality, and implications. *Med Care* 2006 May;44(5):491-496. [doi: [10.1097/01.mlr.0000207917.28988.dd](https://doi.org/10.1097/01.mlr.0000207917.28988.dd)] [Medline: [16641669](https://pubmed.ncbi.nlm.nih.gov/16641669/)]
18. Huygens MW, Vermeulen J, Friele RD, van Schayck OC, de Jong JD, de Witte LP. Internet services for communicating with the general practice: barely noticed and used by patients. *Interact J Med Res* 2015 Nov 24;4(4):e21 [FREE Full text] [doi: [10.2196/ijmr.4245](https://doi.org/10.2196/ijmr.4245)] [Medline: [26601596](https://pubmed.ncbi.nlm.nih.gov/26601596/)]
19. Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW. A research agenda for personal health records (PHRs). *J Am Med Inform Assoc* 2008;15(6):729-736 [FREE Full text] [doi: [10.1197/jamia.M2547](https://doi.org/10.1197/jamia.M2547)] [Medline: [18756002](https://pubmed.ncbi.nlm.nih.gov/18756002/)]
20. Logue MD, Effken JA. Modeling factors that influence personal health records adoption. *Comput Inform Nurs* 2012 Jul;30(7):354-362. [doi: [10.1097/NXN.0b013e3182510717](https://doi.org/10.1097/NXN.0b013e3182510717)] [Medline: [22525046](https://pubmed.ncbi.nlm.nih.gov/22525046/)]
21. Najaftorkaman M, Ghapanchi A, Talaei-Khoei A. Analysis of Research in Adoption of Person-Centred Healthcare Systems: the Case of Online Personal Health Record. In: 25th Australasian Conference on Information Systems. 2014 Presented at: CIS'14; December 8-10, 2014; Auckland, New Zealand.
22. Or CK, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. *J Am Med Inform Assoc* 2011;18(1):51-59 [FREE Full text] [doi: [10.1136/jamia.2010.007336](https://doi.org/10.1136/jamia.2010.007336)] [Medline: [21131605](https://pubmed.ncbi.nlm.nih.gov/21131605/)]
23. Sanders MR, Winters P, Fortuna RJ, Mendoza M, Berliant M, Clark L, et al. Internet access and patient portal readiness among patients in a group of inner-city safety-net practices. *J Ambul Care Manage* 2013;36(3):251-259. [doi: [10.1097/JAC.0b013e31829702f9](https://doi.org/10.1097/JAC.0b013e31829702f9)] [Medline: [23748274](https://pubmed.ncbi.nlm.nih.gov/23748274/)]
24. Abd-Alrazaq AA, Bewick B, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: a systematic review. *Int J Med Inform* 2019 Jun;126:164-175. [doi: [10.1016/j.ijmedinf.2019.03.014](https://doi.org/10.1016/j.ijmedinf.2019.03.014)] [Medline: [31029258](https://pubmed.ncbi.nlm.nih.gov/31029258/)]
25. Andrews L, Gajanayake R, Sahama T. The Australian general public's perceptions of having a personally controlled electronic health record (PCEHR). *Int J Med Inform* 2014 Dec;83(12):889-900. [doi: [10.1016/j.ijmedinf.2014.08.002](https://doi.org/10.1016/j.ijmedinf.2014.08.002)] [Medline: [25200198](https://pubmed.ncbi.nlm.nih.gov/25200198/)]
26. Assadi V. Adoption of Integrated Personal Health Record Systems: a Self-determination Theory Perspective. McMaster University. 2013. URL: [https://www.researchgate.net/publication/318739488\\_Consumer\\_Adoption\\_of\\_Personal\\_Health\\_Record\\_Systems\\_A\\_Self-Determination\\_Theory\\_Perspective](https://www.researchgate.net/publication/318739488_Consumer_Adoption_of_Personal_Health_Record_Systems_A_Self-Determination_Theory_Perspective) [accessed 2020-09-11]
27. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. *J Med Internet Res* 2012 Nov 5;14(6):e150 [FREE Full text] [doi: [10.2196/jmir.2278](https://doi.org/10.2196/jmir.2278)] [Medline: [23128775](https://pubmed.ncbi.nlm.nih.gov/23128775/)]
28. Or CK, Karsh B. A systematic review of patient acceptance of consumer health information technology. *J Am Med Inform Assoc* 2009;16(4):550-560 [FREE Full text] [doi: [10.1197/jamia.M2888](https://doi.org/10.1197/jamia.M2888)] [Medline: [19390112](https://pubmed.ncbi.nlm.nih.gov/19390112/)]
29. Abramson E, Patel V, Edwards A, Kaushal R. Consumer perspectives on personal health records: a 4-community study. *Am J Manag Care* 2014 Apr;20(4):287-296 [FREE Full text] [Medline: [24884860](https://pubmed.ncbi.nlm.nih.gov/24884860/)]
30. Nambisan P. Factors that impact patient web portal readiness (PWPR) among the underserved. *Int J Med Inform* 2017 Jun;102:62-70. [doi: [10.1016/j.ijmedinf.2017.03.004](https://doi.org/10.1016/j.ijmedinf.2017.03.004)] [Medline: [28495349](https://pubmed.ncbi.nlm.nih.gov/28495349/)]
31. Ozok AA, Wu H, Gurses AP. Exploring patients' use intention of personal health record systems: implications for design. *Int J Hum-Comput Int* 2017 Jan 6;33(4):265-279. [doi: [10.1080/10447318.2016.1277637](https://doi.org/10.1080/10447318.2016.1277637)]



32. Razmak J, Bélanger C. Using the technology acceptance model to predict patient attitude toward personal health records in regional communities. *Info Technol People* 2018 Apr 3;31(2):306-326. [doi: [10.1108/ITP-07-2016-0160](https://doi.org/10.1108/ITP-07-2016-0160)]
33. Gebauer L, Söllner M, Leimeister J. Towards Understanding the Formation of Continuous It Use. In: 34th International Conference on Information Systems. 2013 Presented at: CIS'13; December 15-18, 2013; Milan, Italy.
34. Legris P, Ingham J, Colletette P. Why do people use information technology? A critical review of the technology acceptance model. *Info Manag* 2003 Jan;40(3):191-204. [doi: [10.1016/S0378-7206\(01\)00143-4](https://doi.org/10.1016/S0378-7206(01)00143-4)]
35. Rodman D. Meaningful Use: the Utilization of Patient Portals. D'Youville College. 2015. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4141030/> [accessed 2020-09-11]
36. Turner M, Kitchenham B, Brereton P, Charters S, Budgen D. Does the technology acceptance model predict actual use? A systematic literature review. *Inform Software Tech* 2010 May;52(5):463-479. [doi: [10.1016/j.infsof.2009.11.005](https://doi.org/10.1016/j.infsof.2009.11.005)]
37. Bhattacharjee A. Social Science Research: Principles, Methods, and Practices. Creative Commons Attribution. 2012. URL: <http://scholarcommons.usf.edu/> [accessed 2020-09-11]
38. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors affecting patients' use of electronic personal health records in England: cross-sectional study. *J Med Internet Res* 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: [10.2196/12373](https://doi.org/10.2196/12373)] [Medline: [31368442](https://pubmed.ncbi.nlm.nih.gov/31368442/)]
39. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. *MIS Q* 2003;27(3):425. [doi: [10.2307/30036540](https://doi.org/10.2307/30036540)]
40. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. *J Med Internet Res* 2016 Mar 2;18(3):e49 [FREE Full text] [doi: [10.2196/jmir.5069](https://doi.org/10.2196/jmir.5069)] [Medline: [26935646](https://pubmed.ncbi.nlm.nih.gov/26935646/)]
41. Taylor S, Todd P. Understanding information technology usage: a test of competing models. *Inf Syst Res* 1995 Jun;6(2):144-176. [doi: [10.1287/isre.6.2.144](https://doi.org/10.1287/isre.6.2.144)]
42. Baird A. Extending Adoption of Innovation Theory with Consumer Influence the Case of Personal Health Records (PHRs) and Patient Portals. Arizona State University. URL: <https://eric.ed.gov/?id=ED549969> [accessed 2020-09-14]
43. Lazard AJ, Watkins I, Mackert MS, Xie B, Stephens KK, Shalev H. Design simplicity influences patient portal use: the role of aesthetic evaluations for technology acceptance. *J Am Med Inform Assoc* 2016 Apr;23(e1):e157-e161 [FREE Full text] [doi: [10.1093/jamia/ocv174](https://doi.org/10.1093/jamia/ocv174)] [Medline: [26635314](https://pubmed.ncbi.nlm.nih.gov/26635314/)]
44. Torres CA. Examining the role of anxiety and apathy in health consumers' intentions to use patient health portals for personal health information management. Florida State University. URL: <http://diginole.lib.fsu.edu/islandora/object/fsu%3A253944> [accessed 2020-09-14]
45. Greene K, Hale J, Rubin D. A test of the theory of reasoned action in the context of condom use and AIDS. *Commun Res* 1997 Jan;10(1):21-33. [doi: [10.1080/08934219709367656](https://doi.org/10.1080/08934219709367656)]
46. Park HS. Relationships among attitudes and subjective norms: testing the theory of reasoned action across cultures. *Commun Stud* 2009 May 22;51(2):- [FREE Full text] [doi: [10.1080/10510970009388516](https://doi.org/10.1080/10510970009388516)]
47. Shepherd G, O'keefe D. Separability of attitudinal and normative influences on behavioral intentions in the Fishbein-Ajzen model. *J Soc Psychol* 2010 Jul;122(2):287-288. [doi: [10.1080/00224545.1984.9713496](https://doi.org/10.1080/00224545.1984.9713496)]
48. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 1989 Sep;13(3):319. [doi: [10.2307/249008](https://doi.org/10.2307/249008)]
49. Hale JL, Householder BJ, Greene KL. The theory of reasoned action. In: *The persuasion Handbook: Developments in Theory and Practice*. Thousand Oaks: Sage Publications; 2002.
50. Bandura A. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs: Prentice Hall; 1986.
51. Chang M, Cheung W. Determinants of the intention to use Internet/WWW at work: a confirmatory study. *Inf Manag* 2001 Nov;39(1):1-14. [doi: [10.1016/S0378-7206\(01\)00075-1](https://doi.org/10.1016/S0378-7206(01)00075-1)]
52. Compeau D, Higgins C. Computer self-efficacy: development of a measure and initial test. *MIS Q* 1995 Jun;19(2):189. [doi: [10.2307/249688](https://doi.org/10.2307/249688)]
53. Taylor S, Todd P. Assessing IT usage: the role of prior experience. *MIS Q* 1995 Dec;19(4):561. [doi: [10.2307/249633](https://doi.org/10.2307/249633)]
54. Munro S, Lewin S, Swart T, Volmink J. A review of health behaviour theories: how useful are these for developing interventions to promote long-term medication adherence for TB and HIV/AIDS? *BMC Public Health* 2007 Jun 11;7:104 [FREE Full text] [doi: [10.1186/1471-2458-7-104](https://doi.org/10.1186/1471-2458-7-104)] [Medline: [17561997](https://pubmed.ncbi.nlm.nih.gov/17561997/)]
55. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. *Dec Sci* 2008 May;39(2):273-315. [doi: [10.1111/j.1540-5915.2008.00192.x](https://doi.org/10.1111/j.1540-5915.2008.00192.x)]
56. Assadi V. Adoption of integrated personal health record systems: a self-determination theory perspective. McMaster University. 2013. URL: <https://macsphere.mcmaster.ca/handle/11375/13530> [accessed 2020-09-15]
57. Bagozzi R. The legacy of the technology acceptance model and a proposal for a paradigm shift. *JAIS* 2007 Apr;8(4):244-254. [doi: [10.17705/1jais.00122](https://doi.org/10.17705/1jais.00122)]
58. Gartrell K. Factors Associated with Electronic Personal Health Record Use among Registered Nurses for Their Own Health Management. University of Maryland. URL: <https://archive.hshsl.umaryland.edu/handle/10713/4174> [accessed 2020-09-14]
59. Rastogi R, Chourey PS, Muhitch MJ. The maize glutamine synthetase GS1-2 gene is preferentially expressed in kernel pedicels and is developmentally-regulated. *Plant Cell Physiol* 1998 Apr;39(4):443-446. [doi: [10.1093/oxfordjournals.pcp.a029388](https://doi.org/10.1093/oxfordjournals.pcp.a029388)] [Medline: [9615467](https://pubmed.ncbi.nlm.nih.gov/9615467/)]

60. Whetstone M, Goldsmith R. Factors influencing intention to use personal health records. *Intl J of Pharm & Health Mkrt* 2009 Apr 03;3(1):8-25. [doi: [10.1108/17506120910948485](https://doi.org/10.1108/17506120910948485)]
61. Or CK. Development of a model of consumer health information technology acceptance of patients with chronic illness. University of Wisconsin-Madison. 2013. URL: [https://books.google.com.qa/books/about/Development\\_of\\_a\\_Model\\_of\\_Consumer\\_Health.html?id=F7gIAQAAMAAJ&redir\\_esc=y](https://books.google.com.qa/books/about/Development_of_a_Model_of_Consumer_Health.html?id=F7gIAQAAMAAJ&redir_esc=y) [accessed 2020-09-25]
62. Cimperman M, Makovec Brenčič M, Trkman P. Analyzing older users' home telehealth services acceptance behavior-applying an Extended UTAUT model. *Int J Med Inform* 2016 Jun;90:22-31. [doi: [10.1016/j.ijmedinf.2016.03.002](https://doi.org/10.1016/j.ijmedinf.2016.03.002)] [Medline: [27103194](https://pubmed.ncbi.nlm.nih.gov/27103194/)]
63. Beekens FH. Acceptance of e-health Technology: A Patient Perspective. Delft, Netherlands: Delft University of Technology; 2011.
64. de Veer AJ, Peeters JM, Brabers AE, Schellevis FG, Rademakers JJ, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. *BMC Health Serv Res* 2015 Mar 15;15:103 [FREE Full text] [doi: [10.1186/s12913-015-0765-8](https://doi.org/10.1186/s12913-015-0765-8)] [Medline: [25889884](https://pubmed.ncbi.nlm.nih.gov/25889884/)]
65. Mekawie NM. Factors affecting adoption of eHealth in Egypt. Middlesex University. 2013. URL: <https://eprints.mdx.ac.uk/12334/1/NMMekawie%20thesis.pdf> [accessed 2020-09-15]
66. Karahanna E, Straub D. The psychological origins of perceived usefulness and ease-of-use. *Info Manag* 1999 Apr;35(4):237-250 [FREE Full text] [doi: [10.1016/s0378-7206\(98\)00096-2](https://doi.org/10.1016/s0378-7206(98)00096-2)]
67. Kulviwat S, Bruner G, Al-Shuridah O. The role of social influence on adoption of high tech innovations: the moderating effect of public/private consumption. *J Bus Res* 2009 Jul;62(7):706-712 [FREE Full text] [doi: [10.1016/j.jbusres.2007.04.014](https://doi.org/10.1016/j.jbusres.2007.04.014)]
68. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. *Manag Sci* 2000 Feb;46(2):186-204. [doi: [10.1287/mnsc.46.2.186.11926](https://doi.org/10.1287/mnsc.46.2.186.11926)]
69. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Q* 2012;36(1):157. [doi: [10.2307/41410412](https://doi.org/10.2307/41410412)]
70. Bateson J, Hui M. Perceived control as a crucial perceptual dimension of the service experience: an experimental study. *J Consum Res* 1991;18(2).
71. Lee J. A key to marketing financial services: the right mix of products, services, channels and customers. *J Serv Mark* 2002 Jun;16(3):238-258. [doi: [10.1108/08876040210427227](https://doi.org/10.1108/08876040210427227)]
72. Koenig-Lewis N, Palmer A, Moll A. Predicting young consumers' take up of mobile banking services. *Int J Bank Mark* 2010 Jul 27;28(5):410-432. [doi: [10.1108/02652321011064917](https://doi.org/10.1108/02652321011064917)]
73. Hair J, Black W, Babin B, Anderson R. *Multivariate Data Analysis: With Readings*. Seventh Edition. Essex, UK: Pearson Education; 2010.
74. Snijders T, Bosker R. *A Beginner's Guide to Structural Equation Modeling*. Third Edition. New York, USA: Routledge; 1999.
75. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol* 2003 Oct;88(5):879-903. [doi: [10.1037/0021-9010.88.5.879](https://doi.org/10.1037/0021-9010.88.5.879)] [Medline: [14516251](https://pubmed.ncbi.nlm.nih.gov/14516251/)]
76. Podsakoff PM, MacKenzie SB, Podsakoff NP. Sources of method bias in social science research and recommendations on how to control it. *Annu Rev Psychol* 2012;63:539-569. [doi: [10.1146/annurev-psych-120710-100452](https://doi.org/10.1146/annurev-psych-120710-100452)] [Medline: [21838546](https://pubmed.ncbi.nlm.nih.gov/21838546/)]
77. Kline R. *Principles and Practice of Structural Equation Modeling*. Fourth Edition. New York, USA: Guilford Press; 2015.
78. Alanazi A. A Comprehensive Model to Explaining Users' Acceptance and Intention to Use Electronic Health Records (EHR) in Rehabilitation Facilities in Saudi Arabia. University of Pittsburgh. 2015. URL: <http://d-scholarship.pitt.edu/25740/> [accessed 2020-09-11]
79. Byrne B. *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming*. New York, USA: Routledge; 2016.
80. Kamarulzaman Y. The Adoption of Internet Shopping for Travel Services. Cardiff University. 2006. URL: <https://orca.cf.ac.uk/55606/1/U584036.pdf> [accessed 2020-09-11]
81. Chin W. The partial least squares approach to structural equation modeling. In: *Modern Methods for Business Research*. Mahwah, USA: Lawrence Erlbaum Associates; 1998:295-336.
82. Field A. *Discovering Statistics Using SPSS*. London, UK: Sage; 2017.
83. Menard S. *Applied logistic regression analysis*. In: *Applied Logistic Regression Analysis*. Thousand Oaks, CA: Sage; 1995.
84. Brown T. *Confirmatory Factor Analysis for Applied Research*. Second Edition. New York, USA: Guilford Publications; 2014.
85. Hsieh H, Kuo Y, Wang S, Chuang B, Tsai C. A study of personal health record user's behavioral model based on the PMT and UTAUT integrative perspective. *Int J Environ Res Public Health* 2016 Dec 23;14(1):- [FREE Full text] [doi: [10.3390/ijerph14010008](https://doi.org/10.3390/ijerph14010008)] [Medline: [28025557](https://pubmed.ncbi.nlm.nih.gov/28025557/)]
86. Zikmund W, Babin B, Carr J, Griffin M. *Business Research Methods*. Ninth Edition. New York, USA: Cengage Learning; 2013.
87. Bhattacharjee A. Understanding information systems continuance: an expectation-confirmation model. *MIS Q* 2001 Sep;25(3):351. [doi: [10.2307/3250921](https://doi.org/10.2307/3250921)]

88. Forquer HA, Christensen JL, Tan AS. Predicting continuance-findings from a longitudinal study of older adults using an eHealth newsletter. *Health Commun* 2014;29(9):937-946 [FREE Full text] [doi: [10.1080/10410236.2013.833580](https://doi.org/10.1080/10410236.2013.833580)] [Medline: [24446900](https://pubmed.ncbi.nlm.nih.gov/24446900/)]
89. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. *J Med Internet Res* 2011 Sep 30;13(3):e71 [FREE Full text] [doi: [10.2196/jmir.1603](https://doi.org/10.2196/jmir.1603)] [Medline: [21959968](https://pubmed.ncbi.nlm.nih.gov/21959968/)]

## Abbreviations

**AVE:** average variance extracted  
**BI:** behavioral intention  
**CHIT:** consumer health information technology  
**EE:** effort expectancy  
**ePHR:** electronic personal health record  
**FC:** facilitating condition  
**GP:** general practice  
**H:** hypothesis  
**PE:** performance expectancy  
**PPS:** perceived privacy and security  
**SEM:** structural equation modeling  
**SI:** social influence  
**UB:** use behavior  
**UTAUT:** Unified Theory of Acceptance and Use of Technology

*Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by A Tarhini, J Tavares, R Fitton; comments to author 05.02.20; revised version received 24.02.20; accepted 11.03.20; published 07.10.20.*

### *Please cite as:*

Abd-Alrazaq A, Alalwan AA, McMillan B, Bewick BM, Househ M, AL-Zyadat AT  
*Patients' Adoption of Electronic Personal Health Records in England: Secondary Data Analysis*  
*J Med Internet Res* 2020;22(10):e17499  
URL: <https://www.jmir.org/2020/10/e17499>  
doi: [10.2196/17499](https://doi.org/10.2196/17499)  
PMID: [33026353](https://pubmed.ncbi.nlm.nih.gov/33026353/)

©Alaa Abd-Alrazaq, Ali Abdallah Alalwan, Brian McMillan, Bridgette M Bewick, Mowafa Househ, Alaa T AL-Zyadat. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 07.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Adoption of a Personal Health Record in the Digital Age: Cross-Sectional Study

Consuela Cheriece Yousef<sup>1,2,3</sup>, PharmD, MPH, BCPS; Abin Thomas<sup>2,4,5</sup>, MSc; Ahmed O Alenazi<sup>1,2,3</sup>, BSc, PharmD, CACP; Sumaya Elgadi<sup>3,6</sup>, PharmD, MSc; Laila Carolina Abu Esba<sup>2,3,7</sup>, BSc, MSc, PharmD; Aeshah AlAzmi<sup>2,3,8</sup>, BSc, PharmD, SCSCP, FISQUA; Abrar Fahad Alhameed<sup>2,3,9</sup>, PharmD, BCPS, BCIDP; Ahmed Hattan<sup>2,3,7</sup>, PharmD; Saleh Almekhloof<sup>2,3,10</sup>, BSc, PharmD; Mohammed A AlShammmary<sup>2,3,11</sup>, BSPHarm; Nazzal Abdullah Alanezi<sup>2,3,12</sup>, BSHIM, MSHSA; Hani Solaiman Alhamdan<sup>3,8</sup>, BSc, MSc; Manal Eldegeir<sup>2,3,13</sup>, MBBS, MSc, MRCPCH, CCT; Rayf Abulezz<sup>2,3,9</sup>, BSc, PharmD, BCPS; Sahal Khoshhal<sup>2,3,9</sup>, BSc, MSc, PharmD; Clara Glynis Masala<sup>2,3,14</sup>, BA CUR; Omaima Ahmed<sup>2,3,15</sup>, MBBS

<sup>1</sup>Pharmaceutical Care Department, Ministry of National Guard-Health Affairs, Dammam, Saudi Arabia

<sup>2</sup>King Saud bin Abdul Aziz University for Health Sciences, Riyadh, Saudi Arabia

<sup>3</sup>King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

<sup>4</sup>Department of Biostatistics and Bioinformatics, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

<sup>5</sup>Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia

<sup>6</sup>Department of Pharmacy Practice, College of Pharmacy, Princess Noura Bint Abdulrahman University, Riyadh, Saudi Arabia

<sup>7</sup>Pharmaceutical Care Department, Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia

<sup>8</sup>Pharmaceutical Care Department, Ministry of National Guard-Health Affairs, Jeddah, Saudi Arabia

<sup>9</sup>Pharmaceutical Care Department, Ministry of National Guard-Health Affairs, Madinah, Saudi Arabia

<sup>10</sup>Pharmaceutical Care Department, Ministry of National Guard-Health Affairs, Al Ahsa, Saudi Arabia

<sup>11</sup>Primary Health Care, Prince Bader Housing Clinic, Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia

<sup>12</sup>Qassim Primary Health Care Center, Ministry of National Guard-Health Affairs, Qassim, Saudi Arabia

<sup>13</sup>Department of Pediatrics, Ministry of National Guard-Health Affairs, Dammam, Saudi Arabia

<sup>14</sup>Department of Nursing, Ministry of National Guard-Health Affairs, Dammam, Saudi Arabia

<sup>15</sup>Department of Pediatric Hematology/Oncology/BMT, Ministry of National Guard-Health Affairs, Jeddah, Saudi Arabia

**Corresponding Author:**

Consuela Cheriece Yousef, PharmD, MPH, BCPS

Pharmaceutical Care Department

Ministry of National Guard-Health Affairs

PO Box 4616

Dammam, 31412

Saudi Arabia

Phone: 966 853 2555 ext 31680

Email: [yousefco@ngha.med.sa](mailto:yousefco@ngha.med.sa)

**Abstract**

**Background:** As health care organizations strive to improve health care access, quality, and costs, they have implemented patient-facing eHealth technologies such as personal health records to better engage patients in the management of their health. In the Kingdom of Saudi Arabia, eHealth is also growing in accordance with Vision 2030 and its National Transformation Program framework, creating a roadmap for increased quality and efficiency of the health care system and supporting the goal of patient-centered care.

**Objective:** The aim of this study was to investigate the adoption of the personal health record of the Ministry of National Guard Health Affairs (MNGHA Care).

**Methods:** A cross-sectional survey was conducted in adults visiting outpatient clinics in hospitals at the Ministry of National Guard Health Affairs hospitals in Riyadh, Jeddah, Dammam, Madinah, and Al Ahsa, and primary health care clinics in Riyadh and Qassim. The main outcome measure was self-reported use of MNGHA Care.



**Results:** In the sample of 546 adult patients, 383 (70.1%) reported being users of MNGHA Care. MNGHA Care users were more likely to be younger ( $P<.001$ ), high school or university educated ( $P<.001$ ), employed ( $P<.001$ ), have a chronic condition ( $P=.046$ ), use the internet to search for health-related information ( $P<.001$ ), and use health apps on their mobile phones ( $P<.001$ ).

**Conclusions:** The results of this study show that there is substantial interest for the use of MNGHA Care personal health record with 70% of participants self-reporting use. To confirm these findings, objective data from the portal usage logs are needed. Maximizing the potential of MNGHA Care supports patient engagement and is aligned with the national eHealth initiative to encourage the use of technology for high-quality, accessible patient-centered care. Future research should include health care provider perspectives, incorporate objective data, employ a mixed-methods approach, and use a theoretical framework.

(*J Med Internet Res* 2020;22(10):e22913) doi:[10.2196/22913](https://doi.org/10.2196/22913)

## KEYWORDS

patient portal; personal health record; eHealth; Middle East; Saudi Arabia

## Introduction

There has been exponential growth in internet penetration globally, including in Saudi Arabia where the internet penetration rate is 93% and that of social media is 72% [1]. Living in the digital age, an increasing number of patients are empowered, computer-literate, and have access to the internet. With advances in health information technology and the now ubiquity of the internet, patients have access to health information and are expected to engage in their care in new ways. Consumer-based health apps have been developed to “transform the paternalistic model of health care into one that is responsive to consumer needs and treats each individual as a copilot in a life-long health care process” [2].

The eHealth movement, which is the delivery of health services and information through the internet or related technology, has been broadly promoted to improve the health status of patients [3,4]. Patient-facing tools have been endorsed by the Institute of Medicine and the World Health Organization to encourage patient- and person-centered care by facilitating patient involvement in medical decision-making [5,6]. The use of digital information for disease and health-related tracking is widespread and considered an “essential and important element in the health care sector” [7]. Innovative eHealth technologies such as personal health records (PHRs) are leveraged to support patients in becoming empowered. Health care organizations adopt PHRs to increase patient engagement in the drive to meet the triple aims of health care: increase access, reduce cost, and improve quality of care [8-10].

PHRs are either standalone, tethered, or integrated [11]. A standalone PHR is owned by the patient and allows a person to “access and coordinate their lifelong health information” [12]. A tethered PHR, or patient portal, is connected to an organization’s electronic health record (EHR) and allows patients to access information from their medical records [13]. An integrated PHR contains patient information from various sources such as pharmacy data, insurance claims, and an EHR. PHR features vary among health care organizations. Basic PHR features include viewing lab results, requesting prescription refills, and scheduling appointments [8,14,15]. More advanced PHR features include personal health-related reminders, secure messaging, eVisits, and social networking [16]. Since the terms “PHR” and “patient portal” have been used interchangeably in

the literature, we here consider the two terms to be synonymous [9].

PHRs are designed to increase patient engagement in managing their health, increase care coordination, and to encourage patient empowerment [16-18]. Engaged patients monitor and update their medication and there is potentially more treatment concordance, leading to positive health outcomes [17,19]. Patient engagement is increasingly recognized as a vital component of safe, person-centered care [6]. Providing patients with access to their EHRs through a PHR is a method for health care systems to promote engagement [20].

Previous studies have shown that PHR use has positive effects on patient adherence, patient self-management skills, and clinical outcomes [16,19,21]. Wade-Vuturo et al [21] found that patients with diabetes mellitus who used a patient portal had better patient-provider communication, more satisfaction with care, greater self-management behaviors, and improved clinical outcomes (ie, hemoglobin A1c, hospital admissions, and emergency room visits). The PHR could bridge the gap due to the limited time and planning devoted to addressing chronic needs since acute issues are the focus of health care visits [22]. Despite the proposed benefits and consumer interest in PHRs, various studies have shown limited adoption and use [8,9,23-25]. According to Abd-Alrazaq et al [8], PHR adoption ranged from 0.13% in the United Kingdom to 10% in the United States.

PHRs are relatively new in the Middle East with less than 12% of health care organizations in the Arab world offering them [26]. Health care organizations in the Kingdom of Saudi Arabia (KSA) have only recently begun to implement PHRs. As with many developed countries, the KSA has invested substantial resources in the implementation of eHealth systems to reduce health care costs and improve care with 4 billion Saudi Arabian Riyals (SAR; US \$1.1 billion) allocated by the government to improve eHealth [27]. The KSA has dedicated enormous funds to enhance national health care systems. To that end, the health care system is one of the priority areas of the National Transformation Program 2020 and Saudi Vision 2030, aiming to provide the highest quality of health care services to the citizens and residents while providing sufficient and efficient health care. The national eHealth initiative focused on improving the quality and efficiency of health care services by enhancing a patient-centered health care culture and increasing patient involvement in their care through technology [28]. In line with



the KSA's eHealth agenda, research on eHealth tools such as PHRs has grown [18,29-32].

The main objectives of this study were as follows: (1) to examine the prevalence of PHR use by region, (2) to categorize the PHR features used most frequently, and (3) to identify predictors of PHR adoption by patients. In addition, comparisons were made between portal users and nonusers according to (1) demographic and clinical characteristics, (2) access to and use of the internet, (3) health literacy and self-reported health status, and (4) online health-related information-seeking behavior.

Research into the actual use of the PHR focusing on the users and features accessed will lay the foundation for future developments of the system and targeted efforts to motivate patients to adopt its use. To reap the proposed benefits and maximize the return on investment, research must be conducted with an eye toward contextual factors and finding methods to promote initial and sustained PHR use among all patient populations.

## Methods

### Study Setting

The Ministry of National Guard Health Affairs (MNGHA) is a large tertiary health care system established in 1983 to provide state-of-the-art medical care to the National Guard's soldiers and their dependents in all regions across the KSA [33]. The MNGHA is a leader in health care services in the Middle East. As a health care leader, MNGHA implemented its PHR known as MNGHA Care in 2018. Some of the features available in MNGHA Care include scheduling appointments, requesting medical reports, viewing radiology reports, checking laboratory results, requesting prescription refills, and providing vaccination reminders. MNGHA Care allows patients to upload personal

health information such as weight, blood pressure, blood sugar, and exercise details. There is also a self-assessment feature where patients can enter information on pain control, performance status, and quality of life. In addition, it contains links to health educational information. MNGHA Care is a powerful tool that is expected to increase health awareness and promote positive health outcomes [34].

### Study Design

A cross-sectional survey design was used. An online survey was constructed and administered through QuestionPro. The study was conducted from December 2019 to February 2020.

### Sample

The target population consisted of adults who visited the outpatient waiting areas at Imam Abdulrahman Bin Faisal Hospital in Dammam, King Abdulaziz Medical City in Riyadh, King Abdulaziz Medical City in Jeddah, Prince Mohammad Bin AbdulAziz Hospital in Madinah, King Abdulaziz Hospital in Al Ahsa, and the primary health clinics in Riyadh and Qassim. The study was carried out at each site independently with each site's research team. Patients or their caregivers were eligible to participate if they were aged 18 years and older and able to read and understand either Arabic or English.

### Sampling Strategy

Stratified random sampling with proportionate allocation was used to draw samples from the study population. The approximate number of patients seen daily in each city was used to determine the target sample for each site. Table 1 shows the proportionate allocation by city with the sample selected in proportion to the size of the population. Since no sampling frame was available, the biostatistician generated a random day and time schedule to be used by each site when beginning data collection.

**Table 1.** Proportionate allocation by region.

Region	Approximate number of patients/day	Minimum N
Dammam	250	27
Riyadh	1120	139
Jeddah	850	75
Medina	200	22
Al Ahsa	750	80

### Sample Size

The effect size was based on 40% internet usage for Usenet, listserv, discussion forums, internet phone, and streaming audio music [35]. This effect size was then applied to find the optimal sample size to detect a proportion of the Saudi population using the internet with a predefined accuracy. We assumed the minimum weekly visit frequency as the population to calculate the overall sample sizes. We further adjusted the sample size according to the proportion of daily visits in each region. With the above assumptions, the study required at least 364 complete records to estimate the proportion at a 95% confidence limit and within 5% precision.

### Participant Recruitment

Each research site used the survey time points and days from the table prepared prior to study initiation. A member of the research team recruited participants in the outpatient pharmacy waiting areas and used a password-protected device to allow the participants to access the survey through QuestionPro, provide consent, and complete the survey. All participants were informed about the study purpose, and were assured anonymity and confidentiality of the information collected. The research team member approached the subjects at randomly selected time points during active working hours until the minimum sample size requirement for the center was reached.

## Instrument

This study used a questionnaire developed by the lead author based on a literature review. The questionnaire contained 41 questions covering demographics, health status, satisfaction with health care, health literacy, mobile phone and internet usage, online health-related information-seeking behavior, and MNGHA Care. There were 13 statements rated on a Likert scale related to MNGHA Care use included but not analyzed for this study. The authors reviewed the questions to ensure readability and appropriateness. The questionnaire was forward-translated by native Arabic speakers and was back-translated by a professional translator and compared to the original. Before study initiation, a pilot test was performed in 20 volunteers at Imam Abdulrahman Bin Faisal Hospital with slight modifications made.

## Measures

Demographic and clinical characteristics were collected to describe the study sample and were self-reported. Demographic characteristics included health care facility, age, gender, marital status, educational level, employment status, and monthly household income. Clinical characteristics included the presence of a medical condition, number and type of medical conditions, self-reported health status, recent hospitalization (<6 months), recent emergency department visit (<6 months), and satisfaction with health care.

To characterize mobile phone and internet usage, one question was asked about smartphone ownership and another question was asked about the frequency of internet use.

Previous researchers have identified a link between health literacy and technology use [36,37].

Therefore, a single-item health literacy screener was used with the following question: “How often do you need to have someone help you when you read instructions, pamphlets, or other written material from your doctor or pharmacy?” [38].

Four questions were related to internet use for seeking health information: (1) Do you use the internet to search for medical information? (2) Are you a member of an online health community? (3) Do you discuss health issues on social media (eg, Facebook, Twitter)? (4) Do you use health apps on your mobile phone?

## Outcome

The outcome variable was the patient-reported use of the PHR, operationalized by asking the patients whether or not they used MNGHA Care.

## Ethics Statement

All participants were informed about the aim of the study and their right to not answer any question they felt uncomfortable answering. The study was approved by the Institutional Review Board (RD19/002/D) at King Abdullah International Medical Research Center in Riyadh, Saudi Arabia.

## Data Analysis

The raw data were downloaded from QuestionPro into Microsoft Excel and were analyzed using SAS 9.4. The proportion of PHR users with the corresponding 95% confidence limit was calculated using the Wilson score method. Descriptive analysis was used to summarize the categorical variables as frequency and percentage and continuous variables as mean and standard deviation. For reporting convenience, the continuous variables were grouped and the frequency and percentage are reported in each group. Further, we compared the relative frequency distribution of all variables across PHR users and nonusers. The chi-square test was used to measure the significance of the association between individual covariates and PHR usage. Finally, multiple logistic regression was used to estimate the odds ratio and the corresponding 95% confidence limit. Throughout the study, we considered any *P* value less than .05 as evidence for a significant result.

## Results

### Sample Characteristics

The baseline characteristics of the participants are shown in Table 2. A total of 546 participants agreed to complete the survey. Most of the participants were men, married, employed, and university graduates. The mean age was 37.39 (SD 11.23) years. The estimated monthly income was greater than 5000 SAR/month (US \$1333/month) for 68.0% (335/493) of the participants. Most participants rated their health as very good or excellent. The majority reported only one medical condition. The most frequently reported medical conditions were diabetes, hypertension, and asthma or chronic obstructive pulmonary disease. About a quarter of the sample reported being hospitalized and approximately half had visited the emergency department in the previous 6 months. All patients were either satisfied or very satisfied with their health care. Most patients either never or rarely need help with reading materials from their physician or pharmacist.

**Table 2.** Baseline demographics of participants (N=546).

Variable	n (%)
<b>Age (years)</b>	
18-29	133 (26.3)
30-39	170 (33.6)
40-50	104 (20.6)
>50	99 (19.57)
Male	280 (53.3)
<b>Health care facility</b>	
King Abdulaziz Medical City (Riyadh)	118 (21.6)
King Abdullah Specialized Children's Hospital (Riyadh)	35 (6.4)
Primary Health Clinic (Riyadh)	32 (5.9)
King Khaled National Guard Hospital (Jeddah)	124 (22.7)
Prince Mohammed Bin Abdulaziz Hospital (Madinah)	32 (5.9)
King Abdulaziz Hospital (Al Ahsa)	145 (26.6)
Imam Abdulrahman Bin Faisal Hospital (Dammam)	27 (4.9)
Primary Health Clinic (Qassim)	33 (6.0)
Married	426 (78.9)
<b>Education</b>	
Elementary or less	65 (11.9)
Middle school	44 (8.1)
High school	177 (32.6)
University	233 (41.1)
Postgraduate	34 (6.3)
Employed	273 (45.3)
<b>Health status</b>	
Excellent	224 (41.0)
Very good	188 (34.3)
Good	94 (17.2)
Fair	27 (5.0)
Poor	13 (2.4)
<b>Number of medical conditions</b>	
0	152 (27.9)
1	280 (51.5)
≥2	112 (20.6)
<b>Type of medical condition</b>	
Diabetes	118 (21.6)
Hypertension	104 (19.0)
Asthma or COPD <sup>a</sup>	60 (11.0)
<b>Satisfaction with health care</b>	
Very satisfied	502 (92.5)
Satisfied	41 (7.6)
Hospitalized within the last 6 months	133 (24.4)
Visited the emergency department within the last 6 months	249 (45.9)

Variable	n (%)
<b>Need help with reading instructions</b>	
Never	334 (61.2)
Rarely	110 (20.1)
Sometimes	81 (14.9)
Often	16 (2.9)
Always	4 (0.7)

<sup>a</sup>COPD: chronic obstructive pulmonary disease.

## Internet Use and Online Health Information–Seeking Behavior

Table 3 shows data from the survey related to internet use and online health information–seeking behavior. An overwhelming majority of the participants reported using a smartphone and

accessing the internet several times a day. A large majority also reported using the internet to search for medical information. However, relatively few participants reported being a member of an online health community or discussing health issues on social media. Over half the participants use health apps on their mobile phones.

**Table 3.** Internet use and online health information–seeking behavior (N=546).

Internet use category	n (%)
Smartphone use	516 (94.5)
<b>Frequency of internet use</b>	
Several times a day	442 (81.0)
About once daily	30 (5.5)
A few times per week	19 (3.5)
A few times per month	22 (4.0)
Rarely or not at all	29 (5.3)
Use the internet to search for medical information	381 (69.8)
Member of online health community	44 (8.1)
Discuss health issues on social media	101 (18.5)
Use health apps on your mobile phone	299 (54.8)

## MNGHA Care Use

Of the 546 participants, 460 (84.7%) were aware of MNGHA Care. As shown in Table 4, most participants were made aware of MNGHA Care through someone from the organization—health care provider or other hospital staff—followed by a family member. Of the 460 participants aware of the PHR, 383 (83.3%) reported using it. The health care facilities were classified into the central (Riyadh and Qassim), eastern (Al Ahsa and Dammam), and western (Jeddah and Madinah) regions. There was a statistically significant difference in use by region ( $P<.001$ ): central region (182/383, 47.5%), eastern region (109/383, 28.5%), and western region (92/383, 24.0%). In response to the question “How long have

you been using MNGHA Care?”, the majority of respondents (267/383, 69.7%) reported using it for the past 12 months with most (160/383, 41.8%) using it only within the last 6 months. Most participants reported using MNGHA Care a few times a month (210/383, 54.8%) to rarely (119/383, 31.1%). PHR features accessed by the participants are shown in Figure 1. Scheduling appointments was the feature used most frequently in all regions. In the western region, use of the PHR to check laboratory results was quite low (6.5%) compared to that reported in the eastern region (48.6%) and central region (34.1%). Prescription refill requests were used the most in the eastern region (38.5%), followed by the western region (22.8%) and central region (15.9%).

**Table 4.** Source of MNGHA Care recommendation among users (N=383).

Recommender	n (%)
Health care provider	179 (46.7)
Other hospital staff	41 (10.7)
Family	90 (23.5)
Friends	47 (12.3)
No one	20 (5.2)

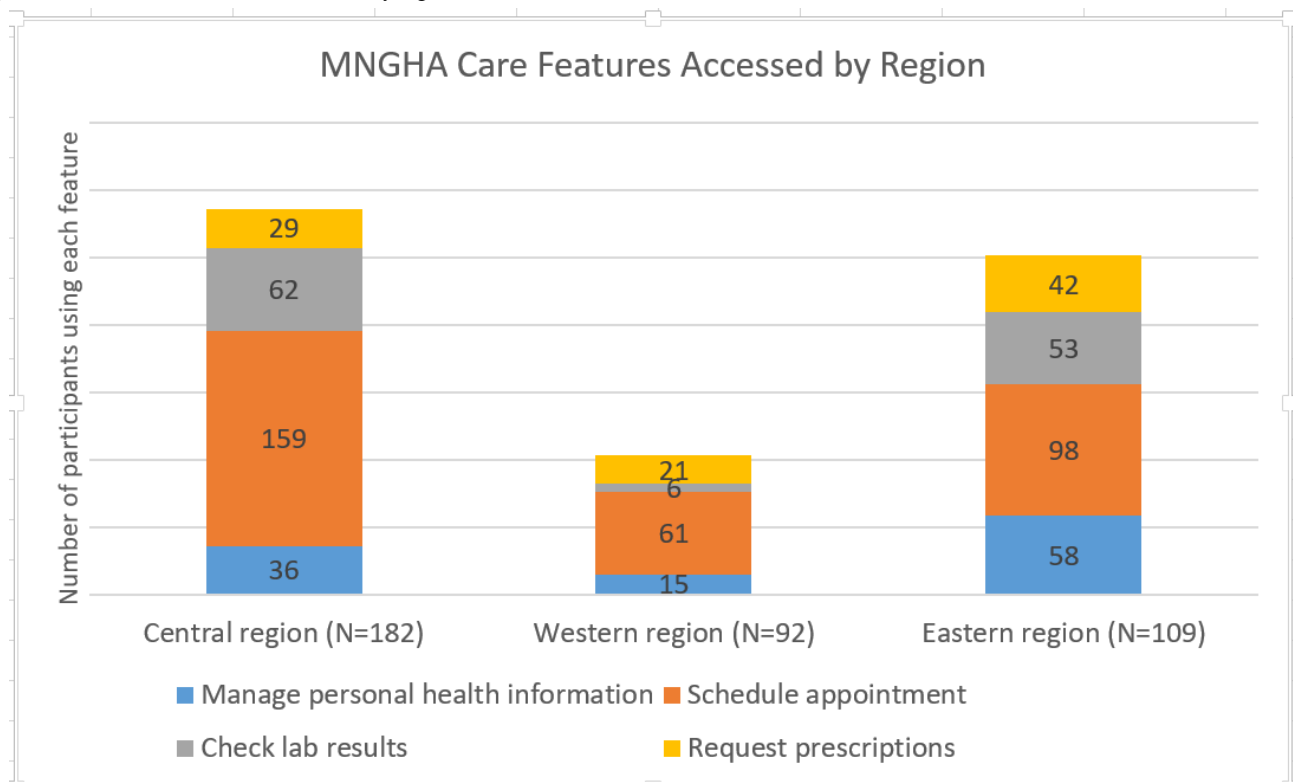
**Figure 1.** MNGHA Care features accessed by region.

Table 5 shows a comparison of the baseline demographics between MNGHA Care users and nonusers. There were statistically significant differences between the groups with respect to age, educational level, and employment status.

MNGHA Care users were younger (18-39 years), had a high school or university education, had at least one medical condition, and larger monthly household incomes relative to those of nonusers.



**Table 5.** Demographic characteristics of personal health record (PHR) users and nonusers.

Characteristic	PHR user, n (%)	PHR nonuser, n (%)	<i>P</i> value
<b>Age group (N=487)</b>			.001
18-29 years	102 (33.0)	30 (21.3)	
30-39 years	128 (41.4)	42 (29.8)	
40-49 years	79 (22.8)	25 (17.7)	
≥50 years	37 (10.7)	44 (31.2)	
<b>Gender (N=523)</b>			.44
Male	194 (52.4)	88 (56.2)	
Female	176 (47.6)	67 (43.8)	
<b>Marital status (N=537)</b>			.30
Married	303 (80.2)	121 (76.1)	
Unmarried	75 (19.8)	38 (23.9)	
<b>Educational level (N=540)</b>			<.001
Elementary school or less	26 (6.8)	38 (23.9)	
Middle school	25 (6.6)	18 (11.3)	
High school	139 (36.5)	38 (23.9)	
University	167 (43.8)	55 (34.6)	
Postgraduate	24 (6.3)	10 (6.3)	
<b>Monthly household income (N=490)</b>			.11
<1333 US \$/month	105 (30.1)	54 (37.6)	
1333-2665 US \$/month	118 (33.8)	50 (35.5)	
>2666 US \$/month	127 (36.1)	38 (27.0)	
<b>Employment status (N=536)</b>			<.001
Employed	207 (54.6)	59 (37.6)	
Unemployed	115 (30.3)	55 (35.0)	
Retired	36 (9.5)	34 (21.7)	
Student	21 (5.5)	9 (5.7)	

Table 6 compares MNGHA Care users and nonusers according to health status. Importantly, the frequency of users among those with a medical condition was more than double that of users

without a medical condition, representing a significant difference between users and nonusers according to medical condition.

**Table 6.** Health status of MNGHA Care users and nonusers.

Question	Users	Nonusers	<i>P</i> value
<b>How do you rate your health? (N=543)</b>			<.001
Poor	5 (1.3)	8 (5.0)	
Fair	13 (3.4)	14 (8.8)	
Good	54 (14.1)	39 (24.4)	
Excellent	170 (44.4)	53 (33.1)	
<b>Do you have any medical condition? (N=541)</b>			.046
Yes	265 (69.4)	124 (78.0)	
No	117 (30.6)	35 (22.0)	
<b>Number of medical conditions (N=541)</b>			.002
0	117 (30.6)	35 (22.0)	
1	201 (52.6)	77 (48.4)	
≥2	64 (16.8)	47 (29.6)	
<b>Hospitalized within the last 6 months (N=542)</b>			.04
Yes	84 (21.9)	49 (30.8)	
No	299 (78.1)	110 (69.2)	
<b>Visited the emergency department within the last 6 months (N=545)</b>			.51
Yes	179 (47.1)	70 (43.8)	
No	201 (52.9)	90 (56.3)	

[Table 7](#) compares MNGHA Care users and nonusers according to health literacy and online health-related information-seeking behavior. There were statistically significant differences in PHR

use in those who use the internet to search for medical information ( $P<.001$ ) and in those who use health apps on their mobile phone ( $P<.001$ ).

**Table 7.** Health literacy and online health information-seeking behavior.

Question	PHR <sup>a</sup> users, n (%)	PHR nonusers, n (%)	<i>P</i> value
<b>How often do you need to have someone help you when you read instructions, pamphlets, or other written material from your doctor or pharmacy? (N=539)</b>			.47
Never	230 (60.1)	102 (64.2)	
Rarely	85 (22.2)	25 (15.7)	
Sometimes	53 (13.8)	27 (17.0)	
Often	12 (3.1)	4 (2.5)	
Always	3 (0.8)	1 (0.6)	
<b>Do you use the internet to search for medical information? (N=543)</b>			<.001
Yes	303 (79.1)	77 (48.1)	
No	80 (20.9)	83 (51.9)	
<b>Are you a member of an online health community? (N=542)</b>			.61
Yes	33 (8.6)	11 (6.9)	
No	350 (91.4)	148 (93.1)	
<b>Do you discuss health issues on social media (e.g., Facebook, Twitter) (N=541)</b>			.07
Yes	79 (20.7)	22 (13.8)	
No	302 (79.3)	138 (86.3)	
<b>Do you use health apps on your mobile phone? (N=533)</b>			<.001
Yes	250 (65.6)	48 (31.6)	
No	131 (34.4)	104 (68.4)	

<sup>a</sup>PHR: personal health record.

Table 8 shows the logistic regression results assessing the influence of various participant characteristics on MNGHA Care use. Of the 9 predictor variables, only 3 were statistically significant: educational level, use of the internet to search for health-related information, and use of health apps on the mobile phone. Although higher educational level was associated with more frequent PHR use, this relationship was only statistically

significant for patients with a high school education, who had a 4.08-times higher odds of using the PHR compared with those having an elementary education ( $P=.002$ ). Patients who use the internet to search for health-related information had a 2.4-times higher odds of using the PHR ( $P=.005$ ). Finally, patients who use health apps on their mobile phones had a 2.1-times higher odds of using the PHR ( $P=.008$ ).

**Table 8.** Predictors of MNGHA Care use.

Variable	Odds ratio (95% CI)	P value
Female	1.426 (0.739-2.750)	.29
<b>Age (years)</b>		
18-29	Reference	
30-39	0.738 (0.335-1.628)	.59
40-49	0.657 (0.258-1.672)	.99
≥50	0.378 (0.129-1.105)	.08
<b>Employment status</b>		
Unemployed	Reference	
Employed	1.083 (0.457-2.569)	.39
Retired	0.703 (0.241-2.051)	.65
Student	0.640 (0.182-2.250)	.58
<b>Household income (US \$/month)</b>		
<1333	Reference	
1333-2665	1.044 (0.491-2.222)	.53
>2666	1.571 (0.641-3.849)	.23
<b>Region</b>		
Central	Reference	
Eastern	0.199 (0.098-0.406)	.09
Northern	0.521 (0.088-3.081)	.49
Southern	0.148 (0.038-0.567)	.14
Western	0.225 (0.105-0.480)	.22
<b>Education</b>		
Elementary school or less	Reference	
Middle school	2.29 (0.695-7.549)	.58
High school	4.08 (1.428-11.662)	.002
University	1.825 (0.590-5.646)	.99
Postgraduate	1.170 (0.267-5.138)	.33
<b>Number of medical conditions</b>		
0	Reference	
1	1.323 (0.722-2.425)	.95
≥2	1.695 (0.740-3.88)	.28
Internet use for health-related information	2.448 (1.32-4.539)	.005
Health apps on mobile phones	2.069 (1.209-3.539)	.008

## Discussion

### Principal Findings

In the sample of 546 adult patients, 383 (70.1%) reported being users of MNGHA Care. The central region (Riyadh and Qassim) had higher use (83.4%) than the western (Jeddah and Madinah, 59.4%) or eastern (Dammam and Al Ahsa, 64.1%) regions. There were 460 participants (84.6%) who were aware of the PHR. Of those, the majority (83.3%) reported using MNGHA Care. Despite the high penetration of internet (93%) and social

media (72%) use in the KSA, few participants reported being a member of an online health community (8.0%) or discussing health issues on social media (18.6%), whereas a high proportion of participants reported using the internet to seek health information [1]. Other researchers in the KSA have found a growing interest in patients using online social networking for health-seeking purposes [39-41].

Prior to the implementation of MNGHA Care in 2018, Al-Sahan and Saddik [31] conducted a study among 424 patients in the outpatient setting at MNGHA Riyadh to gauge the acceptance of the PHR. In their study, most patients were interested (25.2%)

or very interested (60.6%) in a PHR. The results of this study appear to be concordant with these previous findings, even though their sample was predominantly female (68.2%) with the majority having no medical condition (68.2%). Technology use was also evaluated, demonstrating high internet use (95.9%), smart device use (92.2%), and computer use (80.7%), with only 15.9% accessing patient electronic services from the MNGHA website (15.9%). Our study showed similar findings with 94.5% having a smartphone and 81% using the internet several times a day.

MNGHA Care users were more likely to be younger (18-39 years of age), high school or university educated, employed, users of the internet to search for health-related information, and users of health apps on their mobile phones. Most PHR studies have been conducted in Western countries and have noted differences in PHR use by age, gender, and ethnic background, with people of a lower socioeconomic status using PHRs less often [19,42-47]. In the systematic review conducted by Abd-Alrazaq et al [8], the factors positively associated with use of a PHR were awareness of the PHR, perceived ease of use, perceived usefulness, internet access, income, and education level. Our study also showed a positive association of PHR use with awareness, internet access, income, and educational level.

Our results indicated that increasing age is associated with lower odds of using the PHR. This is consistent with the existing literature [44,46,48-51]. Even though older patients have more chronic conditions and are in the greatest need of support in disease self-management, many do not use a PHR for a variety of reasons. Disparities in PHR use by age have frequently been cited in the literature with many studies showing that older individuals are less likely to use a PHR. Confounding factors include low computer literacy, low eHealth literacy, or less inclination to use technology, ultimately leading to more difficulties with advanced technologies such as PHRs [17,32,48]. In a study using a simulated PHR in adults aged 40 years and above, the authors concluded that adults with age-related declines in reasoning and cognitive abilities were more likely to have difficulties completing more complex health management tasks using a PHR [52]. In a study conducted in Saudi Arabia using a simulated PHR to perform various health management tasks, the authors found an increased probability to watch the help video with each 1-year increase in age, but did not note difficulties by age in completing simple compared to complex tasks [32]. The authors suggested embedding aids such as help videos in PHRs to improve the comprehension of numeric health information [53]. Special considerations should be made in the design of a PHR with targeted training sessions for the older population. Allowing patients to participate in the design of the PHR through focus groups is a strategy employed by some health care organizations [50].

Another area where the digital divide has been evident with the use of PHRs relates to education level [9,42,54]. Our study showed that a high school or university education was associated with use of the PHR; however, there was no difference in use for those with a postgraduate degree. This is inconsistent with previous research, which has shown increasing PHR use with higher levels of education. In the systematic review conducted by Zhao et al [9] examining barriers and facilitators to PHR

use, the authors recognized that one of the most common barriers is lack of a user-friendly interface. They identified 17 studies that mentioned redesigning the PHR and patient portal interfaces so that they are “easy-to use, easy-to-navigate interfaces and simpler language” [9]. Honein-AbouHaidar et al [26] evaluated the acceptance of the patient portal in Lebanon and noted the importance of simplifying and tailoring messages to the target population. Focusing on the patients and how the information is presented in the PHR will prevent the widening of health disparities.

Other studies have found that patients with chronic medical conditions are more frequent PHR users than those without [26,44,51,55]. In this study, 389 (71.9%) of the participants had a medical condition, 265 (68.1%) of whom reported using MNGHA Care. Similar to the literature, patients with a chronic medical condition had a higher prevalence of PHR use. Diabetes mellitus was the most frequently reported medical condition in our participants. Numerous studies have evaluated PHR use in this population to improve diabetes care, increase self-management, and optimize health outcomes [14,21,37,49,56-61]. Belcher et al [29] conducted a 12-week study in 31 patients with diabetes mellitus in the eastern province of Saudi Arabia who were sent twice-weekly messages through the patient portal. They found a reduction in hemoglobin A1c (11% to 9%) and fasting blood sugar (198 to 173 mg/dL). Secure messaging, a feature not available in MNGHA Care, has been associated with improved glycemic control in studies of PHR use in patients with diabetes mellitus [13]. In the study by Al Sahan and Sadek [31], 74.1% of participants reported a desire to communicate with a physician. Since patients showed an interest in secure messaging and studies have shown positive patient outcomes, it may be the right time to consider adding this feature in phases across the organization to support patient-centered communication.

This study showed that the MNGHA Care feature used most commonly was appointment scheduling (83.0%). In the study by Al Sahan and Sadek [31], participants were interested in accessing laboratory results (91.7%), radiology results (82.9%), and appointment scheduling (90.5%). With real-world use, our participants underutilized the features for managing personal health information (28.5%), checking laboratory results (31.6%), and requesting prescription refills (24.0%). Although the previous study showed a high level of patient interest in specific PHR features 2 years before MNGHA Care was implemented in the organization, utilization was less than would be expected after the implementation. Indeed, interest in PHRs and their features has been found to be higher than the actual adoption, with 80% of US respondents surveyed indicating interest in PHRs but only 2.2% of the population actually used a PHR [53]. Having the technology available is one challenge to be overcome but it is also necessary to monitor the process, find ways to connect patients to the technology, and investigate their concerns regularly postimplementation.

An explanation for the low percentage of participants requesting prescription refills is that the pharmacies in each region chose when to activate this feature. Not all pharmacies in all regions had activated the refill function at the time of performing this study. With respect to checking laboratory results and managing



personal information, patients may have disliked the interface or had problems interacting with the system. Zhao et al [9] highlighted the need for health care providers to work with patients and demonstrate the features as a way to increase the value of the PHR from the patients' perspective. There also could have been differences in the way the PHR was rolled out in each region. The central region possibly implemented more strategies aimed at encouraging patient awareness compared to the other regions. Finally, it is recognized that health care providers play an influential role in endorsing PHR use [8,9,44,47]. This is evident in our study, in which the health care provider (47.9%) or hospital staff (10.8%) was responsible for recommending the use of MNGHA Care to most participants. For better utility of the PHR, all health care providers and staff need proper training to support their patients. It is hoped that an impact of this study will be to disseminate information on the availability and benefits of the PHR and its various functions to patients and health care providers.

### Limitations

A major limitation of this study is the use of only self-reported data. Several biases are associated with self-reported data, including social desirability, recall, and nonresponse bias [62]. There may have been overreporting of PHR use by participants because they felt that this was the answer expected of them. In the systematic literature review and meta-analysis conducted by Fraccaro et al [25], the researchers found an adoption rate of 71% in controlled experiments and 23% in real-world experiments. In a cross-sectional study conducted in the Netherlands, there was a 32.1% adoption rate [44]. Objective PHR data from system usage logs can be used along with self-reported data; however, such data were not available to the research team for this study. Systems data would provide information on the total MNGHA population registered to use the PHR, patient logins, and use of specific features. In addition,

those who agreed to participate may have been more likely to use MNGHA Care compared with those who did not participate, resulting in skewed results. Another limitation is the use of only a quantitative approach for analysis. This study would have benefited from combining quantitative data with qualitative data in a mixed-methods approach. The advantage would be gaining greater insight into patient acceptance and concerns with use of MNGHA Care. Such analysis would allow for a more in-depth examination of some of the barriers and facilitators to adoption within the KSA. Another limitation is that the study was conducted in a single health system. However, the large sample size of patients from across the country who were visiting different departments should increase the generalizability of the findings. Finally, the lack of a theoretical framework possibly limits the interpretation of interrelationships between patient factors and PHR characteristics.

### Conclusions

The results of this study show that there is a great deal of interest in use of MNGHA Care with 70% of participants self-reporting use. To our knowledge, this study is the first to report on the adoption of a PHR in a real-world setting in the KSA. To confirm these findings, objective data from the portal usage logs are needed. With the COVID-19 pandemic, the entire world is learning many lessons as the eHealth landscape transforms rapidly. One lesson we are learning is how to engage in health care remotely and efficiently via electronic apps. Remote personal health engagement has become the new normal. Maximizing the potential of MNGHA Care supports patient engagement and is aligned with the national eHealth initiative to encourage the use of technology for high-quality, accessible patient-centered care. Future research should include health care provider perspectives, incorporate objective data, employ a mixed-methods approach, and use a theoretical framework.

### Acknowledgments

The authors would like to thank the data collectors Aliya Najeeb Balbaid, Raneem Mohammad Aljohani, Faten Mohammed Alasiri, Mahmoud Ismail Nouh, and Ismail Hasanin. We also greatly appreciate the assistance of the research coordinators Nadia Alotaibi, Mohammed Al Arab, and Dr. Shahinaz Bashir. Finally, we thank Marjorie Palcat for her administrative assistance at King Abdullah International Medical Research Center, Dammam.

### Conflicts of Interest

None declared.

### References

1. Digital 2020: Saudi Arabia. DataReportal. 2020 Feb 12. URL: <https://datareportal.com/reports/digital-2020-saudi-arabia> [accessed 2020-09-26]
2. Meier CA, Fitzgerald MC, Smith JM. eHealth: extending, enhancing, and evolving health care. *Annu Rev Biomed Eng* 2013;15:359-382. [doi: [10.1146/annurev-bioeng-071812-152350](https://doi.org/10.1146/annurev-bioeng-071812-152350)] [Medline: [23683088](https://pubmed.ncbi.nlm.nih.gov/23683088/)]
3. DeLuca JM, Enmark R. E-health: the changing model of healthcare. *Front Health Serv Manage* 2000;17(1):3-15. [Medline: [11184427](https://pubmed.ncbi.nlm.nih.gov/11184427/)]
4. Gerber T, Olazabal V, Brown K, Pablos-Mendez A. An agenda for action on global e-health. *Health Aff (Millwood)* 2010 Feb;29(2):233-236. [doi: [10.1377/hlthaff.2009.0934](https://doi.org/10.1377/hlthaff.2009.0934)] [Medline: [20348066](https://pubmed.ncbi.nlm.nih.gov/20348066/)]
5. Global strategy on digital health. World Health Organization eHealth. 2019. URL: <https://www.who.int/ehealth/en/> [accessed 2020-07-01]

6. Berwick D, Snair M, Nishtar S. Crossing the Global Health Care Quality Chasm: A Key Component of Universal Health Coverage. *JAMA* 2018 Oct 02;320(13):1317-1318. [doi: [10.1001/jama.2018.13696](https://doi.org/10.1001/jama.2018.13696)] [Medline: [30177995](https://pubmed.ncbi.nlm.nih.gov/30177995/)]
7. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: An extension of the UTAUT model. *Int J Med Inform* 2017 May;101:75-84. [doi: [10.1016/j.ijmedinf.2017.02.002](https://doi.org/10.1016/j.ijmedinf.2017.02.002)] [Medline: [28347450](https://pubmed.ncbi.nlm.nih.gov/28347450/)]
8. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: A systematic review. *Int J Med Inform* 2019 Jun;126:164-175. [doi: [10.1016/j.ijmedinf.2019.03.014](https://doi.org/10.1016/j.ijmedinf.2019.03.014)] [Medline: [31029258](https://pubmed.ncbi.nlm.nih.gov/31029258/)]
9. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, Facilitators, and Solutions to Optimal Patient Portal and Personal Health Record Use: A Systematic Review of the Literature. *AMIA Annu Symp Proc* 2017;2017:1913-1922 [FREE Full text] [Medline: [29854263](https://pubmed.ncbi.nlm.nih.gov/29854263/)]
10. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
11. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. *BMC Med Inform Decis Mak* 2008 Oct 06;8:45 [FREE Full text] [doi: [10.1186/1472-6947-8-45](https://doi.org/10.1186/1472-6947-8-45)] [Medline: [18837999](https://pubmed.ncbi.nlm.nih.gov/18837999/)]
12. Toscos T, Daley C, Heral L, Doshi R, Chen Y, Eckert G, et al. Impact of electronic personal health record use on engagement and intermediate health outcomes among cardiac patients: a quasi-experimental study. *J Am Med Inform Assoc* 2016 Jan;23(1):119-128. [doi: [10.1093/jamia/ocv164](https://doi.org/10.1093/jamia/ocv164)] [Medline: [26912538](https://pubmed.ncbi.nlm.nih.gov/26912538/)]
13. Heath S. How do patient portals and personal health records differ? Internet. Patient Engagement HIT. 2017 Feb 17. URL: <https://patientengagementhit.com/features/how-do-patient-portals-and-personal-health-records-differ> [accessed 2020-07-03]
14. Coughlin SS, Williams LB, Hatzigeorgiou C. A systematic review of studies of web portals for patients with diabetes mellitus. *Mhealth* 2017 Jun 12;3:23. [doi: [10.21037/mhealth.2017.05.05](https://doi.org/10.21037/mhealth.2017.05.05)] [Medline: [28736732](https://pubmed.ncbi.nlm.nih.gov/28736732/)]
15. Ammenwerth E. From eHealth to ePatient: The Role of Patient Portals in Fostering Patient Empowerment. *Eur J Biomed Inf* 2018;14(2):20-23. [doi: [10.24105/ejbi.2018.14.2.4](https://doi.org/10.24105/ejbi.2018.14.2.4)]
16. Ammenwerth E, Hoerbst A, Lannig S, Mueller G, Siebert U, Schnell-Inderst P. Effects of Adult Patient Portals on Patient Empowerment and Health-Related Outcomes: A Systematic Review. *Stud Health Technol Inform* 2019 Aug 21;264:1106-1110. [doi: [10.3233/SHTI190397](https://doi.org/10.3233/SHTI190397)] [Medline: [31438096](https://pubmed.ncbi.nlm.nih.gov/31438096/)]
17. Kruse C, Argueta D, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. *J Med Internet Res* 2015 Feb 20;17(2):e40 [FREE Full text] [doi: [10.2196/jmir.3703](https://doi.org/10.2196/jmir.3703)] [Medline: [25707035](https://pubmed.ncbi.nlm.nih.gov/25707035/)]
18. Alanazi A, Anazi YA. The Challenges in Personal Health Record Adoption. *J Healthc Manag* 2019;64(2):104-109. [doi: [10.1097/JHM-D-17-00191](https://doi.org/10.1097/JHM-D-17-00191)] [Medline: [30845058](https://pubmed.ncbi.nlm.nih.gov/30845058/)]
19. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. *J Med Internet Res* 2015 Feb 10;17(2):e44 [FREE Full text] [doi: [10.2196/jmir.3171](https://doi.org/10.2196/jmir.3171)] [Medline: [25669240](https://pubmed.ncbi.nlm.nih.gov/25669240/)]
20. Irizarry T, Shoemaker J, Nilsen ML, Czaja S, Beach S, DeVito Dabbs A. Patient Portals as a Tool for Health Care Engagement: A Mixed-Method Study of Older Adults With Varying Levels of Health Literacy and Prior Patient Portal Use. *J Med Internet Res* 2017 Mar 30;19(3):e99 [FREE Full text] [doi: [10.2196/jmir.7099](https://doi.org/10.2196/jmir.7099)] [Medline: [28360022](https://pubmed.ncbi.nlm.nih.gov/28360022/)]
21. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. *J Am Med Inform Assoc* 2013 May 01;20(3):519-525 [FREE Full text] [doi: [10.1136/amiainl-2012-001253](https://doi.org/10.1136/amiainl-2012-001253)] [Medline: [23242764](https://pubmed.ncbi.nlm.nih.gov/23242764/)]
22. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. *JAMA* 2002 Oct 09;288(14):1775-1779. [doi: [10.1001/jama.288.14.1775](https://doi.org/10.1001/jama.288.14.1775)] [Medline: [12365965](https://pubmed.ncbi.nlm.nih.gov/12365965/)]
23. Nambisan P. Factors that impact Patient Web Portal Readiness (PWPR) among the underserved. *Int J Med Inform* 2017 Jun;102:62-70. [doi: [10.1016/j.ijmedinf.2017.03.004](https://doi.org/10.1016/j.ijmedinf.2017.03.004)] [Medline: [28495349](https://pubmed.ncbi.nlm.nih.gov/28495349/)]
24. Zhong X, Park J, Liang M, Shi F, Budd P, Sprague J, et al. Characteristics of Patients Using Different Patient Portal Functions and the Impact on Primary Care Service Utilization and Appointment Adherence: Retrospective Observational Study. *J Med Internet Res* 2020 Feb 25;22(2):e14410 [FREE Full text] [doi: [10.2196/14410](https://doi.org/10.2196/14410)] [Medline: [32130124](https://pubmed.ncbi.nlm.nih.gov/32130124/)]
25. Fraccaro V, Vigo M, Balatsoukas P, Buchan IE, Peek N, van der Veer SN. Patient Portal Adoption Rates: A Systematic Literature Review and Meta-Analysis. *Stud Health Technol Inform* 2017;245:79-83. [Medline: [29295056](https://pubmed.ncbi.nlm.nih.gov/29295056/)]
26. Honein-AbouHaidar GN, Antoun J, Badr K, Hlais S, Nazaretian H. Users' acceptance of electronic patient portals in Lebanon. *BMC Med Inform Decis Mak* 2020 Feb 17;20(1):31 [FREE Full text] [doi: [10.1186/s12911-020-1047-x](https://doi.org/10.1186/s12911-020-1047-x)] [Medline: [32066425](https://pubmed.ncbi.nlm.nih.gov/32066425/)]
27. AlBar AM, Hoque MR. Patient Acceptance of e-Health Services in Saudi Arabia: An Integrative Perspective. *Telemed J E Health* 2019 Sep;25(9):847-852. [doi: [10.1089/tmj.2018.0107](https://doi.org/10.1089/tmj.2018.0107)] [Medline: [30452333](https://pubmed.ncbi.nlm.nih.gov/30452333/)]
28. Vision 2030. National Transformation Program. URL: <https://vision2030.gov.sa/en/programs/NTP> [accessed 2020-06-15]
29. Belcher T, Vess J, Johnson E. Using Patient Portal Messaging to Improve Glycemic Control in Adult Patients with Diabetes in Saudi Arabia. HIMSS. 2019. URL: <https://www.himss.org/resources/using-patient-portal-messaging-improve-glycemic-control-adult-patients-diabetes-saudi> [accessed 2020-05-07]

30. Alhammad O. Assessing outpatients' attitudes and expectations in secondary and tertiary hospitals in Riyadh, Saudi Arabia, Masters Thesis. Hamilton, Ontario, Canada: McMaster University; 2017. URL: <https://macsphere.mcmaster.ca/handle/11375/21250> [accessed 2020-05-07]
31. Al-Sahan A, Saddik B. Perceived challenges for adopting the Personal Health Record (PHR) at Ministry of National Guard Health Affairs (MNGHA)- Riyadh. *Online J Public Health Inform* 2016;8(3):e205 [FREE Full text] [doi: [10.5210/ojphi.v8i3.6845](https://doi.org/10.5210/ojphi.v8i3.6845)] [Medline: [28210426](https://pubmed.ncbi.nlm.nih.gov/28210426/)]
32. Al-Mifgai BF, Sharit J, Onar-Thomas A, Asfour S. Examining attitudes toward and ability to interact with an online personal health record: A case study in Saudi Arabia. *Int J Healthc* 2019 Aug 15;6(1):1. [doi: [10.5430/ijh.v6n1p1](https://doi.org/10.5430/ijh.v6n1p1)]
33. MNGHA Organization Profile. Ministry of National Guard-Health Affairs. 2018. URL: <https://ngha.med.sa/English/AboutNGHA/Pages/profile.aspx> [accessed 2019-09-06]
34. MNGHA Care mobile app. Ministry of National Guard-Health Affairs. URL: <https://ngha.med.sa/English/eServices/Pages/MobileServices/patientportal.aspx> [accessed 2020-04-05]
35. Sait S, Al-Tawil K. Impact of internet usage in Saudi Arabia: A social perspective. *Int J Inf Technol Web Eng* 2007;2(2):81-115. [doi: [10.4018/jitwe.2007040104](https://doi.org/10.4018/jitwe.2007040104)]
36. Davis SE, Osborn CY, Kripalani S, Goggins KM, Jackson GP. Health Literacy, Education Levels, and Patient Portal Usage During Hospitalizations. *AMIA Annu Symp Proc* 2015;2015:1871-1880 [FREE Full text] [Medline: [26958286](https://pubmed.ncbi.nlm.nih.gov/26958286/)]
37. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. *Curr Diab Rep* 2010 Dec;10(6):422-435 [FREE Full text] [doi: [10.1007/s11892-010-0151-1](https://doi.org/10.1007/s11892-010-0151-1)] [Medline: [20890688](https://pubmed.ncbi.nlm.nih.gov/20890688/)]
38. Morris NS, MacLean CD, Chew LD, Littenberg B. The Single Item Literacy Screener: evaluation of a brief instrument to identify limited reading ability. *BMC Fam Pract* 2006 Mar 24;7:21 [FREE Full text] [doi: [10.1186/1471-2296-7-21](https://doi.org/10.1186/1471-2296-7-21)] [Medline: [16563164](https://pubmed.ncbi.nlm.nih.gov/16563164/)]
39. Al-Qahtani M, Alsaffar A, Alshammasi A, Alsanni G, Alyousef Z, Alhussaini M. Social media in healthcare: Advantages and challenges perceived by patients at a teaching hospital in eastern province, Saudi Arabia. *Saudi J Health Sci* 2018;7(2):116. [doi: [10.4103/sjhs.sjhs\\_36\\_18](https://doi.org/10.4103/sjhs.sjhs_36_18)]
40. ALrukban A. The health related uses of social media among users in Saudi Arabia. *Int J Med Sci Public Health* 2014;3(12):1492. [doi: [10.5455/ijmsph.2014.170920141](https://doi.org/10.5455/ijmsph.2014.170920141)]
41. Alzahrani A, Alanzi T. Social Media Use By People With Diabetes In Saudi Arabia: A Survey About Purposes, Benefits And Risks. *Diabetes Metab Syndr Obes* 2019;12:2363-2372. [doi: [10.2147/DMSO.S208141](https://doi.org/10.2147/DMSO.S208141)] [Medline: [32009810](https://pubmed.ncbi.nlm.nih.gov/32009810/)]
42. Tavares J, Oliveira T. Electronic Health Record Patient Portal Adoption by Health Care Consumers: An Acceptance Model and Survey. *J Med Internet Res* 2016 Mar 02;18(3):e49 [FREE Full text] [doi: [10.2196/jmir.5069](https://doi.org/10.2196/jmir.5069)] [Medline: [26935646](https://pubmed.ncbi.nlm.nih.gov/26935646/)]
43. Tavares J, Oliveira T. Electronic Health Record Portal Adoption: a cross country analysis. *BMC Med Inform Decis Mak* 2017 Jul 05;17(1):97 [FREE Full text] [doi: [10.1186/s12911-017-0482-9](https://doi.org/10.1186/s12911-017-0482-9)] [Medline: [28679423](https://pubmed.ncbi.nlm.nih.gov/28679423/)]
44. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the Users of a Patient Portal: Cross-Sectional Study. *J Med Internet Res* 2018 Sep 17;20(9):e262 [FREE Full text] [doi: [10.2196/jmir.9418](https://doi.org/10.2196/jmir.9418)] [Medline: [30224334](https://pubmed.ncbi.nlm.nih.gov/30224334/)]
45. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. *Diabetes Technol Ther* 2014 Nov;16(11):784-793. [doi: [10.1089/dia.2014.0078](https://doi.org/10.1089/dia.2014.0078)] [Medline: [24999599](https://pubmed.ncbi.nlm.nih.gov/24999599/)]
46. Peacock S, Reddy A, Leveille S, Walker J, Payne T, Oster NV, et al. Patient portals and personal health information online: perception, access, and use by US adults. *J Am Med Inform Assoc* 2017 Apr 01;24(e1):e173-e177. [doi: [10.1093/jamia/ocw095](https://doi.org/10.1093/jamia/ocw095)] [Medline: [27413120](https://pubmed.ncbi.nlm.nih.gov/27413120/)]
47. Vreugdenhil M, Ranke S, de Man Y, Haan M, Kool R. Patient and Health Care Provider Experiences With a Recently Introduced Patient Portal in an Academic Hospital in the Netherlands: Mixed Methods Study. *J Med Internet Res* 2019 Aug 20;21(8):13743 [FREE Full text] [doi: [10.2196/13743](https://doi.org/10.2196/13743)] [Medline: [31432782](https://pubmed.ncbi.nlm.nih.gov/31432782/)]
48. Khaneghah P, Miguel-Cruz A, Bentley P, Liu L, Stroulia E, Ferguson-Pell M. Users' Attitudes Towards Personal Health Records: A Cross-Sectional Pilot Study. *Appl Clin Inform* 2016;7(2):573-586 [FREE Full text] [doi: [10.4338/ACI-2015-12-RA-0180](https://doi.org/10.4338/ACI-2015-12-RA-0180)] [Medline: [27437062](https://pubmed.ncbi.nlm.nih.gov/27437062/)]
49. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Patients' Experiences with and Attitudes towards a Diabetes Patient Web Portal. *PLoS One* 2015;10(6):e0129403 [FREE Full text] [doi: [10.1371/journal.pone.0129403](https://doi.org/10.1371/journal.pone.0129403)] [Medline: [26086272](https://pubmed.ncbi.nlm.nih.gov/26086272/)]
50. Showell C. Barriers to the use of personal health records by patients: a structured review. *PeerJ* 2017;5:e3268. [doi: [10.7717/peerj.3268](https://doi.org/10.7717/peerj.3268)] [Medline: [28462058](https://pubmed.ncbi.nlm.nih.gov/28462058/)]
51. Bauer A, Rue T, Munson S, Ghomi R, Keppel G, Cole A, et al. Patient-oriented health technologies: Patients' perspectives and use. *J Mob Technol Med* 2017 Aug;6(2):1-10 [FREE Full text] [doi: [10.7309/jmtm.6.2.1](https://doi.org/10.7309/jmtm.6.2.1)] [Medline: [28936236](https://pubmed.ncbi.nlm.nih.gov/28936236/)]
52. Taha J, Czaja SJ, Sharit J, Morrow DG. Factors affecting usage of a personal health record (PHR) to manage health. *Psychol Aging* 2013 Dec;28(4):1124-1139 [FREE Full text] [doi: [10.1037/a0033911](https://doi.org/10.1037/a0033911)] [Medline: [24364414](https://pubmed.ncbi.nlm.nih.gov/24364414/)]
53. Househ MS, Borycki EM, Rohrer WM, Kushniruk AW. Developing a framework for meaningful use of personal health records (PHRs). *Health Policy Techn* 2014 Dec;3(4):272-280. [doi: [10.1016/j.hlpt.2014.08.009](https://doi.org/10.1016/j.hlpt.2014.08.009)]
54. Tsai J, Rosenheck R. Use of the internet and an online personal health record system by US veterans: comparison of Veterans Affairs mental health service users and other veterans nationally. *J Am Med Inform Assoc* 2012;19(6):1089-1094 [FREE Full text] [doi: [10.1136/amiajnl-2012-000971](https://doi.org/10.1136/amiajnl-2012-000971)] [Medline: [22847305](https://pubmed.ncbi.nlm.nih.gov/22847305/)]

55. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. *Ann Fam Med* 2014;12(5):418-426. [doi: [10.1370/afm.1691](https://doi.org/10.1370/afm.1691)] [Medline: [25354405](https://pubmed.ncbi.nlm.nih.gov/25354405/)]
56. Coughlin SS, Heboyan V, Williams LB, Hatzigeorgiou C. Use of a web portal among adult clinic patients seen for type 2 diabetes mellitus. *J Hosp Manag Health Policy* 2018 Jan;2:1 [FREE Full text] [doi: [10.21037/jhmhp.2017.12.04](https://doi.org/10.21037/jhmhp.2017.12.04)] [Medline: [29377041](https://pubmed.ncbi.nlm.nih.gov/29377041/)]
57. Sun R, Burke LE, Saul MI, Korytkowski MT, Li D, Sereika SM. Use of a Patient Portal for Engaging Patients with Type 2 Diabetes: Patterns and Prediction. *Diabetes Technol Ther* 2019 Oct;21(10):546-556. [doi: [10.1089/dia.2019.0074](https://doi.org/10.1089/dia.2019.0074)] [Medline: [31335206](https://pubmed.ncbi.nlm.nih.gov/31335206/)]
58. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. *J Am Med Inform Assoc* 2013 May 01;20(3):526-534 [FREE Full text] [doi: [10.1136/amiajnl-2012-001263](https://doi.org/10.1136/amiajnl-2012-001263)] [Medline: [23171659](https://pubmed.ncbi.nlm.nih.gov/23171659/)]
59. Tao D, Or CK. Effects of self-management health information technology on glycaemic control for patients with diabetes: a meta-analysis of randomized controlled trials. *J Telemed Telecare* 2013 Apr;19(3):133-143. [doi: [10.1177/1357633X13479701](https://doi.org/10.1177/1357633X13479701)] [Medline: [23563018](https://pubmed.ncbi.nlm.nih.gov/23563018/)]
60. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. *Arch Intern Med* 2008 Sep 08;168(16):1776-1782 [FREE Full text] [doi: [10.1001/archinte.168.16.1776](https://doi.org/10.1001/archinte.168.16.1776)] [Medline: [18779465](https://pubmed.ncbi.nlm.nih.gov/18779465/)]
61. Grant RW, Wald JS, Poon EG, Schnipper JL, Gandhi TK, Volk LA, et al. Design and implementation of a web-based patient portal linked to an ambulatory care electronic health record: patient gateway for diabetes collaborative care. *Diabetes Technol Ther* 2006 Oct;8(5):576-586 [FREE Full text] [doi: [10.1089/dia.2006.8.576](https://doi.org/10.1089/dia.2006.8.576)] [Medline: [17037972](https://pubmed.ncbi.nlm.nih.gov/17037972/)]
62. Timmins F. Nursing Research Generating and Assessing Evidence for Nursing Practice. *Nurse Educ Pract* 2013 Nov;13(6):e29. [doi: [10.1016/j.nepr.2013.04.001](https://doi.org/10.1016/j.nepr.2013.04.001)]

## Abbreviations

**EHR:** electronic health record

**KSA:** Kingdom of Saudi Arabia

**MNGHA:** Ministry of National Guard Health Affairs

**PHR:** personal health record

**SAR:** Saudi Arabian Riyal

*Edited by G Eysenbach; submitted 30.07.20; peer-reviewed by N Fareed, J Pecina; comments to author 25.09.20; revised version received 26.09.20; accepted 30.09.20; published 28.10.20.*

### Please cite as:

Yousef CC, Thomas A, Alenazi AO, Elgadi S, Abu Esba LC, AlAzmi A, Alhameed AF, Hattan A, Almekhloof S, AlShammary MA, Alanezi NA, Alhamdan HS, Eldegeir M, Abulezz R, Khoshhal S, Masala CG, Ahmed O  
Adoption of a Personal Health Record in the Digital Age: Cross-Sectional Study

*J Med Internet Res* 2020;22(10):e22913

URL: <http://www.jmir.org/2020/10/e22913/>

doi: [10.2196/22913](https://doi.org/10.2196/22913)

PMID: [32998854](https://pubmed.ncbi.nlm.nih.gov/32998854/)

©Consuela Cheriece Yousef, Abin Thomas, Ahmed O Alenazi, Sumaya Elgadi, Laila Carolina Abu Esba, Aeshah AlAzmi, Abrar Fahad Alhameed, Ahmed Hattan, Saleh Almekhloof, Mohammed A AlShammary, Nazzal Abdullah Alanezi, Hani Solaiman Alhamdan, Manal Eldegeir, Rayf Abulezz, Sahal Khoshhal, Clara Glynis Masala, Omaira Ahmed. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Original Paper

# Evaluation of the Perceived Persuasiveness Questionnaire: User-Centered Card-Sort Study

Nienke Beerlage-de Jong<sup>1,2</sup>, PhD; Hanneke Kip<sup>1,3</sup>, MSc; Saskia Marion Kelders<sup>1,4</sup>, PhD

<sup>1</sup>Department of Psychology, Health and Technology, Center for eHealth and Wellbeing Research, University of Twente, Enschede, Netherlands

<sup>2</sup>Department of Medical Microbiology and Infection Prevention, University Medical Centre Groningen, Groningen, Netherlands

<sup>3</sup>Department of Research, Stichting Transfore, Deventer, Netherlands

<sup>4</sup>Optentio Research Focus Area, North-West University, Vanderbijlpark, South Africa

**Corresponding Author:**

Nienke Beerlage-de Jong, PhD

Department of Psychology, Health and Technology

Center for eHealth and Wellbeing Research

University of Twente

P O Box 217

7500AE

Enschede

Netherlands

Phone: 31 534893517

Email: [n.beerlage-dejong@utwente.nl](mailto:n.beerlage-dejong@utwente.nl)

## Abstract

**Background:** eHealth technologies aim to change users' health-related behavior. *Persuasive* design and system features can make an eHealth technology more motivating, engaging, or supportive to its users. The Persuasive Systems Design (PSD) model incorporates software features that have the possibility to increase the persuasiveness of technologies. However, the effects of specific PSD software features on the effectiveness of an intervention are still largely unknown. The Perceived Persuasiveness Questionnaire (PPQ) was developed to gain insight into the working mechanisms of persuasive technologies. Although the PPQ seems to be a suitable method for measuring subjective persuasiveness, it needs to be further evaluated to determine how suitable it is for measuring perceived persuasiveness among the public.

**Objective:** This study aims to evaluate the face and construct validity of the PPQ, identify points of improvement, and provide suggestions for further development of the PPQ.

**Methods:** A web-based closed-ended card-sort study was performed wherein participants grouped existing PPQ items under existing PPQ constructs. Participants were invited via a Massive Open Online Course on eHealth. A total of 398 people (average age 44.15 years, SD 15.17; 251/398, 63.1% women) completed the card sort. Face validity was evaluated by determining the item-level agreement of the original PPQ constructs. Construct validity was evaluated by determining the construct in which each item was placed most often, regardless of the original placement and how often 2 items were (regardless of the constructs) paired together and what interitem correlations were according to a cluster analysis.

**Results:** Four PPQ constructs obtained relatively high face validity scores: perceived social support, use continuance, perceived credibility, and perceived effort. Item-level agreement on the other constructs was relatively low. Item-level agreement for almost all constructs, except perceived effort and perceived effectiveness, would increase if items would be grouped differently. Finally, a cluster analysis of the PPQ indicated that the strengths of the newly identified 9 clusters varied strongly. Unchanged strong clusters were only found for perceived credibility support, perceived social support, and use continuance. The placement of the other items was much more spread out over the other constructs, suggesting an overlap between them.

**Conclusions:** The findings of this study provide a solid starting point toward a redesigned PPQ that is a true asset to the field of persuasiveness research. To achieve this, we advocate that the redesigned PPQ should adhere more closely to what persuasiveness is according to the PSD model and to the mental models of potential end users of technology. The revised PPQ should, for example, *enquire if* the user thinks anything is done to provide task support but not *how* this is done exactly.

(J Med Internet Res 2020;22(10):e20404) doi:[10.2196/20404](https://doi.org/10.2196/20404)



**KEYWORDS**

eHealth; behavior change support systems; card sort; perceived persuasiveness; persuasive systems design; mental model; questionnaire evaluation

**Introduction**

Owing to the increasing pressure on health care systems, an increasing number of health-related interventions are making use of technology. This is also referred to as eHealth: the use of technology to support health, well-being, and (the organization of) health care [1]. Many eHealth technologies aim to change their users' health-related attitudes or behaviors. A known pitfall of such technologies is that users stop using them rather quickly or do not use the technology as intended by the developers [2]. This issue of nonadherence can limit the effectiveness of technology because of the dose-response relationship, that is, the more a technology is used, the more effective it is [2]. Furthermore, limited use of technology also hampers sustainability. To deal with this downside of eHealth, it is pivotal that users of eHealth technologies are continuously motivated and engaged by the technology itself. To that end, eHealth technologies can make use of certain design and system features that can make technology more motivating, engaging, and supportive to its users, in turn helping them become *persuasive* eHealth technologies [3-7]. Persuasive (eHealth) technologies may be defined as "computerized software or information systems designed to reinforce, change or shape attitudes or behaviors or both without using coercion or deception" [8]. Despite the potential benefits of persuasion, insight into whether, why, for whom, and how persuasive eHealth technology works remains limited. Most of the existing research has focused on whether persuasive eHealth technology as a whole is effective, which provides little insight into what it is that makes it effective and what the role of persuasiveness is [9]. Thus, there is a need to open the black box of persuasive eHealth technologies to gain insight into whether and how persuasive technology actually increases adherence and whether persuasiveness is related to the effectiveness of such technologies.

**How Does Persuasion Work?**

Although we currently have a broad sense of which persuasive strategies are preferred by certain user groups (eg, based on personality traits, age, or gender) [10-14], more research is needed to unravel why and how persuasive technologies work for different types of users. To evaluate the effect of persuasion, we need to gain insight into whether technology users actually experience the included persuasive software features (eg, personalization and social comparison) as persuasive. After all, the persuasive strategies that designers apply in the software may not be perceived as such by the users [15]. It has been stated that the persuasive powers of a technology are determined by an individual's subjective evaluation of that technology and its impact on the self [16]. This highlights the importance of the subjective experience of persuasiveness as opposed to merely checking whether certain persuasive features are present according to the researchers or developers of the technology. It would therefore be extremely helpful for persuasive technology researchers and developers to have an instrument at hand that

enables them to reliably check the persuasive aspects of technologies, not from a designer's perspective but from a user's perspective. This would not only support developers in making their designs more persuasive but also provide the necessary step to evaluate whether technology increases adherence and effectiveness through the persuasiveness of the technology as theorized.

**Toward Designing Persuasive Technologies**

For developers and designers, the Persuasive Systems Design (PSD) model was created as a guideline for the development of persuasive eHealth technologies [17]. It provides 4 categories of software features that can be applied to persuasive systems: *primary task support*, *dialogue support*, *social support*, and *credibility support* [17] (Table 1). The model posits that all these features have the ability to increase the persuasiveness of technologies. However, as previously indicated, the effects of specific software features on the persuasiveness of an intervention are still largely unknown. Thus, measuring persuasion would require a measurement tool to evaluate the subjective (*perceived*) persuasion of a technology for the individual user. There are 3 newly developed questionnaires that might play a role in achieving this goal. First, there is the Persuadability Inventory [18]; this is a questionnaire aimed at measuring how susceptible an individual is to certain persuasive strategies such as rewards or social comparison. Although this is an interesting focus area, it does not measure the effects of specific software features on different individuals and, therefore, cannot be used for evaluating the persuasiveness of a system. The second related questionnaire is the Persuasive Potential Questionnaire [19]. This questionnaire is aimed at measuring the potential of different systems to persuade users, especially at the stage when the system is not fully developed. As such, it also does not evaluate the effect of specific software features within a system and seems to focus on how persuadable an individual is. Therefore, this questionnaire seems ill suited for the goal of evaluating the persuasiveness of a system. The third questionnaire is the Perceived Persuasiveness Questionnaire (PPQ) [16,20]. The PPQ is a 31-item scale that assesses perceived persuasiveness according to the 4 categories of the PSD (primary task support, dialogue support, social support, and credibility support) and the 4 related constructs: unobtrusiveness, effort, effectiveness, and (overall) perceived persuasiveness [16,20].

Although the goal of the PPQ is not clearly articulated in the first papers that it appeared in, it has been used to evaluate how users perceive the persuasiveness of a certain system (a digital weight loss intervention) and how the different included constructs relate to each other within a structural equation model. Later studies have used the PPQ to evaluate the persuasive effects of different virtual agents [21] in a web-based weight loss intervention [22] and in an antibiotic information app for nurses [23]. Thus, the questionnaire has been applied to evaluate interventions as a whole or specific persuasive elements within such interventions, and it has been applied to

the general public and health care professionals as well. Overall, as the intent of the PPQ is to measure perceived persuasiveness, it is interesting to evaluate to what extent the scale can be used to assess the perceived persuasiveness of technologies.

Therefore, this study addresses questions related to how to evaluate and measure the dimensions of persuasion in eHealth. The PPQ is recognized as a promising low-threshold instrument to measure perceived persuasiveness among a heterogeneous set of people [16,20]. However, to live up to its potential, it is important to evaluate whether the PPQ truly matches the underlying PSD model and the mental models of its broad range of potential target groups [15,24]. A mental model is a person's simplified representation of how something works in the real world. As the PPQ focuses on how persuasiveness is *perceived*, it is even more essential that its items actually match its users' mental models. Consequently, this study aims to evaluate the

face and construct validity of the PPQ constructs to identify points of improvement for a potentially updated version. To achieve this, a broad target group with an interest in eHealth was asked to evaluate the face and construct validity of the PPQ constructs by means of a card-sort study. This provides insight into the extent to which the PPQ items are grouped under the PPQ constructs as intended by its developers and as such provide face validity. Furthermore, we explore the extent to which participants see the constructs included in the PPQ as separate from each other and the items as fitting with each other, which provides information on construct validity. Together, this provides insight into whether the PPQ items are capable of rendering a valid impression of a technology's perceived persuasiveness and whether the PPQ items match the mental models of the users of the technology. The outcomes of this study can be used to establish whether any changes are needed to the PPQ constructs and items.

**Table 1.** Overview of the categories of the Persuasive Systems Design model.

Category	Description	Examples of design features
Primary task support	What the technology does to support the user in carrying out his primary task	Reduction, tunneling, tailoring, personalization, self-monitoring, simulation, rehearsal
Dialogue support	How the technology supports its users via computer-human communication	Praise, rewards, reminders, suggestions, similarity, liking, social role
Social support	How the technology uses social influence to motivate its users	Social facilitation, social comparison, normative influence, social learning, cooperation, competition, recognition
Credibility support	How the technology's design contributes to instilling trust in its users	Trustworthiness, expertise, surface, credibility, real-world feel, third-party endorsements, verifiability

## Methods

### Design

Traditionally, the card-sorting methodology is used to create and evaluate a fit between an information structure and the mental models of its target group such that it enables them to easily find, interpret, and apply information [25-27]. However, previous research has indicated that it can also yield valuable insights into how mental models fit with other kinds of structuring of information, such as the items of a questionnaire [28].

Specifically, in this study, we applied the card-sort methodology to evaluate the face and construct validity of a questionnaire. The questionnaire's face validity refers to the extent to which its items subjectively (at first glance) seem to actually cover the constructs they are intended to measure [29]. In the card sort, this was ascertained by evaluating how many participants had grouped each individual item in the intended construct. The questionnaire's construct validity represents the extent to which the included items are actually related to each other and the construct they intend to measure [30]. In the card sort, this was evaluated by analyzing how the items are grouped and placed in constructs by the respondents themselves, according to their own mental models regardless of the constructs that the items originally belonged to. This cross-sectional study used a closed-ended card sort to evaluate the face and construct validity of the PPQ. In this type of card-sort study, the categories that have to be used to group items have already been defined

beforehand. During the card-sort study, participants were asked to sort cards with, in this case, questionnaire items into the PPQ constructs (primary task support, dialogue support, perceived credibility, perceived social support, perceived unobtrusiveness, perceived persuasiveness, perceived effort, perceived effectiveness, and use continuance) in a way that was logical or meaningful to them. The resulting groups render information about the participants' mental models concerning these constructs, including an agreement or disagreement among users [31].

### Participants

Participants were invited to join the study via a Massive Open Online Course (MOOC) on *eHealth: Combining Psychology, Technology and Health* that was hosted by the research group of the authors of this paper [32]. Anyone with an interest in eHealth technology can join the MOOC, free of charge. It is targeted at an international audience via the FutureLearn platform [33]. Participants of the MOOC were invited to participate in the fourth week (of a total of 6 weeks) of the MOOC. This week of the course was entirely focused on the persuasive design of eHealth technology. In addition, at that point of the web-based course, participants had already finished reading and practicing with information about eHealth, human-centered design, and behavior change. This means that they had acquired at least some basic knowledge about eHealth and persuasive design. After completing the lessons of week 4, participants were presented with a short introductory text including disclosure about the research ([Multimedia Appendix](#)

1). They were invited to participate on a voluntary basis in this study. Data were collected during 10 runs of the MOOC, which ran from May 23, 2016, to December 8, 2019. A total of 12,439 people actively participated in the MOOC during this time span. A total of 602 people started filling in the questionnaire, of which 398 (251/398, 63.1% women; 147/398, 36.1% men) completed the study and were included. The average response rate per MOOC run was 3.35% (398/11,880) of the active learners of the MOOC. The participants' mean age was 44.15 years (SD 15.17) with participants' age ranging from 18 to 81 years. Most participants (288/398, 72.4%) were *somewhat familiar* with eHealth. However, despite having joined the MOOC, some (17/398, 4.3%) participants indicated that they were not familiar with eHealth at all. The first participant completed the card sort on June 11, 2016, and the last participant completed the card sort on June 25, 2019. Most (121/398, 30.4%) of the participants had joined the MOOC during its very first run. Participants' data were included in the data analysis if all demographic questions were answered and the card sort was completed.

## Materials and Procedure

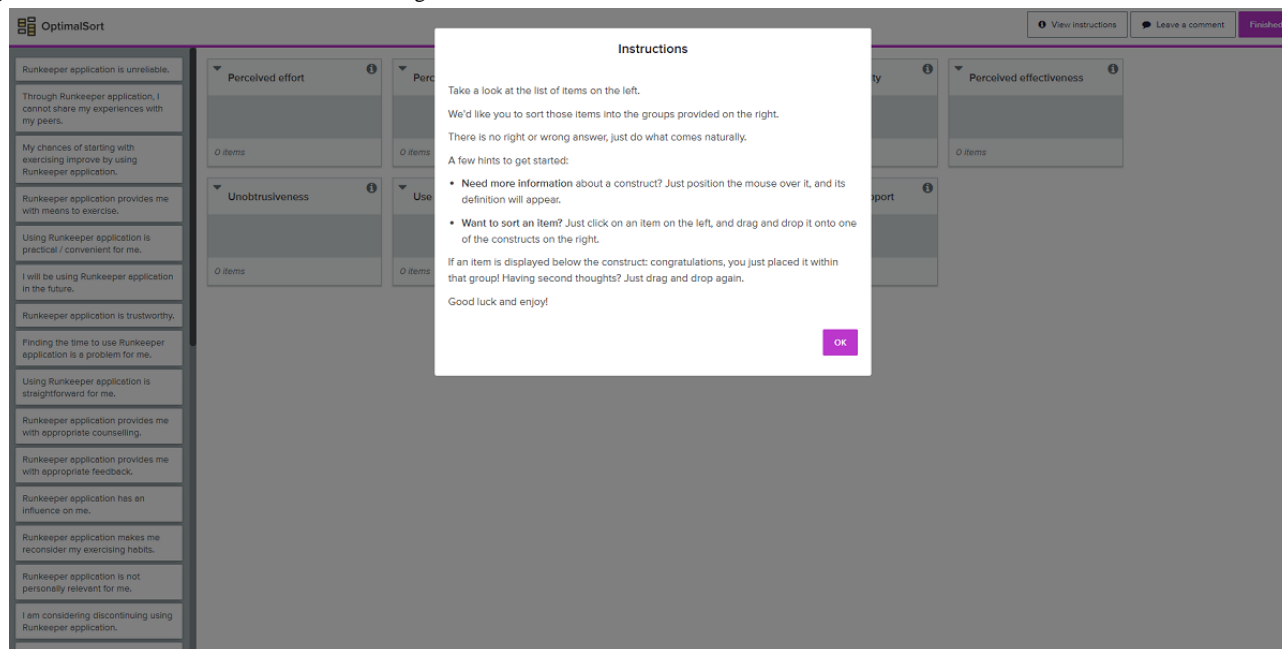
All participants received an explanation of the background, purpose, and methodology of the study. Optimal workshop software [34] was used for this study. This software allowed us to inquire about the participants' demographics (ie, gender, age, and familiarity with eHealth) and complete the card sort at a single (web) location. The software provided participants with instructions on how to perform the card sort and how to sort the

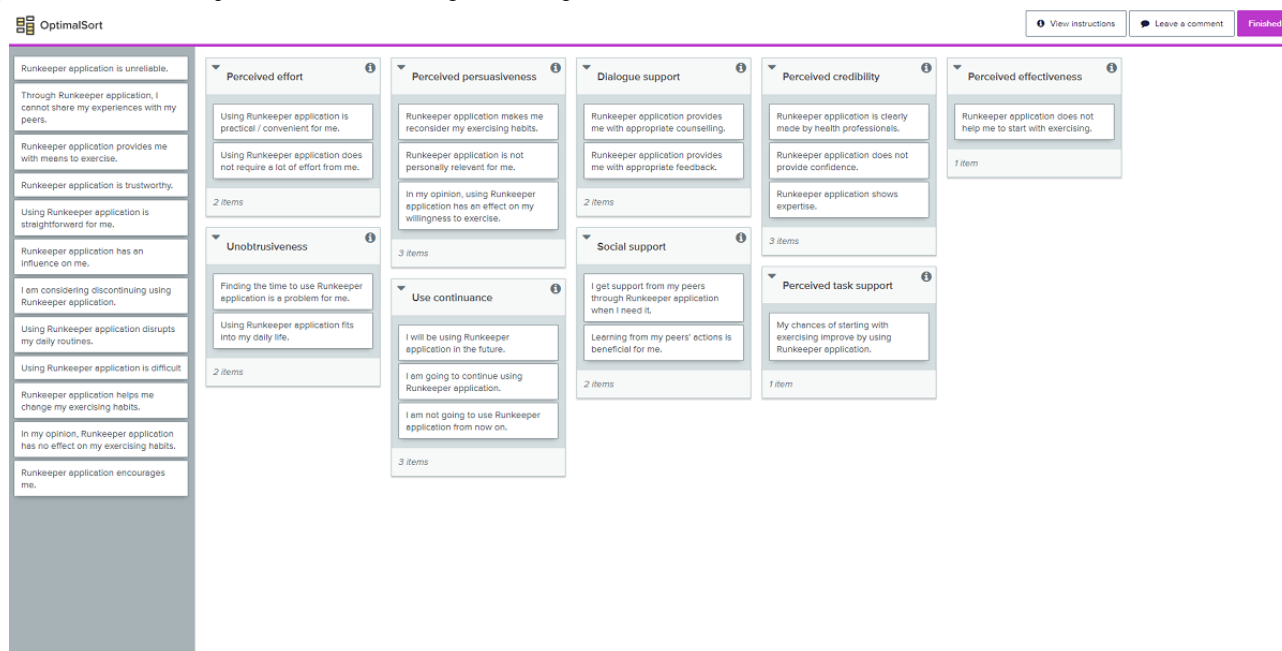
cards by a simple drag-and-drop functionality. Screenshots of the card-sort software used are provided in [Multimedia Appendix 2](#). [Figure 1](#) portrays the instructions page; [Figure 2](#) shows a screenshot during the actual sorting task.

The focal point of this study's card sort was the PPQ. Overall, the PPQ consists of 31 items divided over 9 constructs. The constructs, their meaning, and an example item are provided in [Table 2](#). A complete overview of the PPQ constructs and their items is provided in [Multimedia Appendix 3](#). To make the items easier to interpret and less abstract, for this study, the PPQ items were formulated around an existing, well-known eHealth technology: the Runkeeper app [35]. However, the items can be adapted to fit other persuasive technologies. Participants did not need to actively work with the Runkeeper app; it was merely used as a way to concretize the PPQ constructs. For example, it allowed a card to state "Runkeeper application does not help me to start with exercising" rather than "XYZ does not help me change [target behavior]."

During the card sort, the participants were asked to *drag and drop* all PPQ items onto the PPQ constructs to group them into a meaningful collection of items. They were assured that there were no right or wrong answers and to *just do what comes naturally*. It was explained to them that they could see the definition of each PPQ construct by hovering the mouse over it. During the card sort, the software ensured that participants had to select a single construct to place an item in. Completing the card-sort task took the participants an average of 10.55 min (SD 8.03).

**Figure 1.** Screenshot of card sort software during instructions.



**Figure 2.** Screenshot of OptimalSort software during card sorting task.**Table 2.** Short overview of Perceived Persuasiveness Questionnaire constructs and examples of their items.

PPQ <sup>a</sup> construct (abbreviation)	Short description	No. of items	Example of an item
Primary task support (TASK)	Whether the technology helps to achieve the goal	3	Runkeeper helps me change my exercising habits
Perceived dialogue support (DIAL)	Whether the technology provides feedback and guidance to the user	3	Runkeeper provides me with appropriate counselling
Perceived credibility (CRED)	The perceived reliability and trustworthiness of the technology	5	Runkeeper is clearly made by health professionals
Perceived social support (SOCI)	Whether the technology allows the user to share with and learn from their peers	3	I get support from my peers through Runkeeper when I need it
Perceived persuasiveness (PERS)	Whether users think that the technology is valuable and has an influence on them	3	Runkeeper has an influence on me
Perceived unobtrusiveness (UNOB)	How disturbing the technology is to daily life	4	Using Runkeeper disrupts my daily routines
Perceived effort (EFO)	The endeavor that the technology entails	3	Using Runkeeper is difficult
Perceived effectiveness (EFFE)	The efficacy of the technology	3	My chances of starting with exercising improve by using Runkeeper
Use continuance (CONT)	Willingness of users to adopt the technology in the future	4	I will be using Runkeeper in the future

<sup>a</sup>PPQ: Perceived Persuasiveness Questionnaire.

## Data Analysis

All card-sort data were exported into an Excel spreadsheet (Microsoft) that was specifically developed for the analysis of card sorts [36]. This spreadsheet was extended, adapted, and used in this study and previous research from our research group [24,26,37].

Face validity was evaluated by determining the item-level agreement of the original PPQ constructs. This was done by evaluating how many participants had grouped each individual item in the intended construct by means of descriptive statistics. As a second step for the data analysis, the focus was on construct

validity of the PPQ constructs. It is of interest to see how the items are grouped and placed in constructs by the respondents themselves, according to their own mental models, regardless of the constructs that the items originally belonged to.

Third, we performed a more in-depth analysis by exploring how often any 2 items (regardless of the constructs) were paired together and by measuring interitem correlations (a measure of how consistently the items of that construct are grouped together). To study the items' coherence regardless of the clusters, we performed a cluster analysis of the items. These analyses provide insight into which items were perceived as

measuring a similar construct. We performed a hierarchical cluster analysis (intervals based on squared Euclidian distance) using IBM SPSS 24.

## Results

### Face Validity—Item-Level Agreement in the Original PPQ

The first research question focused on the face validity of the PPQ constructs, revolving around the question of the extent to which the participants' sorts fit the original PPQ constructs. Item-level agreement on these constructs is shown in [Table 3](#).

Some of the original PPQ constructs demonstrated a relatively strong (average >50%) item-level agreement, indicating that most of the participants grouped the *correct* items within these constructs. Agreement about the items of the constructs perceived social support, use continuance, perceived credibility, and perceived effort was high. On the other hand, item-level

agreements for the constructs perceived task support, perceived dialogue support, perceived usefulness, perceived persuasiveness, and perceived effectiveness were relatively low. Agreement on these items was more diffused with respondents frequently mingling with these constructs. Some interesting observations from [Table 3](#) are as follows:

- The items of the perceived task support construct were grouped relatively often into the constructs of perceived persuasiveness and perceived effectiveness.
- The items of perceived dialogue support were grouped relatively often under perceived task support.
- The items of perceived unobtrusiveness were also grouped under perceived effort.
- The items of perceived persuasiveness were also grouped relatively often under perceived effectiveness and perceived task support.
- The items of perceived effectiveness were most frequently also grouped under the construct perceived persuasiveness and perceived task support.



**Table 3.** Item-level agreement in original Perceived Persuasiveness Questionnaire placements.

Original con- struct	Item	TASK <sup>a</sup> (%)	DIAL <sup>b</sup> (%)	CRED <sup>c</sup> (%)	SOCI <sup>d</sup> (%)	UNOB <sup>e</sup> (%)	PERS <sup>f</sup> (%)	EFFO <sup>g</sup> (%)	EFFE <sup>h</sup> (%)	CONT <sup>i</sup> (%)
TASK	24	56	4	2	0	2	11	5	16	4
TASK	15	25	4	3	1	6	15	7	36	4
TASK	23	20	3	2	1	1	22	4	43	4
DIAL	26	22	63	2	3	0	4	1	5	1
DIAL	21	22	54	5	9	1	3	1	5	1
DIAL	11	17	29	1	2	0	33	5	8	5
CRED	29	1	1	90	0	1	4	1	2	1
CRED	6	2	2	61	0	7	4	5	14	6
CRED	4	3	2	88	1	1	3	1	2	1
CRED	16	11	7	46	1	6	12	3	12	3
CRED	17	3	1	89	1	1	3	0	2	0
SOCI	22	3	6	1	88	1	1	0	1	1
SOCI	20	3	8	1	78	4	1	2	1	3
SOCI	30	2	4	1	88	1	1	0	2	1
UNOB	10	5	2	1	1	69	3	8	5	8
UNOB	14	4	1	1	1	64	2	17	4	7
UNOB	1	5	2	2	1	36	6	25	16	8
UNOB	28	3	1	1	1	35	2	39	3	15
PERS	18	5	4	1	1	2	53	3	28	3
PERS	5	10	3	5	1	11	23	6	29	12
PERS	19	17	7	3	0	1	42	3	24	3
EFFO	2	4	2	1	0	15	2	71	3	3
EFFO	7	6	3	3	1	18	5	49	9	6
EFFO	31	7	3	1	1	11	4	65	5	5
EFFE	3	18	2	3	0	2	26	6	38	6
EFFE	12	11	1	2	1	3	42	5	32	4
EFFE	25	8	2	4	0	4	9	5	63	4
CONT	27	1	1	2	0	1	5	3	5	83
CONT	13	2	1	3	0	1	6	2	8	76
CONT	8	2	1	1	1	5	3	4	8	76
CONT	9	2	1	2	0	5	4	6	7	74

<sup>a</sup>TASK: primary task support.<sup>b</sup>DIAL: perceived dialogue support.<sup>c</sup>CRED: perceived credibility.<sup>d</sup>SOCI: perceived social support.<sup>e</sup>UNOB: perceived unobtrusiveness.<sup>f</sup>PERS: perceived persuasiveness.<sup>g</sup>EFFO: perceived effort.<sup>h</sup>EFFE: perceived effectiveness.<sup>i</sup>CONT: use continuance.

### Optimal Item-Level Agreement

As a second step, focusing on construct validity, we calculated in what constructs the items were most often placed regardless

of the original placements of the items. An overview of the average agreement within the constructs for the original constructs versus the constructs as they should be according to popular placements is given in [Table 4](#).

**Table 4.** Overview of the agreement within the original constructs versus popular placement constructs.

Construct	Average agreement within the original PPQ <sup>a</sup> constructs	Average agreement within constructs as defined by popular placement
Perceived task support	33.7	73.7
Perceived dialogue support	48.7	73.7
Perceived credibility	74.8	74.8
Perceived social support	84.7	84.7
Unobtrusiveness	51.0	56.3
Perceived persuasiveness	39.3	42.5
Perceived effort	61.7	56.0
Perceived effectiveness	44.3	41.8
Use continuance	77.3	77.3

<sup>a</sup>PPQ: Perceived Persuasiveness Questionnaire.

According to popular placement ([Table 5](#)), the items of the constructs perceived task support and perceived dialogue support should be merged into a single category with a high average agreement (>70%). This merged category showed much higher agreement than the items within the separate constructs according to the original PPQ. Existing constructs with high agreement according to popular placement were perceived credibility support, perceived social support, and use continuance. All these constructs had the same average

agreement as when using the original placement, with their items remaining unchanged. Agreement with popular placement for unobtrusiveness and perceived persuasiveness was somewhat higher than agreement within these constructs of the original PPQ. Simultaneously, agreement with popular placement for perceived effort and perceived effectiveness was slightly lower than agreement within the original PPQ. For all these constructs and for the merged construct, the composition of items would have to change based on popular placements.

**Table 5.** Popular placement matrix.

Item	TASK <sup>a</sup> (%)	DIAL <sup>b</sup> (%)	CRED <sup>c</sup> (%)	SOCI <sup>d</sup> (%)	UNOB <sup>e</sup> (%)	PERS <sup>f</sup> (%)	EFFO <sup>g</sup> (%)	EFFE <sup>h</sup> (%)	CONT <sup>i</sup> (%)
24	56	4	2	0	2	11	5	16	4
26	22	63	2	3	0	4	1	5	1
21	22	54	5	9	1	3	1	5	1
29	1	1	90	0	1	4	1	2	1
17	3	1	89	1	1	3	0	2	0
4	3	2	88	1	1	3	1	2	1
6	2	2	61	0	7	4	5	14	6
16	11	7	46	1	6	12	3	12	3
22	3	6	1	88	1	1	0	1	1
30	2	4	1	88	1	1	0	2	1
20	3	8	1	78	4	1	2	1	3
10	5	2	1	1	69	3	8	5	8
14	4	1	1	1	64	2	17	4	7
1	5	2	2	1	36	6	25	16	8
18	5	4	1	1	2	53	3	28	3
12	11	1	2	1	3	42	5	32	4
19	17	7	3	0	1	42	3	24	3
11	17	29	1	2	0	33	5	8	5
2	4	2	1	0	15	2	71	3	3
31	7	3	1	1	11	4	65	5	5
7	6	3	3	1	18	5	49	9	6
28	3	1	1	1	35	2	39	3	15
25	8	2	4	0	4	9	5	63	4
23	20	3	2	1	1	22	4	43	4
3	18	2	3	0	2	26	6	38	6
15	25	4	3	1	6	15	7	36	4
5	10	3	5	1	11	23	6	29	12
27	1	1	2	0	1	5	3	5	83
8	2	1	1	1	5	3	4	8	76
13	2	1	3	0	1	6	2	8	76
9	2	1	2	0	5	4	6	7	74

<sup>a</sup>TASK: primary task support.<sup>b</sup>DIAL: perceived dialogue support.<sup>c</sup>CRED: perceived credibility.<sup>d</sup>SOCI: perceived social support.<sup>e</sup>UNOB: perceived unobtrusiveness.<sup>f</sup>PERS: perceived persuasiveness.<sup>g</sup>EFFO: perceived effort.<sup>h</sup>EFFE: perceived effectiveness.<sup>i</sup>CONT: use continuance.

### Cluster Analysis of the Items

The results described in the sections *Face Validity—Item-Level Agreement in the Original PPQ* and *Optimal Item-Level*

*Agreement* have indicated that some changes in the PPQ would be advisable based on the mental models of the participants. In the following section, we therefore performed an additional cluster analysis of the items to study their coherence regardless

of the constructs. [Multimedia Appendix 4](#) shows an item-item matrix of correlations. Within the table, we indicate which construct most of the items in each cluster belong to. There are 3 relatively strong clusters that are identical to the original perceived credibility support, perceived social support, and use continuance. However, the placement of the other items was much more spread out over the other clusters ([Multimedia Appendix 4](#)). Some interesting observations are as follows:

- The original perceived effort and perceived unobtrusiveness could still be found unchanged in the item-item matrix. However, the items of perceived effort also showed overlap with 2 other items (items 28 and 1) that were mainly grouped in the latter. Merging the 2 constructs would lead to an interitem correlation of 0.898, which is slightly lower than perceived effort and higher than perceived unobtrusiveness.
- On the basis of cluster analysis, the original construct of perceived dialogue support could be reduced to only 2 items. This change slightly improved the interitem correlation from 0.715 to 0.731.
- The items of the original constructs perceived task support, perceived effectiveness, and perceived persuasiveness showed rather strong overlap and were grouped into a single cluster.
- One item of perceived dialogue support (item 11) had been moved to a left over cluster alongside item 24, which was originally part of perceived task support.

## Discussion

### Principal Findings

This study set out to evaluate the PPQ by means of a card sort of its items to determine whether the current version of the PPQ can be used to assess the overall perceived persuasiveness of a technology according to the mental model of its users. The results show that some constructs within the PPQ seem to have high face and construct validity in their current form (ie, perceived credibility, social support, and use continuance), whereas other constructs (ie, perceived task support, perceived dialogue support, perceived effort, unobtrusiveness, effectiveness, and persuasiveness) have less face validity and overlap with other PPQ constructs. It appears that the PPQ in its current form does not fit the mental models of the participants, which sheds doubt on the usefulness of the scale to assess perceived persuasiveness because of the subjective nature of this concept. The results of this study provide a first step toward a thorough redesign of the PPQ to more robustly and validly measure constructs related to perceived persuasiveness in a manner that fits the perceptions and experiences of its potential users.

First, it is noticeable that the constructs perceived task support and perceived dialogue support show strong overlap. Both constructs are part of the PSD model but differ in their theoretical underpinning, meaning, and effect [16,17]. However, it seems as if the items are not able to distinguish between the constructs. The PPQ should be able to make a distinction

between these constructs as their theoretical aims differ from one another: perceived task support covers the features of the technology that aim to support its user in achieving its goals, whereas perceived dialogue support covers computer-human interactions that are enabled by the technology [16,17]. In line with this, it has been shown that the effect of including elements of perceived task support and perceived dialogue support in technologies is different; for example, perceived dialogue support has been shown to be related to adherence, although this relationship was not found for perceived task support. Thus, the items within these constructs need to be redefined to enable adequate measurement of these distinct constructs. When looking at the items, it is apparent that the conceptual level of the items in the different constructs varies. For perceived task support, the items are aimed at the overall goal of the concept (eg, “Runkeeper helps me change my exercising habits”), whereas for perceived dialogue support, the items are framed on separate elements within the concept of perceived dialogue support (eg, “Runkeeper provides me with appropriate counseling”). It might be that the focus on separate elements in perceived dialogue support does not sufficiently address the overall goal of the category, in this case supporting the dialogue between the system and the user. Therefore, we suggest redefining all items of the PPQ such that they represent a similar level of abstraction, focusing on the purpose of the construct rather than on how it might have been applied in a technology.

Second, the constructs perceived effort and perceived unobtrusiveness also show strong overlap. Ideally, the PPQ should include all constructs that are part of a technology’s persuasiveness. However, it should omit constructs that seem to predict rather than being a part of persuasiveness. Perceived effort revolves around how strenuous it is to use the technology, regardless of its context. In addition, perceived effort is seen to be part of usability, which is a precondition for persuasion but not persuasion itself [16]. Therefore, we argue that perceived effort should not be a part of the PPQ. Perceived unobtrusiveness, on the other hand, concerns the extent to which the technology can be used as a *seamless part of daily routines* [16]. From many previous studies, we know that the success of a technology not only depends on the technology itself but also on its fit with the use context [38,39]. A main way of technology to be persuasive is, therefore, to increase the fit of the technology in daily life and by improving the way in which it is integrated into the work or thought process of its users. Therefore, we advocate that perceived unobtrusiveness should remain part of the PPQ and that *unobtrusiveness support* might even be seen as a fifth category of the PSD model.

Third, perceived effectiveness, perceived persuasiveness, and (to a lesser extent) perceived task support display strong overlap. This might be explained by the fact that they all cover the technology’s support toward achieving one’s goals. However, it is noticeable that of these constructs, only perceived task support is part of the PSD model. Furthermore, the construct of perceived task support is focused on how supportive the technology is, whereas perceived effectiveness and perceived persuasion are concerned with the extent to which technology-driven support has actually contributed to a behavior change. As the latter two are more of an outcome measure of

persuasion than a form of persuasion itself, we advocate that they should not be part of the PPQ. For perceived persuasiveness, this is also confirmed in the literature, where perceived persuasiveness was shown to be the result of perceived task support, perceived dialogue support, perceived credibility, and perceived unobtrusiveness. Yet, perceived persuasiveness itself did not influence any variables except the (intended) use of the technology, suggesting that it is an outcome rather than a part of persuasion [16]. This confirms our belief that perceived persuasiveness should merely be seen as an outcome of persuasion, rather than being part of it. Moreover, it seems peculiar to have a construct called perceived persuasiveness as a part of a questionnaire with the same name. This suggests that the other constructs are not part of the same theoretical concept and should thus not be included in the questionnaire.

Fourth, the construct use continuance has and maintains strong face and construct validity. However, following the same reasoning as described above, it is questionable whether this construct is part of perceived persuasion or is more of an outcome or effect of perceiving persuasion. Moreover, we argue that use continuance would be a measure of adherence rather than persuasion and therefore advocate that it should not be part of the PPQ.

### Using a Card-Sort Study to Evaluate the Questionnaire

In this study, we opted to use the card-sort method to investigate the face and construct validity of the PPQ instead of more traditional methods such as interviews and focus groups. A large advantage of the card-sort method is that the data of more people can be combined in a robust and statistical manner, more so than using qualitative methods [40]. In this manner, we could statistically analyze whether participants agreed with the original grouping of the PPQ items and which items were clustered together in new user-centered concepts. A disadvantage of the card-sort method is that the reasoning behind participants' clustering is not taken into account, whereas in a qualitative study, a researcher is able to focus on this more. Therefore, a conscious decision on which method to use in a particular context should be made.

Although the card sort yielded many valuable results in this study, it would also have been useful to use this closed-ended card sort at an earlier stage: during the development of the PPQ, before its release. In this manner, a card sort can be used to verify whether assumptions made by the researchers on the structure of the questionnaire and clustering of the items are in line with the users' mental models. Consequently, card sorting is not merely a suitable method to evaluate existing questionnaires but can also be used as a development tool when creating new questionnaires.

Furthermore, in this study, prospective end users of the PPQ were invited to participate in the card sort study. An advantage of this approach is that the resulting structure of a questionnaire fits the mental models of users, who might lack the necessary expertise to make decisions based on theoretical underpinning of the constructs. Consequently, we recommend combining user-based card sorts with expert-based card sorts to combine both perspectives, resulting in a theory-based and user-centered questionnaire.

Finally, in this study, we used a closed-ended card sort, which means that the *final* categories were presented to the users. For a new and improved version of the PPQ, in a later stage of development, an open-ended card sort might be used, in which participants are asked to cluster items together individually and provide them with a label of their own choice to cross-reference the construct validity.

Overall, card sort is a very promising method for questionnaire construction and evaluation. However, more research that applies this method and reflects on its benefits and barriers is required to study its full potential at every stage of questionnaire development, redesign, and evaluation.

### Future Research

Further steps must be taken to redesign and validate the PPQ. First, an expert evaluation of the concept of perceived persuasiveness should be carried out to assess what subconstructs need to be included in this broad concept and which items can be used to assess these subconstructs. Second, measures such as Cronbach alpha and interitem correlations are relevant for its reliability, and a factor analysis is needed to evaluate its (convergent and divergent) construct validity. Furthermore, the predictive validity of the PPQ for adherence or effectiveness should be investigated.

Following such research, it would be highly relevant to also explore the relatedness of perceived persuasiveness with other factors, such as engagement [41], adherence [42] and the effectiveness of technologies. This would allow future research to address the following relevant questions: To what extent does perceived persuasiveness matter? Does it actually make its users more adherent? Does it actually improve the effects that are achieved through using the technology? Can the PPQ be used at early stages during the developmental stages of eHealth technologies to make them as persuasive as possible?

### Strengths and Limitations

A strength of this study is that the PPQ is assessed from the user perspective and insight is yielded in the mental models of users for this valuable concept. However, the card sort method did not allow us to assess the theoretical content validity of the scale and the subconstructs. The results showed that this is a step that still needs to be taken. However, this study did allow us to explore whether the PPQ holds enough potential value to invest more resources in a more extensive validation process.

The results of this study should be interpreted with caution. First, the participants of this study were no experts in the field of persuasion. Thus, although definitions of the constructs were provided, it is not certain whether they had enough skills to truly grasp the meaning of the PPQ constructs and items. However, the choice to include these nonexperts as participants was a conscious one. After all, the PPQ is intended to be filled in by technology users just like them, not by experts in the field of persuasion.

Simultaneously, because of the highly international nature of the participants, it is not known if language barriers might have influenced the results. However, given the fact that the participants were recruited via an English language MOOC, it



can be assumed that their English language skills are at least sufficient to understand the meaning of written text.

Finally, it is noticeable that the participants were not as consistent in their groupings as they could have been. For example, items 25 and 12 are reversed versions of the same item but are not most frequently grouped together (by only 131/398, 32.9% of the respondents). However, this could also be a symptom of the overlap that participants perceived between different constructs.

## Conclusions

In summary, we believe that the original PPQ in combination with the findings of this study, provide a solid starting point toward a redesigned PPQ that is a true asset to the field of persuasiveness research. It has the potential to contribute to answering the all-important question of *what works when for whom*. However, to be able to achieve this, we advocate that the redesigned PPQ should adhere more closely to what

persuasiveness is and to the mental models of potential end users of technology.

In its current form, the PPQ covers many broad constructs, with items that are very specific and diverse. We suggest altering the items to be less focused on details and specific design features and more focused on what the technology is intended to do per category of the PSD model. In other words, the PPQ should enquire *if* the user thinks anything is done to provide task support, but not *how* this is done exactly. This would also increase the reliability and external validity of PPQ measurements. After all, considering the breadth of technologies and the endless varieties of possible target groups, aims, and contexts of these technologies, it is nearly impossible to cover all possible ways of providing (for example) task support. Therefore, there is a need for a new, more abstract version of the PPQ that closely matches the concept of perceived persuasiveness and fits with the mental models of its users.

---

## Acknowledgments

All authors contributed to the design and execution of the study and data analysis. All authors contributed to writing the manuscript and read and approved its final version for submission.

---

## Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Card sort invitation from the Massive Open Online Course .

[DOCX File , 15 KB - [jmir\\_v22i10e20404\\_app1.docx](#) ]

---

### Multimedia Appendix 2

Screenshots of the card sort procedure.

[DOCX File , 250 KB - [jmir\\_v22i10e20404\\_app2.docx](#) ]

---

### Multimedia Appendix 3

Original Perceived Persuasiveness Questionnaire constructs and associated cards.

[DOCX File , 15 KB - [jmir\\_v22i10e20404\\_app3.docx](#) ]

---

### Multimedia Appendix 4

Item×Item groupings according to cluster analysis (in percentage of respondents).

[DOCX File , 73 KB - [jmir\\_v22i10e20404\\_app4.docx](#) ]

---

## References

1. van Gemert-Pijnen J, Kelders S, Kip H, Sanderman R. Ehealth Research, Theory and Development: A Multi-Disciplinary Approach. Abingdon, UK: Routledge; 2018.
2. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52 [FREE Full text] [doi: [10.2196/jmir.1772](#)] [Medline: [21821503](#)]
3. Fogg B. Persuasive Technology: Using Computers to Change What We Think and Do. New York, USA: Ubiquity; 2002.
4. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from critical interpretive synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: [10.1089/tmj.2011.0062](#)] [Medline: [22381060](#)]
5. Kelders S. Using Log-Data as a Starting Point to Make eHealth More Persuasive. New York, USA: Springer; 2013.

6. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuisen H, van Gemert-Pijnen JE. Key components in ehealth interventions combining self-tracking and persuasive ecoaching to promote a healthier lifestyle: a scoping review. *J Med Internet Res* 2017 Aug 1;19(8):e277 [FREE Full text] [doi: [10.2196/jmir.7288](https://doi.org/10.2196/jmir.7288)] [Medline: [28765103](https://pubmed.ncbi.nlm.nih.gov/28765103/)]
7. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. *J Med Internet Res* 2012 Nov 14;14(6):e152 [FREE Full text] [doi: [10.2196/jmir.2104](https://doi.org/10.2196/jmir.2104)] [Medline: [23151820](https://pubmed.ncbi.nlm.nih.gov/23151820/)]
8. Oinas-Kukkonen H, Harjumaa M. Towards Deeper Understanding of Persuasion in Software and Information Systems. In: First International Conference on Advances in Computer-Human Interaction.: IEEE; 2008 Presented at: ACHI'08; February 10-15, 2008; Sainte Luce, Martinique. [doi: [10.1109/ACHI.2008.31](https://doi.org/10.1109/ACHI.2008.31)]
9. Hamari J, Koivisto J, Pakkanen T. Do Persuasive Technologies Persuade? - A Review of Empirical Studies. In: International Conference on Persuasive Technology. 2014 Presented at: PERSUASIVE'08; May 21-23, 2014; Padua, Italy. [doi: [10.1007/978-3-319-07127-5\\_11](https://doi.org/10.1007/978-3-319-07127-5_11)]
10. Orji R, Mandryk R, Vassileva J. Gender, Age, and Responsiveness to Cialdini's Persuasion Strategies. In: International Conference on Persuasive Technology. 2020 Presented at: CPT'25; August 6, 2015; Chicago, IL.
11. Oyibo K, Orji R, Vassileva J. Investigation of the Influence of Personality Traits on Cialdini's Persuasive Strategies. In: International Conference on Persuasive Technology. 2017 Presented at: PERSUASIVE'17; April 3-6, 2017; Amsterdam, the Netherlands.
12. Smith K, Dennis M, Masthoff J. Personalizing reminders to personality for melanoma self-checking. In: Proceedings of the Conference on User Modeling Adaptation and Personalization. 2016 Presented at: UMAP'16; 29 June - 3 July, 2016; Dublin, Ireland. [doi: [10.1145/2930238.2930254](https://doi.org/10.1145/2930238.2930254)]
13. Beerlage-de Jong JN, Wrede C, van Gemert-Pijnen J, Sieverink F, de Vries P, Van Rompay T. Storyboarding Persuasion to Match Personality Traits. In: 12th International Conference on Persuasive Technology. 2017 Presented at: PERSUASIVE'17; April 4-6, 2017; Amsterdam, the Netherlands.
14. Halko S, Kientz J. Personality and persuasive technology: an exploratory study on health-promoting mobile applications. In: International Conference on Persuasive Technology. 2010 Presented at: PERSUASIVE'10; June 7-10, 2010; Copenhagen, Denmark. [doi: [10.1007/978-3-642-13226-1\\_16](https://doi.org/10.1007/978-3-642-13226-1_16)]
15. De Jong N, Wentzel J, Kelders S, Oinas-Kukkonen H, van Gemert-Pijnen J. Evaluation of perceived persuasiveness constructs by combining user tests and expert assessments. In: Proceedings of the Second International Workshop on Behavior Change Support Systems (BCSS). 2014 Presented at: Persuasive'14; 2014; Padova, Italy.
16. Lehto T, Oinas-Kukkonen H, Drozd F. Factors affecting perceived persuasiveness of a behavior change support system. In: Thirty Third International Conference on Information Systems. 2012 Presented at: CIS'12; 2012; Orlando, USA.
17. Oinas-Kukkonen H, Harjumaa M. Persuasive Systems Design: Key Issues, Process Model, and System Features. In: Thirty Third International Conference on Information Systems. 2009 Presented at: CIAS'09; 2009; Orlando, USA. [doi: [10.17705/1cais.02428](https://doi.org/10.17705/1cais.02428)]
18. Busch M, Schrammel J, Tscheligi M. Personalized persuasive technology? Development and validation of scales for measuring persuadability. In: International Conference on Persuasive Technology. 2013 Presented at: PERSUASIVE'13; 2013; Sydney, Australia. [doi: [10.1007/978-3-642-37157-8\\_6](https://doi.org/10.1007/978-3-642-37157-8_6)]
19. Meschtscherjakov A, Gärtner M, Mirnig A, Rödel C, Tscheligi M. Challenges, drawbacks, and lessons learned. In: International Conference on Persuasive Technology. The persuasive potential questionnaire (ppq); 2016 Presented at: PERSUASIVE'16; 2016; Salzburg, Austria. [doi: [10.1007/978-3-319-31510-2\\_14](https://doi.org/10.1007/978-3-319-31510-2_14)]
20. Drozd F, Lehto T, Oinas-Kukkonen H. Exploring Perceived Persuasiveness of a Behavior Change Support System: A Structural Model. New York, USA: Springer; 2012.
21. Kantharaju RB, De Franco D, Pease A, Pelachaud C. Is Two Better than One? Effects of Multiple Agents on User Persuasion. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. 2018 Presented at: CIVA'18; 2018; Sydney. [doi: [10.1145/3267851.3267890](https://doi.org/10.1145/3267851.3267890)]
22. Lehto T, Oinas-Kukkonen H. Explaining and predicting perceived effectiveness and use continuance intention of a behaviour change support system for weight loss. *Behav Inf Technol* 2014 Jan 13;34(2):176-189. [doi: [10.1080/0144929x.2013.866162](https://doi.org/10.1080/0144929x.2013.866162)]
23. Wentzel J, van Drie-Pierik R, Nijdam L, Geesing J, Sanderma R, van Gemert-Pijnen JE. Antibiotic information application offers nurses quick support. *Am J Infect Control* 2016 Jun 1;44(6):677-684. [doi: [10.1016/j.ajic.2015.12.038](https://doi.org/10.1016/j.ajic.2015.12.038)] [Medline: [26905792](https://pubmed.ncbi.nlm.nih.gov/26905792/)]
24. Beerlage-de Jong N, Kulyk O, Kuonanoja L, Wentzel J, Oinas-Kukkonen H, van Gemert-Pijnen J. Evaluation of the Perceived Persuasiveness Questionnaire. Enschede, the Netherlands: University of Twente; 2016.
25. Maguire M. Methods to support human-centred design. *Int J Hum-Comput Stud* 2001 Oct;55(4):587-634. [doi: [10.1006/jjhc.2001.0503](https://doi.org/10.1006/jjhc.2001.0503)]
26. Wentzel J, Müller F, Beerlage-de Jong N, van Gemert-Pijnen J. Card sorting to evaluate the robustness of the information architecture of a protocol website. *Int J Med Inform* 2016 Feb;86:71-81. [doi: [10.1016/j.ijmedinf.2015.12.003](https://doi.org/10.1016/j.ijmedinf.2015.12.003)] [Medline: [26725698](https://pubmed.ncbi.nlm.nih.gov/26725698/)]
27. Rosenfeld L, Morville P. Information Architecture for the World Wide Web. New York, USA: O'Reilly Media Inc; 2002.

28. Beerlage-de Jong N, Wentzel M, Kelders S, Oinas-Kukkonen H, van Gemert-Pijnen J. Perceived persuasiveness questionnaire; Evaluation of Perceived Persuasiveness Constructs by Combining User Tests and Expert Assessments. In: Proceedings of the International conference on Persuasive Technology 2014. 2014 Presented at: PERSUASIVE'14; 2014; Padua, Italy. [doi: [10.2196/preprints.20404](https://doi.org/10.2196/preprints.20404)]
29. Holden R. The Corsini Encyclopedia of Psychology, Face Validity. Hoboken, New Jersey: Wiley; 2010.
30. Construct validity in psychological tests. Psychol Bull 1955 Jul;52(4):281-302. [doi: [10.1037/h0040957](https://doi.org/10.1037/h0040957)] [Medline: [13245896](https://pubmed.ncbi.nlm.nih.gov/13245896/)]
31. Wentzel J, van Velsen L, van Limburg M, de Jong N, Karreman J, Hendrix R, et al. Participatory eHealth development to support nurses in antimicrobial stewardship. BMC Med Inform Decis Mak 2014 Jun 5;14:45 [FREE Full text] [doi: [10.1186/1472-6947-14-45](https://doi.org/10.1186/1472-6947-14-45)] [Medline: [24898694](https://pubmed.ncbi.nlm.nih.gov/24898694/)]
32. Kelders S, Sieverink F, Kip H, van Gemert-Pijnen J. MOOC on eHealth: Combining Psychology, Technology and Health. Future Learn. URL: [https://www.futurelearn.com/courses/ehealth?utm\\_campaign=twente\\_ehealth\\_may\\_2018](https://www.futurelearn.com/courses/ehealth?utm_campaign=twente_ehealth_may_2018) [accessed 2020-10-01]
33. FutureLearn. URL: <https://www.futurelearn.com/> [accessed 2020-10-01]
34. Optimal Workshop. URL: <https://www.optimalworkshop.com/> [accessed 2020-10-01]
35. RunKeeper. 6.6 ed. 2016. URL: <https://runkeeper.com/cms/> [accessed 2020-10-01]
36. Spencer D. Instructions for Use: Card Sort Analysis Spreadsheet. Maad Mob Interaction Design. 2007. URL: [http://www.rosenfeldmedia.com/books/downloads/cardsorting/cardsort\\_analysis.pdf](http://www.rosenfeldmedia.com/books/downloads/cardsorting/cardsort_analysis.pdf) [accessed 2020-10-01]
37. Beerlage-de Jong N, Kulyk O, Wentzel M, Oinas-Kukkonen H, van Gemert-Pijnen J. 'Sorting out' the PPQ: A mixed-methods approach to evaluate Perceived Persuasiveness Questionnaire constructs. In: 10th International Conference on Persuasive Technology. 2015 Presented at: PERSUASIVE'15; 2015; Chicago, USA. [doi: [10.4324/9781315385907-11](https://doi.org/10.4324/9781315385907-11)]
38. Beerlage-de Jong N, Eikelenboom-Boskamp A, Voss A, Sanderma R, van Gemert-Pijnen J. Combining user-centered design with the persuasive systems design model; the development process of a web-based registration and monitoring system for healthcare-associated infections in nursing homes. Int J Adv Life Sci 2014;6:262-271.
39. Laurie J, Blandford A. Making time for mindfulness. Int J Med Inform 2016 Dec;96:38-50. [doi: [10.1016/j.ijmedinf.2016.02.010](https://doi.org/10.1016/j.ijmedinf.2016.02.010)] [Medline: [26965526](https://pubmed.ncbi.nlm.nih.gov/26965526/)]
40. Faiks A, Hyland N. Gaining user insight: a case study illustrating the card sort technique. CRL 2000 Jul 1;61(4):349-357. [doi: [10.5860/crl.61.4.349](https://doi.org/10.5860/crl.61.4.349)]
41. Kelders S, Kip H. Development and initial validation of a scale to measure engagement with eHealth technologies. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI'19; 2019; Glasgow, UK. [doi: [10.1145/3290607.3312917](https://doi.org/10.1145/3290607.3312917)]
42. Sieverink F, Kelders SM, van Gemert-Pijnen J. Clarifying the concept of adherence to ehealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 6;19(12):e402 [FREE Full text] [doi: [10.2196/jmir.8578](https://doi.org/10.2196/jmir.8578)] [Medline: [29212630](https://pubmed.ncbi.nlm.nih.gov/29212630/)]

## Abbreviations

**MOOC:** Massive Open Online Course  
**PPQ:** Perceived Persuasiveness Questionnaire  
**PSD:** persuasive systems design

*Edited by G Eysenbach; submitted 18.05.20; peer-reviewed by E Toki, D Bossen; comments to author 23.06.20; revised version received 02.07.20; accepted 27.07.20; published 23.10.20.*

### *Please cite as:*

Beerlage-de Jong N, Kip H, Kelders SM  
Evaluation of the Perceived Persuasiveness Questionnaire: User-Centered Card-Sort Study  
J Med Internet Res 2020;22(10):e20404  
URL: <http://www.jmir.org/2020/10/e20404/>  
doi:[10.2196/20404](https://doi.org/10.2196/20404)  
PMID:[33095173](https://pubmed.ncbi.nlm.nih.gov/33095173/)

©Nienke Beerlage-de Jong, Hanneke Kip, Saskia Marion Kelders. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org/>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Exergaming With Beat Saber: An Investigation of Virtual Reality Aftereffects

Ancret Szpak<sup>1</sup>, PhD; Stefan Carlo Michalski<sup>1</sup>, BPsych(Hons); Tobias Loetscher<sup>1</sup>, PhD

University of South Australia, Adelaide, Australia

**Corresponding Author:**

Ancret Szpak, PhD

University of South Australia

St Bernards Road, Magill

South Australia

Adelaide, 5072

Australia

Phone: 61 08 8302 ext 4336

Email: [ancret.szpak@unisa.edu.au](mailto:ancret.szpak@unisa.edu.au)

## Abstract

**Background:** Virtual reality (VR) exergaming has the potential to target sedentary behavior. Immersive environments can distract users from the physical exertion of exercise and can motivate them to continue exergaming. Despite the recent surge in VR popularity, numerous users still experience VR sickness from using head-mounted displays (HMDs). Apart from the commonly assessed self-reported symptoms, depth perception and cognition may also be affected. Considering the potential benefits of VR exergaming, it is crucial to identify the adverse effects limiting its potential and continued uptake.

**Objective:** This study aims to investigate the consequences of playing one of the most popular VR exergames for 10 and 50 min on aspects of vision, cognition, and self-reported VR sickness.

**Methods:** A total of 36 participants played an exergame, called Beat Saber, using an HMD. A repeated measures within-subject design was conducted to assess changes in vision, cognition, and well-being after short (10 min) and long (50 min) durations of VR exposure. We measured accommodation, convergence, decision speed, movement speed, and self-reported sickness at 3 test periods—before VR, immediately after VR, and 40 min after VR (late).

**Results:** Beat Saber was well tolerated, as there were no dropouts due to sickness. For most participants, any immediate aftereffects were short-lived and returned to baseline levels after 40 min of exiting VR. For both short and long exposures, there were changes in accommodation ( $F_{1,35}=8.424$ ;  $P=.006$ ) and convergence ( $F_{1,35}=7.826$ ;  $P=.008$ ); however, in the late test period, participants returned to baseline levels. Measures on cognition revealed no concern. The total simulator sickness questionnaire (SSQ) scores increased immediately after VR ( $F_{1,35}=26.515$ ;  $P<.001$ ) and were significantly higher for long compared with short exposures ( $t_{35}=2.807$ ;  $P=.03$ ), but there were no differences in exposure duration in the late test period, with scores returning to baseline levels. Although at a group level, participants' sickness levels returned to baseline 40 min after VR exposure, approximately 14% of the participants still reported high levels of sickness in the late test period after playing 50 min of Beat Saber. We also showed that the participants who experienced a high level of sickness after a short exposure were almost certain to experience a high level of symptoms after a longer exposure.

**Conclusions:** Irrespective of the duration of exposure, this study found no strong evidence for adverse symptoms 40 min after exiting VR; however, some individuals still reported high levels of VR sickness at this stage. We recommend that users commit to a waiting period after exiting VR to ensure that any aftereffects have deteriorated. Exergames in HMDs have the potential to encourage people to exercise but are understudied, and the aftereffects of exergaming need to be closely monitored to ensure that VR exergames can reach their full potential.

(*J Med Internet Res* 2020;22(10):e19840) doi:[10.2196/19840](https://doi.org/10.2196/19840)

**KEYWORDS**

virtual reality; motion sickness; exercise; sedentary behavior; depth perception



## Introduction

### Background

Exergaming combines exercise and gameplay into a virtual environment with the intention of promoting greater physical activity among users. Exergaming is a valid method for targeting sedentary behaviors in children, adolescents [1], and adults [2,3]. The physiological health benefits of exergames are comparable with exercise such as running and aerobic dancing [1]. Several studies show that some people get greater enjoyment and feel more positive toward exergaming compared with other forms of physical exercise [1,4,5]. Greater enjoyment of exercise and adherence are linked to the psychological experience of flow during physical activities [6]. Exergaming facilitates the flow experience by providing an opportunity to become easily absorbed in the goals and challenges in the game [7,8]. A combination of physiological and psychological benefits of physical activities makes exergaming an appealing strategy to encourage sedentary people to exercise.

A potentially promising avenue to boost the effectiveness of exergaming is virtual reality (VR). One main reason why exergaming is so successful is that it can distract the user from the physical exertion of exercise [4,9]. Users become deeply involved in the game (flow state) and are more motivated to continue engaging in the tasks and narrative of the game even if it requires physical and mental effort. When it comes to involvement and absorption, VR excels in this domain. By offering a combination of realistic 3D environments, immersive 360° spaces, and body tracking, these features allow users to feel like they exist in the virtual world and are ready to face all the challenges and experiences it affords. Typically, presence is not a prerequisite for exergames but can provide added benefits [9]. However, in VR, presence is the cornerstone of a successful virtual experience, with greater presence leading to better task performance and stronger physiological responses compared with other screen-based activities [10-12]. Similarly, immersion and presence in VR exergames play a major role in the motivation of users and their continued engagement [9].

Reviews of VR exergames are an important indicator of user attitudes and receptivity of using VR experiences as a motivator to exercise. A recent study [9] examined reviews of popular VR exergames to identify the positive and negative factors contributing to their continued engagement. Generally, users said that they enjoyed immersive exergames that distracted them from the intensity of the exercise and felt that VR games, such as Beat Saber, provided a similar level of physical activity they would typically get from exercising in the real world [9]. Negative reviews identified *motion sickness* as a hindrance to their in-game performance and perceived enjoyment [9].

*Motion sickness*, or VR sickness, has plagued VR since its early days. While the VR industry has seen a surge in accessibility, greater reliability, and lower prices even for casual consumers, many users are still affected by symptoms of VR sickness [13-15]. High levels of VR sickness symptoms such as nausea, disorientation, and visual disturbances have been estimated to result in an average dropout rate of 15.6% [14]. VR sickness has been shown to decrease or break presence [10,11], impact

motivation and enjoyment [9,16], and influence task performance [17-19]. Apart from the commonly assessed self-reported symptoms, depth perception and cognition may also be affected when users experience VR sickness [15,20,21]; however, little is known about these aftereffects. Considering the potential benefits of serious VR gaming, it is crucial to better understand the adverse effects that limit its potential and continued uptake.

There is no consensus on the etiology of VR sickness [22-24]. One prominent theory is the mismatch between virtual and real worlds. A visual-vestibular conflict can be introduced in VR when the visual experience does not match the physical or bodily experience [13,25]. The integration of visual and vestibular (ie, bodily) sensations plays a fundamental role in an individual's ability to move about and interact with their environment [26,27]. If there is a conflict between visual and vestibular sensations relayed back to the brain, an individual may experience a disturbance in sensory integration leading to the occurrence of motion sickness symptoms such as nausea or disorientation. An individual may also experience an increase in oculomotor symptoms resulting from vergence-accommodation conflicts from HMDs [15,20,28]. Vergence and accommodation are essential oculomotor functions that facilitate the accurate use of depth cues [15]. It is unclear whether vergence-accommodation conflicts are responsible for VR sickness in certain individuals or whether it compounds the severity of sickness symptoms.

Considering the factors that induce VR aftereffects, it is clear that both content and device characteristics play a key role in the onset and progression of VR sickness. To address the role of content, the Oculus store (Facebook Technologies LLC, 2012) includes comfort ratings for most VR games, which are either rated *comfortable*, *moderate*, or *intense* to warn users of potential adverse side effects [29]. These ratings are based on the amount of camera movement, user motion, and occasional disorienting content that can be perceived under a typical experience. However, these ratings do not factor in the length of the experience and how longer exposures may contribute to VR aftereffects [14,30]. Similarly, device manufacturers of HMDs address VR sickness or side effects in their instruction manuals and make safety recommendations for VR usage [31-33]. These safety guidelines are reliant on the self-assessment of their users to gauge the symptoms of VR sickness. Recommendations for VR usage durations range from 30 min to hour-long exposures, and some devices do not specify but leave it up to the user to decide [31-33]. As VR is highly immersive, users can easily lose track of time, resulting in longer exposure times. Research shows that longer exposures lead to a higher prevalence of serious self-reported symptoms [14,30]. Publicly available research supporting manufacturer recommendations for VR exposure durations is limited.

Research investigating the impact of VR exposure duration primarily focuses on self-reported symptoms and shows that the length of time spent in VR is a critical factor in the development and severity of aftereffects [14,34,35]. However, a recent review suggests that the relationship between VR content and exposure duration may not be straightforward [14]. The authors found that lower self-reported symptoms were



recorded for exposures less than 10 min compared with longer VR studies between 10 and 20 min. Intriguingly, studies that were longer than 20 min on average reported less severe symptoms than studies between 10 and 20 min. The authors suggested that the distribution of the types of VR content (ie, 360° videos, game, minimalist, scenic) in each of the time categories may have contributed to this nonlinear pattern of results [14]. There are few studies that directly compare aftereffects in the same content with shorter and longer exposure durations. Hence, it is difficult to answer the question of whether users who experience short exposures to a particular VR content will experience similar or worse symptoms for longer exposures with the same content.

Another important safety issue to consider is the duration of aftereffects after exiting VR and the impact the duration of aftereffects may have on recovery. There is a gradual degradation of VR sickness symptoms, and it is often recommended that users experiencing symptoms stop VR and wait until they have recovered [32,36]. The duration of these symptoms may vary depending on the time spent in VR and the initial severity of symptoms [34,37]. A recent review suggests that although there is an increase in symptoms with exposure durations, the persistence of aftereffects varies considerably from short periods of time (10 min) to longer periods (4 hours) [34]. Furthermore, recovery from aftereffects may take longer if a user experiences severe symptoms [34].

## Objectives

With the growth in VR exergaming and entertainment, an increasing number of people will likely experience symptoms from longer exposures. Using a popular VR exergame, this study addresses the influence of exposure durations on a user's well-being, and on aspects of vision and cognition. One of the most successful commercial VR exergames is called Beat Saber, with over 2 million copies sold worldwide [38]. According to the Virtual Reality Institute of Health and Exercise [39], the time spent playing Beat Saber is comparable with the energy spent playing tennis in the real world. Therefore, Beat Saber provides a compelling test case for studying the aftereffects of VR exergaming. This study examined VR aftereffects from exergaming through long (50 min) and short (10 min) exposure durations. For both exposure durations, measures of near-point convergence and accommodation, reaction times, and self-reported symptoms were taken before, immediately after VR, and 40 min after VR (late measurements).

## Methods

### Participants

A power analysis [40] was performed using a medium effect size (0.54) from the average effect sizes for vision and cognitive tests in Szpak et al [11]. Accordingly, the power analysis for a one-sample *t* test (difference from constant) with  $\alpha=.05$  and  $1-\beta=.80$  suggested that a minimum sample of 29 participants was required. We tested all participants who signed up when the study was advertised, which resulted in a total of 44 participants.

A total of 44 English-speaking participants were recruited and provided informed consent for participation. Participants were reimbursed AUD \$20 (USD \$14.6) per hour of participation. One participant withdrew (*male*) from the study, which was not due to VR sickness, and another 7 participants (*male*=5; *female*=2) were excluded because of stereoacuity of 100 arcseconds or worse. The remaining 36 participants (*male* 21; *female*=15) were included in the main analyses (mean age 20.55, SD 2.29 years). Of these 36 participants, 17 (47%) self-reported to play computer/console games on a weekly or daily basis, 50% played on a monthly basis or less, and 1 person did not specify. The average total Motion Sickness Susceptibility Questionnaire Short-form (MSSQ-short) score was relatively low at 8.97 (SD 6.14) compared with other studies and norm data [25,26]. The Human Research Ethics Committee at the University of South Australia granted ethics approval for this study.

### Materials and Apparatus

#### Virtual Reality Setup

A commercially available HTC Vive Pro HMD was used to administer a VR rhythm exergame—*Beat Saber* (developed by Beat Games). Beat Saber was selected because it is a best-selling exergame with a large user base and it offers a high-quality, responsive, and enjoyable game that participants could engage with for at least 60 min. A high-end laptop with an Intel Quad-Core i7-7820HK processor at 2.90 GHz, 16 GB RAM, and an Nvidia GeForce GTX 1080 8 GB graphics card, which ensures that participants experience the game at optimal performance. Using motion tracking, Beat Saber simulates handheld controllers as light sabers, whereby users slash targets and must actively avoid incoming obstacles to the rhythm of the beat. The game provided haptic, auditory, and performance feedback, thereby giving participants an immersive experience.

#### Visual Measures

Stereo vision, near vision, and distance vision were measured to screen participants' visual and stereoacuity. The Snellen [41] and Fonda-Anderson [42] charts were used to assess distance vision and near vision, respectively. The Butterfly Stereo Acuity test (Vision Assessment Corporation, 2007) was utilized to ensure participants could see the virtual environment correctly. Furthermore, accommodation and vergence were measured to investigate changes in participants' depth perception and vision. The Royal Air Force (RAF) near-point rule [43] was used to assess the near point of convergence and the near point of accommodation before and after VR exposure. The RAF near-point rule is composed of a 500-mm ruler-like square tube with a slider attachment bracketing a 4-sided rotating cuboid. In this study, we used only 2 of the 4 sides: the Times Roman typeface to measure accommodation and a small black dot to measure convergence. At one end of the RAF rule, there is a plastic 60-mm V-shaped cheek rest to comfortably sit on a participant's cheek and fit around his or her nose [43]. Participants' accommodation and vergence measurements were measured in millimeters.

### Cognitive Measures

The CANTAB 5-choice reaction time task (RTI) was administered on an iPad 2 using the CANTAB app [44]. The CANTAB version of the 5-choice RTI focuses on measuring participants' speeded responses so that movement and cognitive factors are dissociable. The 5-choice RTI requires a participant to monitor 5 locations. The RTI consists of a circle (button) on the lower half of the screen and 5 circles on the top of the screen. The participant must press the button located at the bottom of the screen and wait for a yellow dot to appear in any of the 5 circles on the top of the screen. When a yellow dot appears, the participant must release the button and touch the yellow dot (target stimulus) as quickly as possible. The reaction time for this task comprises 2 components: decision and movement speed. In this study, decision speed is the median duration from the time the target stimulus appeared to the moment the participant released the button. The movement speed is the median duration of the release of the button to the touch of the target stimulus. Only correct responses were used to calculate these components.

### Self-Report Questionnaires

The MSSQ-short was employed to measure how susceptible participants are to motion sickness [45,46]. The MSSQ-short [46] has a high correlation with the long version ( $r=0.93$ ) [45,46] and has an internal consistency of  $\alpha=.91$  [46].

The Simulator Sickness Questionnaire (SSQ) [47] is the most commonly used questionnaire in simulator and VR studies [14,25] and has a good internal consistency  $\alpha=.87$  [48]. The SSQ was used in this study to measure self-reported symptoms of VR sickness [47,49]. The SSQ comprises a 16-symptom inventory with a 4-point rating scale from 0 (none) to 3 (severe). Each symptom cluster was divided into 3 categories: nausea, oculomotor, and disorientation. The nausea cluster comprises 7 symptoms associated with feelings of stomach sickness, such as increased salivation, burping, and stomach awareness. The oculomotor cluster consists of 7 symptoms related to eyestrain,

fatigue, and focus. The disorientation cluster includes 7 symptoms related to dizziness and vertigo. The 3 subscales include overlapping symptoms from the other subscales. Raw scores are weighted differently for total scores and the subscales [12,38,39].

### Procedure

First, each participant was given verbal instructions and guided through the consent process. All participants were screened to ensure that they had normal vision and stereoacuity using the Butterfly Stereo Acuity test, the Snellen chart, and the Fonda-Anderson chart. Participants completed several questionnaires that included questions regarding demographics (eg, age, gender, handedness), gaming experience, vision history, and motion sickness susceptibility (MSSQ-short).

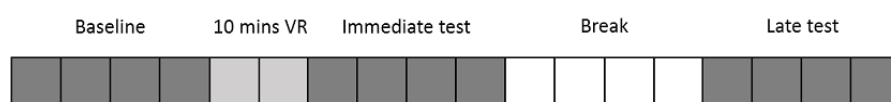
At baseline, participants' accommodation and convergence were measured with the RAF rule, and they also completed the CANTAB RTI on an iPad and paper-based SSQ. Participants were then immersed in a VR exposure that required them to play Beat Saber using the HTC Vive Pro HMD. During this time, participants had either a 10-min or 50-min exposure in the VR. All participants completed both a 10-min and 50-min exposure on 2 separate days. The order of exposure duration was counterbalanced across participants.

Participants were instructed to play the in-game tutorial to understand how the game is played. The tutorial was completed only during the participant's first intervention period. Once the tutorial was completed, the researcher explained that participants had an opportunity for *free play*, that is, they could play the game at their leisure and choose any song and difficulty of their choice.

Immediately after each VR exposure, participants completed the same measures they completed before VR in the following order: accommodation, vergence, RTI, and SSQ. Participants then had a 20-min break and, finally, completed these same measures in a late test period 40 min after VR exposure (Figure 1).

**Figure 1.** The study design of both days of participation. One square represents 5 min. Dark gray squares represent assessment periods, light gray squares represent virtual reality (VR) exposure, and white squares show when participants took a 20-min break. The order of short and long VR exposures on days 1 and 2 was counterbalanced between participants.

#### Short exposure



#### Long exposure



### Analyses

The following analyses used difference scores that were calculated for all measures to demonstrate the change from baseline to immediate or late measurements. All difference

scores were analyzed with the repeated measures analysis of variance with test periods (immediate and late) and exposure durations (10 min and 50 min) as within-subject factors. Bonferroni corrections were employed for all planned pairwise post hoc comparisons. Furthermore, one-sample *t* tests were

performed to determine whether the difference scores significantly changed from zero, that is, represent a significant difference from baseline (Table 1). Alluvial plots were generated from weighted SSQ scores and categorized from a modified version of Kennedy et al [49], resulting in the following 3 sickness levels: low (0-10), mid (>10 to 20), and high (>20).

## Results

Participants' data are available on the Open Science Framework [50].

### Visual Measurements

For accommodation measurements, there was a significant main effect of the test period ( $F_{1,35}=8.424$ ;  $P=.006$ ; partial  $\eta^2=.0.194$ ), with larger accommodation changes immediately after VR (mean 12.43, SE 3.06) compared with the late (mean 4.10, SE 1.92) measurement test period (Figure 2). There was no effect of the exposure duration ( $F_{1,35}=2.974$ ;  $P=.09$ ; partial  $\eta^2=.078$ ), and the interaction was not significant ( $F_{1,35}=0.035$ ;  $P=.85$ ; partial  $\eta^2=.001$ ).

For convergence measurements, there was a significant main effect of test period ( $F_{1,35}=7.826$ ;  $P=.008$ ; partial  $\eta^2=.0.183$ ) with larger convergence changes immediately after VR (mean 13.68, SE 3.65) compared with the late (mean 5.00, SE 2.93)

measurement (Figure 2). There was no effect of the exposure duration ( $F_{1,35}=2.159$ ;  $P=.15$ ; partial  $\eta^2=.0.058$ ), and the interaction was not significant ( $F_{1,35}=1.334$ ;  $P=.26$ ; partial  $\eta^2=.0.037$ ).

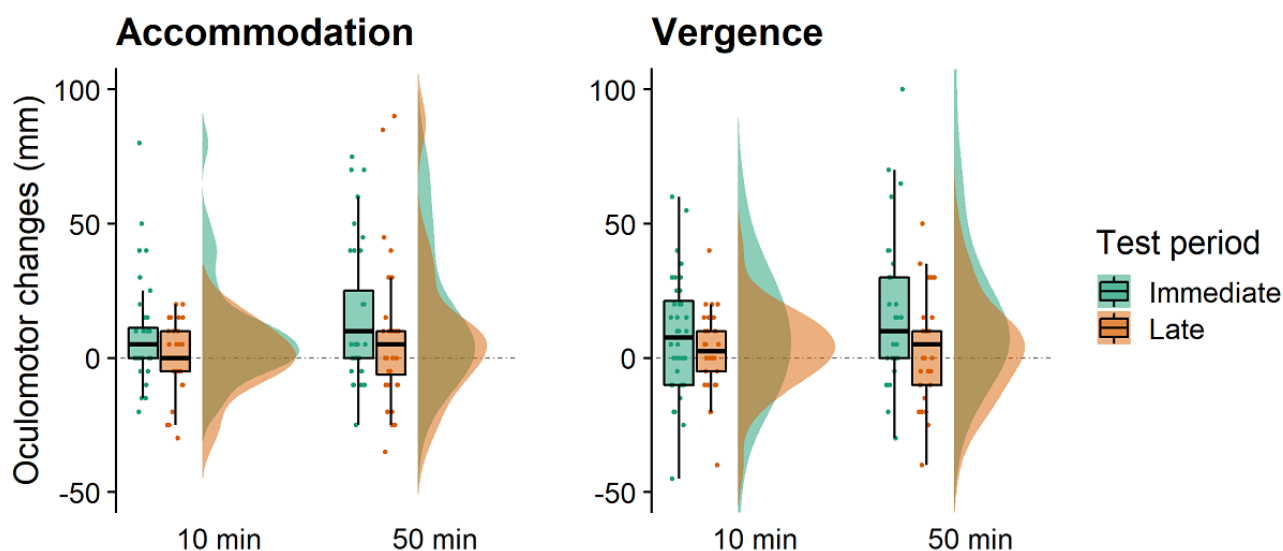
### Cognitive Measurements

The CANTAB 5-choice RTI captured both decision times and movement speeds. Accordingly, these components were analyzed separately to determine which aspect of the reaction speeds may be affected by VR.

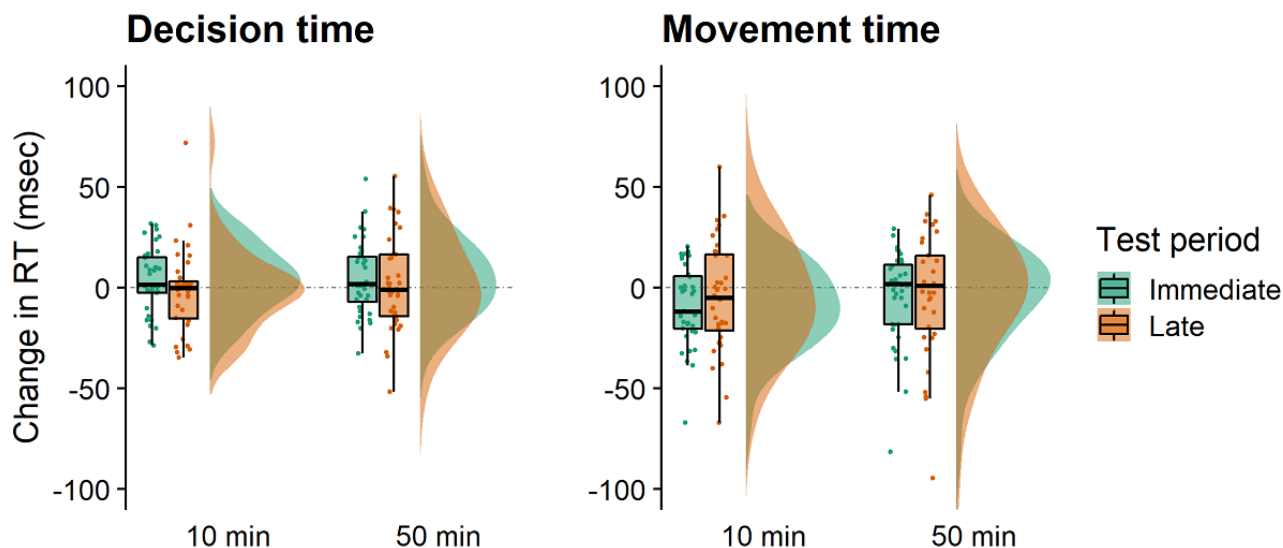
Regarding decision times, there was a significant main effect of the test period ( $F_{1,35}=4.671$ ;  $P=.04$ ; partial  $\eta^2=.118$ ) with larger changes (slower reaction times [RTs]) in the decision speed immediately after VR (mean 4.472, SE 2.229) compared with the late (mean .257, SE 2.638) measurements (Figure 3). There was no effect of the exposure duration ( $F_{1,35}=0.440$ ;  $P=.51$ ; partial  $\eta^2=.0.012$ ), and the interaction was also not significant ( $F_{1,35}=0.668$ ;  $P=.42$ ; partial  $\eta^2=.0.019$ ).

Regarding movement times, there was no main effects of the test period ( $F_{1,35}=1.506$ ;  $P=.23$ ; partial  $\eta^2=.0.041$ ) or the exposure duration ( $F_{1,35}=0.206$ ;  $P=.65$ ; partial  $\eta^2=.0.006$ ), and the interaction was not significant either ( $F_{1,35}=0.784$ ;  $P=.38$ ; partial  $\eta^2=.0.022$ ; Figure 3).

**Figure 2.** Raincloud plots for accommodation (left) and vergence (right) measures showing the different test periods and virtual reality exposure times. Positive and negative scores, respectively, indicate an increase (further) or decrease (nearer) change in accommodation or vergence from baseline measurements.



**Figure 3.** Raincloud plots for the cognitive 5-choice reaction time task showing reaction time (RT) difference scores for the different test periods and virtual reality exposure times. This task captures both decision speeds (left) and movement speeds (right). Positive and negative scores indicate the participant's RTs becoming slower or faster from baseline measurements. RT: reaction time.



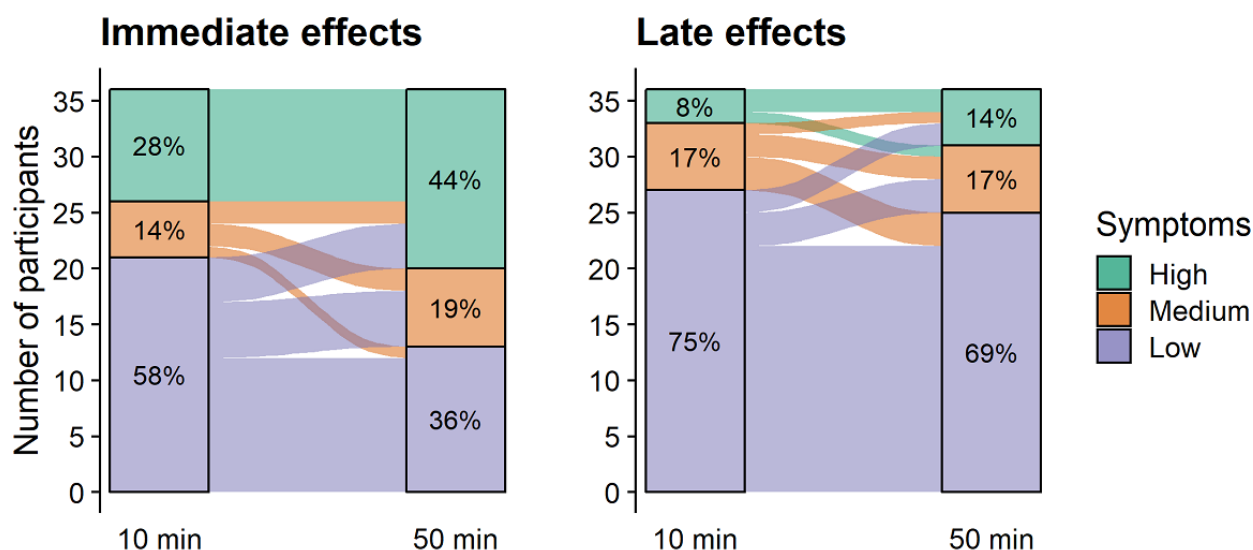
## SSQ

Total SSQ scores and subscales (nausea, oculomotor, and disorientation) were weighted according to Kennedy et al [47]. Difference scores were calculated from these weightings. Alluvial plots were created to visualize changes in SSQ symptom levels in participants across the different exposure times and test periods (Figure 4).

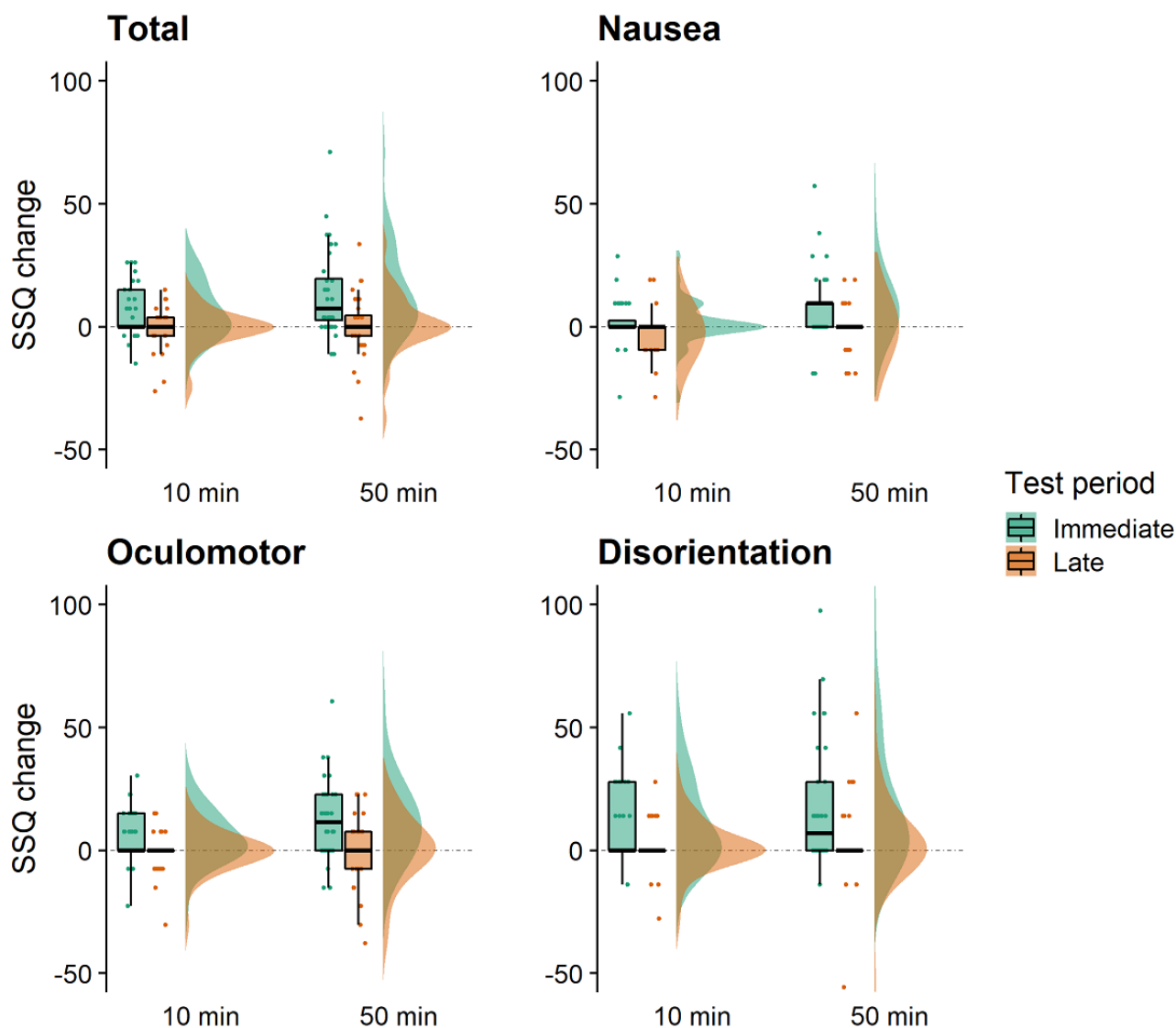
For total SSQ difference scores, there was a significant main effect of the test period ( $F_{1,35}=26.515$ ;  $P<.001$ ; partial  $\eta^2=0.431$ ), with larger changes immediately after VR (mean 9.869, SE 1.952) compared with the late measurement (mean  $-0.052$ , SE 1.272; Figure 5). There was also an effect of exposure ( $F_{1,35}=4.816$ ;  $P=.04$ ; partial  $\eta^2=0.121$ ), with a 50-min VR

exposure (mean 7.220, SE 2.038), leading to larger changes than a 10-min exposure (mean 2.597, SE 1.280). The significant interaction ( $F_{1,35}=4.738$ ;  $P=.04$ ; partial  $\eta^2=0.119$ ) was followed up with a number of post hoc paired  $t$  test comparisons. They showed that for 10 min of VR, immediate measurements (mean 6.026, SE 1.325) were significantly different from late measurements (mean  $-0.831$ , SE 1.325;  $t_{35}=3.868$ ;  $P=.002$ ). Comparisons for 50 min of VR showed that immediate measurements (mean 13.713, SE 2.878) and late measurements (mean 0.727, SE 2.037) were also statistically different ( $t_{35}=4.522$ ;  $P<.001$ ). A comparison of the immediate measurements between the 10 min and 50 min sessions of VR were also significantly different ( $t_{35}=2.807$ ;  $P=.03$ ) from each other. However, the late measurements taken after 10 min and 50 min of VR were not significantly different ( $t_{35}=0.675$ ;  $P=.99$ ).

**Figure 4.** Alluvial plots showing how participants' flow from one virtual reality sickness category to another on the basis of exposure duration and test period. Bars indicate the percentage of participants who are in each of the high, mid, and low virtual reality sickness symptom categories for 10-min and 50-min exposures. Left and right panels show the flow of the categories for the immediate and late test period, respectively.



**Figure 5.** Raincloud plots for the total Simulator Sickness Questionnaire and subscale difference scores showing test periods and exposure duration. Positive and negative scores, respectively, indicate an increase and decrease in sickness symptoms compared with baseline. SSQ: Simulator Sickness Questionnaire.



For nausea difference scores, there was a significant main effect of the test period ( $F_{1,35}=22.379$ ;  $P<.001$ ; partial  $\eta^2=0.390$ ), and the effect of the exposure duration ( $F_{1,35}=4.549$ ;  $P=.04$ ; partial  $\eta^2=0.115$ ) and the interaction was also significant ( $F_{1,35}=4.334$ ;  $P=.045$ ; partial  $\eta^2=0.110$ ; Figure 5). Paired  $t$  test comparisons showed that for 10 min of VR, immediate measurements (mean 1.590, SE 1.496) were significantly different from late measurements (mean  $-2.385$ , SE 1.441;  $t_{35}=2.667$ ;  $P=.048$ ). Comparisons for 50 min of VR showed that immediate (mean 8.480, SE 2.336) and late measurements (mean  $-0.530$ , SE 1.419) were also statistically different ( $t_{35}=4.261$ ;  $P<.001$ ). A comparison of immediate measurements between the 10 min and 50 min sessions of VR were also significantly different ( $t_{35}=2.745$ ;  $P=.04$ ) from each other. However, the late measurements taken after 10 min and 50 min of VR were not significantly different ( $t_{35}=0.827$ ;  $P=.99$ ).

For oculomotor difference scores, there was a significant main effect of the test period ( $F_{1,35}=21.173$ ;  $P<.001$ ; partial  $\eta^2=0.377$ )

and the effect of the exposure duration failed to reach significance ( $F_{1,35}=3.795$ ;  $P=.06$ ; partial  $\eta^2=0.098$ ); however, the interaction was significant ( $F_{1,35}=4.817$ ;  $P=.04$ ; partial  $\eta^2=0.121$ ; Figure 5). Paired  $t$  test comparisons showed that after 10 min of VR, immediate measurements (mean 5.263; SE 1.677) were significantly different from late measurements (mean  $-0.632$ , SE 1.329;  $t_{35}=3.500$ ;  $P=.004$ ). Comparisons for 50 min of VR showed that immediate (mean 12.002, SE 2.587) and late measurements (mean 0.211, SE 2.229) were also statistically different ( $t_{35}=4.128$ ;  $P<.001$ ). A comparison of immediate measurements between the 10 min and 50 min sessions of VR were also significantly different ( $t_{35}=2.652$ ;  $P=.048$ ) from each other. However, the late measurements taken after 10 min and 50 min of VR were not significantly different ( $t_{35}=0.388$ ;  $P=.99$ ).

For disorientation difference scores, there was a significant main effect of the test period ( $F_{1,35}=19.875$ ;  $P<.001$ ; partial  $\eta^2=0.362$ ) but no effect of the exposure duration ( $F_{1,35}=2.577$ ;  $P=.12$ ;



partial  $\eta^2=0.069$ ) and no interaction ( $F_{1,35}=1.628$ ;  $P=.21$ ; partial  $\eta^2=0.044$ ; Figure 5).

In Table 1, a summary of one-sample  $t$  tests is reported to establish whether the difference scores for each measure

significantly changed from zero, that is, represent a significant difference from baseline. Several visual, cognitive, and self-reported measures demonstrated changes from baseline immediately after VR; however, at the group level, all changes returned to preexposure levels at late test periods.

**Table 1.** One-sample  $t$  tests summarizing the measures that are significantly different from zero (baseline) for both test periods (immediate and late) and exposure duration.

Measure and exposure duration	Before to immediately after VR <sup>a</sup>			Before to 40 min (late) after VR		
	$t$ test ( $df$ )	$P$ value	Cohen $d$	$t$ test ( $df$ )	$P$ value	Cohen $d$
<b>Accommodation</b>						
10 min	2.824 (35)	.008	0.471	0.137	.89	0.023
50 min	3.709 (35)	<.001	0.618	1.804	.08	0.301
<b>Convergence</b>						
10 min	2.205 (35)	.03	0.367	1.172	.25	0.195
50 min	3.494 (35)	.001	0.582	1.324	.19	0.221
<b>RT<sup>b</sup> (decision)</b>						
10 min	1.457 (35)	.15	0.243	−0.541	.59	−0.090
50 min	1.649 (35)	.11	0.275	0.602	.55	0.100
<b>RT (movement)</b>						
10 min	−2.725 (35)	.01	−0.454	−1.028	.31	−0.171
50 min	−1.206 (35)	.24	−0.201	−0.714	.48	−0.119
<b>Total SSQ<sup>c</sup></b>						
10 min	3.427 (35)	.002	0.571	−0.627	.54	−0.105
50 min	4.765 (35)	<.001	0.794	0.357	.72	0.060
<b>Nausea</b>						
10 min	1.063 (35)	.295	0.177	−1.655	.11	−0.276
50 min	3.630 (35)	<.001	0.605	−0.373	.71	−0.062
<b>Oculomotor</b>						
10 min	3.140 (35)	.003	0.523	−0.475	.64	−0.079
50 min	4.639 (35)	<.001	0.773	0.094	.93	0.016
<b>Disorientation</b>						
10 min	3.894 (35)	<.001	0.649	1.000	.32	0.167
50 min	4.049 (35)	<.001	0.675	1.136	.26	0.189

<sup>a</sup>VR: virtual reality.

<sup>b</sup>RT: reaction time.

<sup>c</sup>SSQ: simulator sickness questionnaire.

## Discussion

### Summary of Findings

This study aimed to investigate the consequences of playing one of the most popular active VR games for 10 min and 50 min on aspects of vision, cognition, and self-reported VR sickness. There were no dropouts due to sickness in this study. Given that the average dropout rate for a VR study using an HMD is approximately 15.6% [14], this suggests that Beat Saber was well tolerated. A low dropout rate is only one indicator of

sickness, and aspects of vision, cognition, and self-reported sickness are discussed in detail in the following sections.

### Visual Measures

Irrespective of the exposure duration, both accommodation and convergence significantly changed immediately after VR compared with baseline measurements. However, the visual measures returned to baseline levels 40 min after exiting VR, suggesting that changes to accommodation and vergence were relatively short-lived.

Interestingly, the exposure duration did not influence any visual measure, suggesting that observable changes in accommodation and vergence did occur within the first 10 min of VR exposure and did not significantly change for exposures up to 50 min. Changes in convergence and accommodation observed after VR are likely to result from the perception of conflicting depth cues in HMDs. Convergence and accommodation are oculomotor functions necessary to achieve a single and clear focus on near objects, respectively. Hence, blur and disparity are essential retinotopic cues assisting accommodation and convergence to achieve a more precise fixation on a near object [51]. Typically, under natural viewing conditions, vergence and accommodation function together in a feedback loop so that changes in one mechanism will lead to concurrent changes in the other. However, in HMDs, vergence and accommodation may be decoupled [15,28], leading to an uncertainty associated with retinotopic cues for depth perception [52,53] and potentially to a range of concomitant symptoms such as headaches, sore eyes, fatigue, and double vision. Although our participants did not report any concurrent clinical visual impairments, it is worth noting that several participants reported large oculomotor changes in both immediate and late test periods (Figure 2). Large oculomotor changes after VR may influence an individual's depth perception in the real world [53], but the risks associated with these changes are not well understood.

### Cognitive Measures

In this study, we assessed whether the ability to react quickly to stimuli is affected after immersion in VR. The time required to initiate motor movements (ie, decision time) was not statistically different from the decision times before VR exposure at either test period (Figure 3). Notably, the participants were slightly faster at the late test period compared with the immediate test period. However, given that the decision times at either of these test periods were not different from their baseline measures (Table 1), this finding remains to be an interesting observation at this stage.

The investigation of the motor movement times (ie, time from button release to the touch of stimuli) revealed no concerns. In fact, after playing Beat Saber for 10 min, participants' movement speeds immediately after VR exposure were slightly faster than before the VR exposure (Figure 3). Whether this improvement is related to practicing fast-paced motor movements required in the game remains to be established. The fact that the observed positive effects were short-lived ties in with research in the field of aerobic exercise, which found that effects of exercise on reaction times disappear quickly after exercise cessation [54]. This study cannot answer the question about the possibility of positive long-term cognitive effects with repeated VR gaming. However, there is research showing that gamers have greater attentional control compared with nongamers [55]. Furthermore, frequent gamers are better at filtering out distractors and have been shown to outperform nongamers on a range of perceptual and attentional tasks [55,56]. Considering the differences between screen-based gaming and VR gaming, it would be interesting to explore whether frequent VR gamers also exhibit a similar attentional advantage.

The literature on the immediate effects of VR exposure on the ability to react quickly to stimuli is highly inconsistent. Some researchers find negative reaction time aftereffects [19,57-59], and others show positive (faster) effects on reaction times [60]. The type of content, the time in VR, and the method of RT measurements are key reasons for the inconsistent results. Notably, although there is inconsistency across studies, the reported positive or negative effects of VR exposure on RT are typically under 50 ms. The degree to which such relatively small changes would have any real-world consequences for activities such as driving is unclear. In simulated driving studies, researchers suggest that the average braking time needs to be between 700 ms and 1200 ms to reduce the negative impact of a collision [61].

### Self-Reported VR Sickness

Self-reported VR sickness scores were higher immediately after VR than 40 min after exiting VR. Immediately after VR, participants reported more nausea, oculomotor, and disorientation symptoms compared with preexposure levels. The reported symptoms returned to preexposure levels during the late testing period.

Longer VR exposure (ie, 50 min vs 10 min) led to more symptoms immediately after VR (Figure 5). The exception was the disorientation subscale, which was not modulated by the exposure duration. An increase in nausea symptoms was observed after 50 min but not after 10 min of VR exposure (Table 1). For all other self-reported measures, 10 min of VR was sufficient to observe an increase in symptoms. All symptoms returned to baseline levels after 40 min of exiting VR (Table 1). During the late test period, the number of reported symptoms was no longer modulated by the duration of the VR exposure.

Although an increase in SSQ scores for longer exposure durations is consistent with the VR literature, few studies have examined VR sickness in HMDs, which makes it difficult to evaluate the role of intense exercise in reporting symptoms. However, considering that there is an overlap between symptoms of VR sickness and symptoms of intense cardio exercise (ie, fatigue, disorientation, sweating, and in some cases nausea), it can be challenging to identify VR sickness during an intense workout in VR. Perhaps more research with VR exergames is needed to better understand the relationship between the exposure duration, VR sickness, and the intensity of physical activity.

### Does the Lack of Prolonged Aftereffects Mean There Is Nothing to Be Concerned About?

Overall, this study found no strong evidence for adverse symptoms of concern 40 min after VR exposure, irrespective of whether people played Beat Saber for 10 or 50 min. However, our findings should *not* be taken as evidence for a clean bill of health for playing VR exergames. Primarily, our participants comprised a young and healthy student population with a below-average history of motion sickness. Hence, it is unclear whether the current findings will hold for the elderly and people with high susceptibility to motion sickness. It is important to keep in mind that the lack of difference between baseline scores

and 40 min after VR is based on group averages. Closer inspection of individual data (Figures 2 and 5) reveals that individual participants still report adverse symptoms that are higher than their baseline scores. Case in point, approximately 1 out of 7 participants (approximately 14%) may still score *high* on the SSQ 40 min after playing Beat Saber for 50 min (Figure 4). The *high* category in this study is based on the study by Kennedy et al [49], who proposed that scores over 20 are indicative of a problem simulator. However, the degree to which a high SSQ score impacts everyday activities is unclear. The link between scores on the SSQ and real-world performance decrements is not understood and constitutes a critical gap in knowledge for advancing the safe use of VR.

We did not follow up with participants after the late measurements and did not know how symptoms developed after the experimental session ended. It is important to note that the peak of symptoms in an individual is not always observed in or immediately after VR. Some users may experience severe latent symptoms occurring up to 24 h later [30,36,62]. There are many challenges with monitoring participants' hours after a VR experiment such as how long should one follow up with participants, what is the best method for measuring latent effects, and should symptoms be weighted higher if they occur hours after VR. Furthermore, when following up with participants' hours later, it is more difficult to establish a causal relationship between their symptoms and VR. For example, headaches and fatigue are symptoms of VR sickness but can also occur for many other reasons not related to VR, that is, dehydration, hunger, and exertion, making it difficult to track the progression of symptoms for each individual.

Individual differences such as age, HMD fit, postural stability, and motion sickness susceptibility may contribute to the likelihood of a person experiencing VR sickness in HMDs [14,63,64]. The alluvial plot (Figure 5) shows subgroup patterns on the relationship between the levels of symptoms and the exposure durations immediately after exiting. When experiencing a high level of symptoms after a short 10-min exposure, a participant was almost certain to also experience a high level of symptoms for a longer 50-min exposure on a different day. Around half of the participants experiencing midlevels or no/low symptoms immediately after a 10-min exposure reported the same level of symptoms after longer exposure durations. The other half of the participants in these categories experienced worse symptoms for a longer 50-min exposure. Experiencing symptoms after short periods in VR may be an indicator of whether an individual may experience serious symptoms after longer exposure periods. There is some evidence to suggest that short repeated exposures may reduce a user's experience of VR sickness in subsequent exposures [65-67], but this strategy may not work for everyone. Additionally, taking more frequent breaks [31-33] could be explored as a strategy to mitigate symptoms.

## Limitations

We used the currently most popular VR exergame; however, to what extent these findings hold for other games is unknown. One major challenge when comparing aftereffects from different VR experiences is the role content has on the progression and

severity of symptoms. For example, several studies have compared sickness symptoms across multiple VR experiences and found that some content induces VR sickness and other content leads to no/minimal symptoms [37,59]. High levels of motion have consistently been shown to increase nausea, disorientation, and oculomotor disturbances [21,37,58]. Other factors such as scene complexity [68], presence [10,11], and locomotion [69] have also been suggested to play a role in the development of VR sickness.

In our Beat Saber study, there was a high level of both visual stimulation and user movement, which would have increased with the level difficulty. Participants were able to select any song or level of difficulty. By directing their own gameplay, participants could choose which levels they felt challenged by. A limitation of this study is that we did not monitor in-game performance, which may have been insightful. If the choice of difficulty increased the visual motion in the exergame, this may have also had an impact on the likelihood of a person experiencing symptoms. Future VR studies should consider how the variation in gameplay may impact sickness outcomes and measure in-game performance if possible.

VR exergames that require physical activity from the user will have higher levels of user movement and visual stimulation, which may contribute to VR sickness. A recent VR bike simulator study [57] found that participants using an HMD displayed substantially higher SSQ scores than participants using a large screen. In their study, SSQ scores increased with the exposure duration and simulated motion [57]. Although HMDs may be a more immersive option for exergames, HMDs may also lead to higher levels of sickness relative to screen-based exergames. If the simulated motion is a major factor leading to VR sickness in exergames, then congruent visual and user motion will likely be better tolerated. A call for more research is needed to investigate the relationship between simulated motion and VR sickness in exergames.

A wide range of possible VR aftereffects exist, and this study only investigated some of them (vision, reaction time, and self-reports). It remains to be established whether or how other symptoms, such as changes in balance, depth perception, and motion drowsiness (sopite symptoms), may have also been affected. This study also targeted a young and healthy group of participants, and it is possible that an older sample may experience different symptoms or challenges when exergaming. There are no clear guidelines and thresholds identifying who will be at risk after using provocative VR applications and what to do if users experience atypical symptoms.

## Conclusions

VR sickness in exergames, particularly in HMDs, is understudied. On the basis of the data in this study, we can make two suggestions for the safe use of VR. First, we recommend that users trial a brief session in VR before committing to longer exposures. Our research shows that if a user experiences a high level of symptoms after a short exposure, they he or she will likely experience similar or worse symptoms for longer exposures. Users with high symptoms for shorter exposures may attempt to take frequent breaks and habituation strategies but should still be cautious as these approaches may not work

for everyone. Second, we recommend that users commit to a waiting period after exiting VR. During this waiting period, users should withdraw from activities that may pose a risk to injury in the event that a person has VR aftereffects, for example, driving a car. In our study, a 40-min wait period was sufficient for most people's symptoms to return to baseline levels. Since content plays a major role in the development of

VR sickness, more research is needed to clarify the relationship between VR sickness, the exposure duration, and the minimum waiting times for different exergames before participants should return to activities that pose an increased risk of injury and accident. Given the increasing popularity of VR exergames and the potential implications of VR aftereffects, it is essential that research in this area continues to propagate.

## Acknowledgments

AS and TL worked together in the conceptualization of this research and wrote the manuscript. SM was involved in data collection and assisted in editing the final drafts of the manuscript. TL was supported by the National Health and Medical Research Council Dementia Research Leadership Fellowship (GNT1136269).

## Conflicts of Interest

None declared.

## References

1. Gao Z, Chen S, Pasco D, Pope Z. A meta-analysis of active video games on health outcomes among children and adolescents. *Obes Rev* 2015 Sep;16(9):783-794. [doi: [10.1111/obr.12287](https://doi.org/10.1111/obr.12287)] [Medline: [25943852](https://pubmed.ncbi.nlm.nih.gov/25943852/)]
2. Huang H, Wong M, Lu J, Huang W, Teng C. Can using exergames improve physical fitness? A 12-week randomized controlled trial. *Computers in Human Behavior* 2017 May;70:310-316. [doi: [10.1016/j.chb.2016.12.086](https://doi.org/10.1016/j.chb.2016.12.086)]
3. Warburton DE, Bredin SS, Horita LT, Zbogor D, Scott JM, Esch BT, et al. The health benefits of interactive video game exercise. *Appl Physiol Nutr Metab* 2007 Aug;32(4):655-663. [doi: [10.1139/H07-038](https://doi.org/10.1139/H07-038)] [Medline: [17622279](https://pubmed.ncbi.nlm.nih.gov/17622279/)]
4. Glen K, Eston R, Loetscher T, Parfitt G. Exergaming: feels good despite working harder. *PLoS One* 2017;12(10):e0186526 [FREE Full text] [doi: [10.1371/journal.pone.0186526](https://doi.org/10.1371/journal.pone.0186526)] [Medline: [29059227](https://pubmed.ncbi.nlm.nih.gov/29059227/)]
5. Monedero J, Lyons EJ, O'Gorman DJ. Interactive video game cycling leads to higher energy expenditure and is more enjoyable than conventional exercise in adults. *PLoS One* 2015;10(3):e0118470 [FREE Full text] [doi: [10.1371/journal.pone.0118470](https://doi.org/10.1371/journal.pone.0118470)] [Medline: [25738290](https://pubmed.ncbi.nlm.nih.gov/25738290/)]
6. Vlachopoulos SP, Karageorghis CI, Terry PC. Hierarchical confirmatory factor analysis of the flow state scale in exercise. *J Sports Sci* 2000 Oct;18(10):815-823. [doi: [10.1080/026404100419874](https://doi.org/10.1080/026404100419874)] [Medline: [11055817](https://pubmed.ncbi.nlm.nih.gov/11055817/)]
7. Sinclair J, Hingston P, Masek M. Considerations for the Design of Exergames. In: Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia. 2007 Presented at: GRAPHITE'07; December 1-4, 2007; Perth, Western Australia. [doi: [10.1145/1321261.1321313](https://doi.org/10.1145/1321261.1321313)]
8. Schmidt S, Ehrenbrink P, Weiss B, Voigt-Antons J, Kojic T, Johnston A, et al. Impact of Virtual Environments on Motivation and Engagement During Exergames. In: Tenth International Conference on Quality of Multimedia Experience. 2018 Presented at: QoMEX'18; May 29-June 1, 2018; Cagliari, Italy. [doi: [10.1109/qomex.2018.8463389](https://doi.org/10.1109/qomex.2018.8463389)]
9. Faric N, Potts HW, Hon A, Smith L, Newby K, Steptoe A, et al. What players of virtual reality exercise games want: thematic analysis of web-based reviews. *J Med Internet Res* 2019 Sep 16;21(9):e13833 [FREE Full text] [doi: [10.2196/13833](https://doi.org/10.2196/13833)] [Medline: [31538951](https://pubmed.ncbi.nlm.nih.gov/31538951/)]
10. Weech S, Kenny S, Lenizky M, Barnett-Cowan M. Narrative and gaming experience interact to affect presence and cybersickness in virtual reality. *International Journal of Human-Computer Studies* 2020 Jun;138:102398. [doi: [10.1016/j.ijhcs.2020.102398](https://doi.org/10.1016/j.ijhcs.2020.102398)]
11. Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: a review. *Front Psychol* 2019;10:158 [FREE Full text] [doi: [10.3389/fpsyg.2019.00158](https://doi.org/10.3389/fpsyg.2019.00158)] [Medline: [30778320](https://pubmed.ncbi.nlm.nih.gov/30778320/)]
12. Cooper N, Milella F, Pinto C, Cant I, White M, Meyer G. The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PLoS One* 2018;13(2):e0191846 [FREE Full text] [doi: [10.1371/journal.pone.0191846](https://doi.org/10.1371/journal.pone.0191846)] [Medline: [29390023](https://pubmed.ncbi.nlm.nih.gov/29390023/)]
13. Lawson BD. Motion sickness symptomatology origins. In: Hale KS, Stanney KM, editors. *Handbook of Virtual Environments: Design, Implementation, and Applications*. Boca Raton, FL: CRC Press/Taylor & Francis; 2014:531-600.
14. Saredakis D, Szpak A, Birkhead B, Keage HA, Rizzo A, Loetscher T. Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. *Front Hum Neurosci* 2020;14:96 [FREE Full text] [doi: [10.3389/fnhum.2020.00096](https://doi.org/10.3389/fnhum.2020.00096)] [Medline: [32300295](https://pubmed.ncbi.nlm.nih.gov/32300295/)]
15. Szpak A, Michalski SC, Saredakis D, Chen CS, Loetscher T. Beyond feeling sick: the visual and cognitive aftereffects of virtual reality. *IEEE Access* 2019;7:130883-130892. [doi: [10.1109/access.2019.2940073](https://doi.org/10.1109/access.2019.2940073)]



16. Shafer DM, Carbonara CP, Korpi MF. Factors affecting enjoyment of virtual reality games: a comparison involving consumer-grade virtual reality technology. *Games Health J* 2019 Feb;8(1):15-23. [doi: [10.1089/g4h.2017.0190](https://doi.org/10.1089/g4h.2017.0190)] [Medline: [30199273](https://pubmed.ncbi.nlm.nih.gov/30199273/)]
17. Mittelstaedt JM, Wacker J, Stelling D. VR aftereffect and the relation of cybersickness and cognitive performance. *Virtual Reality* 2018 Nov 10;23(2):143-154. [doi: [10.1007/s10055-018-0370-3](https://doi.org/10.1007/s10055-018-0370-3)]
18. Bittner AC, Gore BF, Hooey BL. Meaningful Assessments of Simulator Performance and Sickness: Can't Have One without the Other? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 1997;41(2):1089-1093. [doi: [10.1177/107118139704100280](https://doi.org/10.1177/107118139704100280)]
19. Matsangas P, McCauley ME, Becker W. The effect of mild motion sickness and sopite syndrome on multitasking cognitive performance. *Hum Factors* 2014 Sep;56(6):1124-1135. [doi: [10.1177/0018720814522484](https://doi.org/10.1177/0018720814522484)] [Medline: [25277021](https://pubmed.ncbi.nlm.nih.gov/25277021/)]
20. Hoffman DM, Girshick AR, Akeley K, Banks MS. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J Vis* 2008 Mar 28;8(3):33.1-3330 [FREE Full text] [doi: [10.1167/8.3.33](https://doi.org/10.1167/8.3.33)] [Medline: [18484839](https://pubmed.ncbi.nlm.nih.gov/18484839/)]
21. Kim J, Kane D, Banks MS. The rate of change of vergence-accommodation conflict affects visual discomfort. *Vision Res* 2014 Dec;105:159-165 [FREE Full text] [doi: [10.1016/j.visres.2014.10.021](https://doi.org/10.1016/j.visres.2014.10.021)] [Medline: [25448713](https://pubmed.ncbi.nlm.nih.gov/25448713/)]
22. Previc FH. Intravestibular balance and motion sickness. *Aerosp Med Hum Perform* 2018 Feb 1;89(2):130-140. [doi: [10.3357/AMHP.4946.2018](https://doi.org/10.3357/AMHP.4946.2018)] [Medline: [29463358](https://pubmed.ncbi.nlm.nih.gov/29463358/)]
23. Stanney KM, Kennedy RS, Drexler JM. Cybersickness is Not Simulator Sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 1997;41(2):1138-1142. [doi: [10.1177/107118139704100292](https://doi.org/10.1177/107118139704100292)]
24. Gavvani AM, Walker F, Hodgson D, Nalivaiko E. A comparative study of cybersickness during exposure to virtual reality and 'classic' motion sickness: are they different? *J Appl Physiol* (1985) 2018 Oct 4:- epub ahead of print. [doi: [10.1152/jappphysiol.00338.2018](https://doi.org/10.1152/jappphysiol.00338.2018)] [Medline: [30284516](https://pubmed.ncbi.nlm.nih.gov/30284516/)]
25. Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. *Virtual Reality* 2016 Apr 26;20(2):101-125. [doi: [10.1007/s10055-016-0285-9](https://doi.org/10.1007/s10055-016-0285-9)]
26. deAngelis GC, Angelaki DE. Visual-vestibular integration for self-motion perception. In: Murray MM, Wallace MT, editors. *The Neural Bases of Multisensory Processes*. Boca Raton, FL: CRC Press/Taylor & Francis; 2012.
27. Gallagher M, Ferrè ER. Cybersickness: a multisensory integration perspective. *Multisens Res* 2018 Jan 1;31(7):645-674. [doi: [10.1163/22134808-20181293](https://doi.org/10.1163/22134808-20181293)] [Medline: [31264611](https://pubmed.ncbi.nlm.nih.gov/31264611/)]
28. Mon-Williams M, Wann JP, Rushton S. Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. *Ophthalmic Physiol Opt* 1993 Oct;13(4):387-391. [doi: [10.1111/j.1475-1313.1993.tb00496.x](https://doi.org/10.1111/j.1475-1313.1993.tb00496.x)] [Medline: [8278192](https://pubmed.ncbi.nlm.nih.gov/8278192/)]
29. Viewing Comfort Rating for Oculus Store Content. Oculus. URL: <https://support.oculus.com/1639053389725739/> [accessed 2020-04-08]
30. Stanney KM, Hale KS, Nahmens I, Kennedy RS. What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. *Hum Factors* 2003;45(3):504-520. [doi: [10.1518/hfes.45.3.504.27254](https://doi.org/10.1518/hfes.45.3.504.27254)] [Medline: [14702999](https://pubmed.ncbi.nlm.nih.gov/14702999/)]
31. Legal Documents. Facebook Technologies LLC. URL: <https://www.oculus.com/legal/health-and-safety-warnings/> [accessed 2020-04-30]
32. Playstation VR Instruction Manual. PlayStation. URL: <https://www.playstation.com/en-us/support/manuals/psvr/> [accessed 2020-04-08]
33. VIVE Pro Safety and Regulatory Guide. HTC Vive. URL: <https://www.vive.com/au/support/vive-pro/> [accessed 2020-04-08]
34. Dužmańska N, Strojny P, Strojny A. Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Front Psychol* 2018;9:2132 [FREE Full text] [doi: [10.3389/fpsyg.2018.02132](https://doi.org/10.3389/fpsyg.2018.02132)] [Medline: [30459688](https://pubmed.ncbi.nlm.nih.gov/30459688/)]
35. Kourtesis P, Collina S, Doumas LA, MacPherson SE. Validation of the virtual reality neuroscience questionnaire: maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology. *Front Hum Neurosci* 2019;13:417 [FREE Full text] [doi: [10.3389/fnhum.2019.00417](https://doi.org/10.3389/fnhum.2019.00417)] [Medline: [31849627](https://pubmed.ncbi.nlm.nih.gov/31849627/)]
36. Champney RK, Stanney KM, Hash PA, Malone LC, Kennedy RS, Compton DE. Recovery from virtual environment exposure: expected time course of symptoms and potential readaptation strategies. *Hum Factors* 2007 Jun;49(3):491-506. [doi: [10.1518/001872007X200120](https://doi.org/10.1518/001872007X200120)] [Medline: [17552312](https://pubmed.ncbi.nlm.nih.gov/17552312/)]
37. Guna J, Geršak G, Humar I, Song J, Drnovšek J, Pogačnik M. Influence of video content type on users' virtual reality sickness perception and physiological response. *Future Generation Computer Systems* 2019 Feb;91:263-276. [doi: [10.1016/j.future.2018.08.049](https://doi.org/10.1016/j.future.2018.08.049)]
38. Lang B. Road to VR - Virtual Reality News. URL: <https://www.roadtovr.com/beat-saber-2-million-units-sales-revenue/> [accessed 2020-04-15]
39. Rating Notes: Beat Saber. The Virtual Reality Institute of Health and Exercise (VR Health Institute). URL: <https://vrhealth.institute/portfolio/beat-saber/> [accessed 2020-04-08]
40. Faul F, Erdfelder E, Lang A, Buchner A. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007 May;39(2):175-191. [doi: [10.3758/bf03193146](https://doi.org/10.3758/bf03193146)] [Medline: [17695343](https://pubmed.ncbi.nlm.nih.gov/17695343/)]
41. Hetherington R. The Shellen chart as a test of visual acuity. *Psychol. Forsch* 1954;24(4):349-357. [doi: [10.1007/bf00422033](https://doi.org/10.1007/bf00422033)]



42. Fonda G, Anderson M. Fonda-Anderson reading chart for normal and low vision. *Ann Ophthalmol* 1988 Apr;20(4):136-139. [Medline: [3377383](#)]
43. Sharma IP. RAF near point rule for near point of convergence—a short review. *Ann Eye Sci* 2018;2(16):1-6 [FREE Full text] [doi: [10.21037/aes.2017.02.05](#)]
44. Reaction Time (RTI). Cambridge Cognition. URL: <https://www.cambridgecognition.com/cantab/cognitive-tests/attention/reaction-time-rti/> [accessed 2020-04-17]
45. Golding JF. Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain Res Bull* 1998 Nov 15;47(5):507-516. [doi: [10.1016/s0361-9230\(98\)00091-4](#)] [Medline: [10052582](#)]
46. Lamb S, Kwok KC. MSSQ-short norms may underestimate highly susceptible individuals: updating the MSSQ-short norms. *Hum Factors* 2015 Jun;57(4):622-633. [doi: [10.1177/0018720814555862](#)] [Medline: [25977322](#)]
47. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int J Aviat Psychol* 1993 Jul;3(3):203-220. [doi: [10.1207/s15327108ijap0303\\_3](#)]
48. Bouchard S, Robillard G, Renaud P. Revising the factor structure of the simulator sickness questionnaire. *Ann Rev Cyberther Telemed* 2007;5(Summer):137.
49. Kennedy RS, Drexler JM, Compton DE, Stanney KM, Lanham DS, Harm DL. Configural scoring of simulator sickness, cybersickness, and space adaptation syndrome: similarities and differences. In: Hettinger LJ, Haas MW, editors. *Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers; 2003.
50. Szpak A, Michalski SC, Loetscher T. Exergaming with Beat Saber: An Investigation of Virtual Reality Aftereffects. OSF Home. URL: <https://osf.io/dpmqw/> [accessed 2020-09-30]
51. Wann JP, White AD, Wilkie RM, Culmer PR, Lodge PA, Mon-Williams M. Measurement of visual aftereffects following virtual environment exposure: implications for minimally invasive surgery. In: Hale KS, Stanney KM, editors. *Handbook of Virtual Environments: Design, Implementation, and Applications*. Boca Raton, FL: CRC Press/Taylor & Francis; 2014:811-833.
52. Tresilian JR, Mon-Williams M, Kelly BM. Increasing confidence in vergence as a cue to distance. *Proc Biol Sci* 1999 Jan 7;266(1414):39-44 [FREE Full text] [doi: [10.1098/rspb.1999.0601](#)] [Medline: [10081157](#)]
53. Harris DJ, Buckingham G, Wilson MR, Vine SJ. Virtually the same? How impaired sensory information in virtual reality may disrupt vision for action. *Exp Brain Res* 2019 Nov;237(11):2761-2766 [FREE Full text] [doi: [10.1007/s00221-019-05642-8](#)] [Medline: [31485708](#)]
54. Audiffren M, Tomporowski PD, Zagrodnik J. Acute aerobic exercise and information processing: energizing motor processes during a choice reaction time task. *Acta Psychol (Amst)* 2008 Nov 15;129(3):410-419 [FREE Full text] [doi: [10.1016/j.actpsy.2008.09.006](#)] [Medline: [18930445](#)]
55. Bavelier D, Achtman R, Mani M, Föcker J. Neural bases of selective attention in action video game players. *Vision Res* 2012 May 15;61:132-143 [FREE Full text] [doi: [10.1016/j.visres.2011.08.007](#)] [Medline: [21864560](#)]
56. Föcker J, Mortazavi M, Khoe W, Hillyard S, Bavelier D. Neural correlates of enhanced visual attentional control in action video game players: an event-related potential study. *J Cogn Neurosci* 2019 Mar;31(3):377-389. [doi: [10.1162/jocn\\_a\\_01230](#)] [Medline: [29308981](#)]
57. Mittelstaedt J, Wacker J, Stelling D. Effects of display type and motion control on cybersickness in a virtual bike simulator. *Displays* 2018 Jan;51:43-50. [doi: [10.1016/j.displa.2018.01.002](#)]
58. Nesbitt K, Davis S, Blackmore K, Nalivaiko E. Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays* 2017 Jul;48:1-8. [doi: [10.1016/j.displa.2017.01.002](#)]
59. Nalivaiko E, Davis SL, Blackmore KL, Vakulin A, Nesbitt KV. Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiol Behav* 2015 Nov 1;151:583-590. [doi: [10.1016/j.physbeh.2015.08.043](#)] [Medline: [26340855](#)]
60. Smith SP, Burd EL. Response activation and inhibition after exposure to virtual reality. *Array* 2019 Sep;3:100010. [doi: [10.1016/j.array.2019.100010](#)]
61. Doecke S, Grant A, Anderson RW. The real-world safety potential of connected vehicle technology. *Traffic Inj Prev* 2015;16 Suppl 1:S31-S35. [doi: [10.1080/15389588.2015.1014551](#)] [Medline: [26027973](#)]
62. Min B, Chung S, Min Y, Sakamoto K. Psychophysiological evaluation of simulator sickness evoked by a graphic simulator. *Appl Ergon* 2004 Nov;35(6):549-556. [doi: [10.1016/j.apergo.2004.06.002](#)] [Medline: [15374762](#)]
63. Arcioni B, Palmisano S, Aporthorp D, Kim J. Postural stability predicts the likelihood of cybersickness in active HMD-based virtual reality. *Displays* 2019 Jul;58:3-11. [doi: [10.1016/j.displa.2018.07.001](#)]
64. Stanney K, Fidopiastis C, Foster L. Virtual reality is sexist: but it does not have to be. *Front Robot AI* 2020 Jan 31;7(4):- . [doi: [10.3389/frobt.2020.00004](#)]
65. Hill KJ, Howarth PA. Habituation to the side effects of immersion in a virtual environment. *Displays* 2000 Mar;21(1):25-30. [doi: [10.1016/s0141-9382\(00\)00029-9](#)]
66. Howarth PA, Hodder SG. Characteristics of habituation to motion in a virtual environment. *Displays* 2008 Mar;29(2):117-123. [doi: [10.1016/j.displa.2007.09.009](#)]

67. Kennedy RS, Stanney KM, Dunlap WP. Duration and Exposure to Virtual Environments: Sickness Curves During and Across Sessions. 2000 Oct;9(5):463-472. [doi: [10.1162/105474600566952](https://doi.org/10.1162/105474600566952)]
68. Pouke M, Tiirio A, LaValle S, Ojala T. Effects of Visual Realism and Moving Detail on Cybersickness. In: Conference on Virtual Reality and 3D User Interfaces. 2018 Presented at: VR'18; March 18-22, 2018; Reutlingen, Germany. [doi: [10.1109/vr.2018.8446078](https://doi.org/10.1109/vr.2018.8446078)]
69. Habgood J, Moore D, Wilson D, Alapont S. Rapid, Continuous Movement Between Nodes as an Accessible Virtual Reality Locomotion Technique. In: Conference on Virtual Reality and 3D User Interfaces. 2018 Presented at: VR'18; March 18-22, 2018; Reutlingen, Germany. [doi: [10.1109/vr.2018.8446130](https://doi.org/10.1109/vr.2018.8446130)]

## Abbreviations

**HMD:** head-mounted display  
**RAF:** Royal Air Force  
**RT:** reaction time  
**RTI:** reaction time task  
**SSQ:** Simulator Sickness Questionnaire  
**VR:** virtual reality

*Edited by G Eysenbach; submitted 04.05.20; peer-reviewed by S Weech, C Johnson; comments to author 16.06.20; revised version received 06.08.20; accepted 06.08.20; published 23.10.20.*

*Please cite as:*

*Szpak A, Michalski SC, Loetscher T*

*Exergaming With Beat Saber: An Investigation of Virtual Reality Aftereffects*

*J Med Internet Res 2020;22(10):e19840*

*URL:* <http://www.jmir.org/2020/10/e19840/>

*doi:* [10.2196/19840](https://doi.org/10.2196/19840)

*PMID:* [33095182](https://pubmed.ncbi.nlm.nih.gov/33095182/)

©Ancrét Szpak, Stefan Carlo Michalski, Tobias Loetscher. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Patients' Convergence of Mass and Interpersonal Communication on an Online Forum: Hybrid Methods Analysis

Remco Sanders<sup>1</sup>, MSc; Theo B Araujo<sup>1</sup>, PhD; Rens Vliegenthart<sup>1</sup>, PhD; Mies C van Eenbergen<sup>2</sup>, MA; Julia C M van Weert<sup>1</sup>, PhD; Annemiek J Linn<sup>1</sup>, PhD

<sup>1</sup>Department of Communication Science, Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

<sup>2</sup>Department of Research, Netherlands Comprehensive Cancer Organisation, Utrecht, Netherlands

**Corresponding Author:**

Remco Sanders, MSc

Department of Communication Science, Amsterdam School of Communication Research

University of Amsterdam

Nieuwe Achtergracht 166

Amsterdam, 1018 WV

Netherlands

Phone: 31 20525 3680

Email: [r.sanders@uva.nl](mailto:r.sanders@uva.nl)

## Abstract

**Background:** Patients are increasingly taking an active role in their health. In doing so, they combine both mass and interpersonal media to gratify their cognitive and affective needs (ie, convergence). Owing to methodological challenges when studying convergence, a detailed view of how patients are using different types of media for needs fulfillment is lacking.

**Objective:** The aim of this study was to obtain insight into the frequency of reported convergence, how convergence affects what posters write online, motives for posting, and the needs posters are trying to fulfill.

**Methods:** Using a hybrid method of content analysis and supervised machine learning, this study used naturally available data to fill this research gap. We analyzed opening posts (N=1708) of an online forum targeting cancer patients and their relatives (Kanker.nl).

**Results:** Nearly one-third of the forum opening posts contained signs of convergence in mass or interpersonal media. Posts containing mass media references disclosed less personal information and were more geared toward community enhancement and sharing experiences compared to posts without convergence. Furthermore, compared to posts without signs of convergence, posts that included interpersonal media references disclosed more personal information, and posters were more likely to ask for the experiences of fellow users to fulfill their needs. Within posts containing signs of convergence, posts including interpersonal media references reported fewer shortages of information, disclosed more information about the disease, and were more active in seeking other posters' experiences compared to posts containing mass media references.

**Conclusions:** The current study highlights the intertwining of media platforms for patients. The insights of this study can be used to adapt the health care system toward a new type of health information-seeking behavior in which one medium is not trusted to fulfill all needs. Instead, providers should incorporate the intertwining of sources by providing patients with reliable websites and forums through which they can fulfill their needs.

(*J Med Internet Res* 2020;22(10):e18303) doi:[10.2196/18303](https://doi.org/10.2196/18303)

**KEYWORDS**

convergence; online health seeking; supervised machine learning; patient needs; machine learning; online forums; patients; media

## Introduction

**Background**

Patients have the need to know and understand (ie, cognitive needs) and the need to feel acknowledged and understood (ie,

affective needs) [1]. Currently, patients take an active role in the management of their health, and in doing so they combine mass and interpersonal communication to gratify their cognitive and affective needs (hereafter referred to collectively as “needs”) [2]. By using mass and interpersonal communication, patients engage in a process that is called “convergence” [3-5].

According to Kreps [3], convergence can be defined as “the sequence of impersonal to interpersonal interactions” (p. 519; type 1 convergence) or the “conduct of interpersonal and peer discussions about health-related issues in virtual discussion spaces of various kinds” (p. 521; type 2 convergence). We adapted and broadened this definition to include convergence between and within mass or interpersonal communication for the current study. That is, we consider convergence as a process in which either one mass communication source and one interpersonal communication source (intermedium convergence) or two mass communication sources (intramedium convergence) are being used to fulfill the user’s needs. For example, patients learn about their disease through a consultation with a medical expert (interpersonal communication) and then validate the advice of the medical expert by visiting a website (mass communication; intermedium convergence) [6-8]. An example of intramedium convergence is the use of treatment experience from a fellow patient’s blog post in one’s own blog post.

These examples show how patients use mass and interpersonal communication to fulfill their needs. However, research into this topic often focuses on one singled-out communication source. As a consequence, current research does not provide insights into how communication sources affect each other and how needs differ depending on the sources used. Answering these questions is important since patients have different communication sources at their disposal. In particular, when it comes to online health information, patients often struggle to understand the complex information online, have difficulties in assessing whether the information is reliable, and might feel overwhelmed or experience information overload [9,10]. Therefore, patients and medical experts should work together in providing, validating, and discussing information. To determine which (online) source fits best, insights are needed on how patients combine sources and how the combination of sources affects needs. Ultimately, part of the costly and limited time of the medical expert could be used for referring to sources that can reliably fulfill part of the patients’ needs.

One explanation for the research gap on how patients combine multiple sources to fulfill their needs can be found in the methodological challenges faced when studying this process. The scarce research in health communication in which both mass and interpersonal communication are taken into account tends to rely on more traditional research methods such as surveys [11,12], interviews [13], and focus groups [14]. These methods can be affected by selection bias, recall bias, and social desirability [11]. By using a hybrid method consisting of content analysis and supervised machine learning (SML), the limitations of these traditional methods can be surmounted [15]. The hybrid method combines a content analysis of real-life communication and SML. We use natural unsolicited data (ie, data that are not obtained as part of research but are instead compiled by the user writing at the time their needs arose). The benefit of this method is that large amounts of already existing natural data can be used to study the current topic.

A good starting point is the analysis of data from online health forums targeting patients. Forums provide a natural database of people’s online activities. In forum opening posts, posters often provide information about their situation at the time of

writing, which often includes previously used communication (ie, signs of convergence), the outcome of this communication effort (ie, motives to start the forum post), and the needs they are trying to fulfill. Additionally, background information about the poster is often included (eg, stage and type of the disease, and whether the poster him/herself or a family member has been diagnosed with the disease) [3,16]. Thus, by using this hybrid method, we are able to gain more insight into how often patients combine mass and interpersonal communication, the reason as to why they engage in convergence, which need they are trying to fulfill, and whether the content of forum posts differs based on the communication sources they used prior to writing the post. The following research aim was central in the present study: What signs of convergence can be detected in forum opening posts, how frequently does convergence occur, what kinds of needs are patients trying to fulfill by engaging in convergence, and how do forum post characteristics (ie, motives, information about the poster, and needs) differ for different kinds of convergence?

In this study, we used forums in the context of cancer. Cancer patients are confronted with many questions and uncertainties during their illness [17,18]. Furthermore, online platforms such as forums and interpersonal communication with a health care provider are the two most important sources of information for cancer patients [19].

## Interpersonal Communication

In general, most patients consider medical experts to be the most trusted source of information [20]. In a review, Shea-Budgell and colleagues [20] highlighted that patients place a high level of trust in medical experts owing to their expertise on topics that patients find most important, including treatment, screening, testing, and detection. Furthermore, their medical and informational training and, to a lesser degree, the emotional support they provide during a consultation are mentioned as factors that instill confidence [21-23]. Nevertheless, between 40% and 90% of patients report unmet needs after their consultation with a medical expert [24,25]. Multiple reasons can be given for these unmet needs. Patient-related reasons include unmentioned concerns, a lack of trust in a particular medical expert, and information overload [26]. Examples of medical expert–related concerns are time constraints and a lack of experience [27]. Therefore, patients also rely on other sources for needs fulfillment, such as online forums. Patients expect their medical experts to discuss the content they found via other sources and to offer their professional take on it [28]. By discussing online health information with their provider, patients engage in intramedium convergence.

## Online Forums

Online forums are often used by patients and their relatives and can be considered as virtual communities. Virtual communities exist in many different areas, cover many topics, and connect groups with a variety of shared characteristics. In this study, we adopt the definition of Rheingold [29], who states that virtual communities are “social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace.” Many patients encounter various



needs and use multiple communication sources to fulfill these needs. As a result, these patients are active in online forums while also often having contact with other sources such as medical experts [30]. Patients use these platforms to gain understanding about their disease but also to connect and exchange experiences and support with others in comparable situations [7,8,31]. By encouraging and enabling active participation (eg, by opening a thread on a topic that is personally relevant), forums have the potential to provide different types of support to the user, such as to receive the support of their peers, to feel empowered through information provision, and to recognize themselves in stories from peers and thereby feel less isolated [11,32]. The malleability of forums in addressing patients' needs and the ability to do this at any time might be a key reason why patients turn to these forums.

### Convergence and Underlying Motives

Many patients decide to combine multiple sources to fulfill their needs. Generally, 25% to 83% of patients search for online health information before or after a consultation with their medical expert [33-36]. Patients seem to use online health information in addition to a consultation to prepare themselves [13], to complement the information given by the medical expert [13,24,25], as well as to validate or challenge the information given by the medical expert [13].

To understand why patients use multiple sources, the optimal matching model can be used [37,38]. This model states that to fulfill patients' needs, these needs should be matched with the right type of support. For example, if the patient feels the need to prepare for a consultation or wants to complement, validate, or challenge the information that is given by the medical expert, this need can be fulfilled by gathering factual information from other sources. By contrast, if the patient feels lost and alone, this need might not be fulfilled by receiving information about the upcoming treatment but rather by receiving emotional support that helps with the emotional aspects of being sick. According to the optimal matching model, patients actively choose the communication channel that they believe has the highest potential to fulfill their current needs. A patient who feels that they should prepare for the consultation is more likely to choose online medical libraries to fulfill these needs, whereas a patient who feels lost will more likely turn to online health forums and blogs, on which interaction with fellow patients is possible [3,31,39-41]. This exploratory research contributes to the optimal matching theory by identifying whether and how patients fulfill their needs by using multiple sources at their disposal and how these sources are intertwined.

### Research Questions

In summary, we believe that forum posts provide a natural database of peoples' communication activities, and these forum posts can offer an opportunity to gain a better understanding of the interplay between communication channels. Therefore, our first research question (RQ1) is proposed as follows: What is the frequency of signs of convergence in forum opening posts? Furthermore, these forum posts can provide a natural registration of the motives for using different media (eg, after seeing a doctor or reading online health information) to fulfill specific needs. Therefore, research question 2 (RQ2) is proposed as follows:

What needs are patients trying to fulfill by opening a forum post? By using a hybrid method to analyze these forums, we are able to also capture other relevant information such as the disclosure of information about the poster and the motive for posting [3,16]. These aspects are important for providing insight into how users in different situations gratify their needs or the needs of relatives by using multiple sources. Therefore, we propose research question 3 (RQ3) as follows: How do motives (3a), information about the poster (3b), and the needs (3c) differ for different types of convergence?

## Methods

### Study Design

We used a hybrid method consisting of a classic social science method (ie, the framework method [42]) and a newer computational social science method (ie, SML). The benefit of this approach is two-fold. First, this method allows us to combine unique features from both approaches. On the one hand, the framework method starts from a theory-based codebook (ie, the use of sensitizing concepts) and is then further developed through an iterative process of (open) coding on a subsample of the data. On the other hand, the coded subsample can then be used to label the whole sample with codes and categories using SML, thereby allowing us to move from open-coded data on a subsample to data that are suitable for quantitative analysis based on the complete dataset, allowing researchers to analyze sample sizes that were impossible to code manually. Second, SML allows us, as well as other researchers and practitioners, to (re)use the trained model on a different dataset or for practical applications. The reuse of the algorithms makes cost-efficient longitudinal research into convergence possible since the models can automatically and consciously be applied to new data.

### Data

We used data retrieved from cancer patients and relatives on the online forum Kanker.nl [43] in the Netherlands. Cancer is the most common disease, with a yearly incidence rate of 439.2 per 100,000 men and women (18,078,567 in 2018) and a yearly rate of 163.5 per 100,000 people dying from cancer (9,555,027 in 2018) [44]. Furthermore, cancer patients and their relatives experience multiple visits with a medical expert and face many questions and uncertainties. In the Netherlands, Kanker.nl is one of the largest and best-known Dutch websites for cancer-related information within an online community [43].

Users are required to register and must provide their name and a valid email address. Participants of all platforms within Kanker.nl give (standard) consent for using their data for research when they register. Ethical approval for the current study was provided by the ethical committee of University of Amsterdam (2016-PC-7547).

For the complete dataset, first, all forum entries (N=9573) were extracted. Second, only the opening posts of the threads were selected (n=1708). The opening posts were chosen because they are most likely to contain a description of the situation and the need the user wants to fulfill. The median number of words for



each thread opening was 608.05 (range 3-20,649). The threads were created between April 2013 and November 2016.

### Phase 1: Framework Method

Of all 1708 thread openings, a random sample of 306 posts (17.92%) was manually coded by the first author (RS) in multiple iterations. First, using two sensitizing concepts derived from the literature (ie, information sources and motives for searching health information), 100 posts were open-coded on signs of convergence (RQ1), motives for opening a forum thread (RQ2), sought-after need (RQ3), and personal characteristics (RQ3); in total, 583 different codes were used for the constructs needed to answer RQ2 and RQ3. Second, these open codes were merged into overarching, latent categories. For example, the distinctions between different medical experts such as general practitioners, oncologists, and nurses were merged into a

medical expert category. Third, the codebook and categories were evaluated on completeness during research meetings with the coauthors (AL, RV, JvW). As a result, several categories were merged again, and more specific categories were added. This process resulted in the following categories: convergence, motive for posting, information on poster, and needs (see [Table 1](#) for the codebook). Fourth, the updated codebook was evaluated by the first author and a trained coder (RS and MB) who double-coded 20% of the sample. The intercoder reliability was good (Lotus range 0.98-1.00); Lotus and standardized Lotus scores are displayed per category in [Table 2](#). Fifth, all 306 exported thread openings were manually coded by the first author (MB) using the codebook. Finally, the manually coded posts were transformed by the first author (MB) into binary variables to be used in the second phase.

**Table 1.** Overview of categories and codes analyzed.

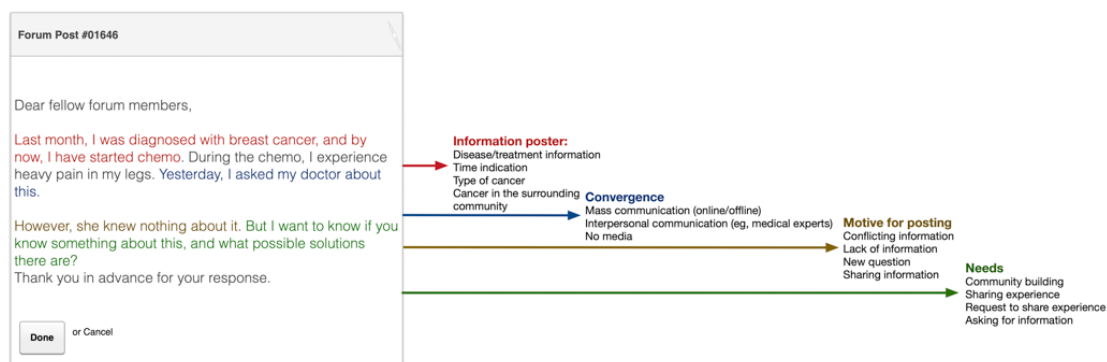
Codes/classifiers	Description	Example
<b>Convergence</b>		
Mass communication	(1) online and (2) offline media. Mass communication was coded as such when communication channels such as internet sources (ie, online) and television and newspapers (ie, offline) were mentioned.	"[...] on a website I read that [...]"; "This article in today's newspaper [...]"
Interpersonal communication	Interpersonal communication was coded as such when one of the following communication sources was mentioned: (1) medical experts (eg, general practitioner, nurse, surgeon), (2) fellow patients, (3) family members, or (4) others.	"My doctor told me that [...]"; "According to my mother [...]"
No media	Posts contained no references to other media.	"I was wondering if any of you knows something about [...]"
<b>Motive for posting</b>		
Conflict	Information is received from a medium that is contradictory to information previously acquired from another medium or contradictory to held beliefs. Resolving this discrepancy is a motive to open a forum thread.	"[...] My doctor told me this treatment is not an option for me, but I heard lots of stories that it was successful [...]"
Shortage of information	Poster indicates that (s)he received little or no information regarding a topic. To fill this information gap, a forum thread is opened.	"[...] There was no time during the consultation to discuss the trajectory of this alternative [...]"
New question	Poster indicates that as a result of information provided during the mentioned communication effort, (s)he has new (follow-up) questions. Answering these questions is a motive for opening a thread.	"[...] the doctor mentioned this medicine can have a lot of side effects, but is it common to experience them?"
Sharing information	Poster wants to share the information/content that was received during the previous communication effort on the forum.	"I read this [website] and thought it might be useful for all of you".
<b>Information on poster</b>		
Disease or treatment information	Specific stages of the disease (eg, stage one) or treatment (eg, after surgery) were described in the post.	"[...] I'm diagnosed with stage one breast cancer"; "[...] After surgery I noticed that [...]"
Time indication disease or treatment	Diseases or treatments that were mentioned at the "disease or treatment information" stage are further specified with a time indicator.	"One year after my surgery, I went back to the hospital [...]"
Type of cancer	Type of cancer is mentioned by the poster.	"I have been diagnosed with lung cancer"
Cancer in the surrounding community	The poster him/herself did not receive a diagnosis but someone close to him or her has.	"My husband has been sick for a few years now, I wonder [...]"
<b>Needs</b>		
Community building	The post is meant as a conversational starter, including rhetorical questions and a direct call for discussion. All without asking for experiences or advice.	"What is your opinion about the quality of care in the Netherlands?"
Sharing experiences	The poster is sharing experiences about the treatment or psychosocial aspects surrounding (living with) cancer.	"For me, this kind of treatment worked very well without too many side effects" or "For me, it worked to limit the number of social activities in a week".
Asking experiences	The poster invites other forum members to share their experiences about a certain topic.	"Who has experience with this?"
Asking for information	The poster asks for more information about a certain topic or asks for referrals to sources where this information can be obtained.	"Who knows where I can find more information about this?"

**Table 2.** Intercoder reliability using Lotus and standardized Lotus (S-Lotus) coefficients per variable.

Concept	Lotus	S-Lotus
<b>Convergence</b>		
Mass communication	1.00	1.00
Interpersonal communication	1.00	1.00
No media	1.00	1.00
<b>Specification of convergence</b>		
Online	1.00	1.00
Offline	1.00	1.00
Medical expert	1.00	1.00
Fellow patients	1.00	1.00
Family members	1.00	1.00
Others	1.00	1.00
<b>Motive for posting</b>		
Conflict	1.00	1.00
Shortage of information	1.00	1.00
New question	0.98	0.96
Sharing information	0.98	0.96
<b>Information on poster</b>		
Disease or treatment information	1.00	1.00
Time indication disease or treatment	1.00	1.00
Type of cancer	0.99	0.97
Cancer in the surrounding community	1.00	1.00
<b>Needs</b>		
Community building	1.00	1.00
Sharing experience	0.99	0.97
Asking experience	1.00	1.00
Asking information	0.99	0.97

## Codebook

Table 1 contains the categories and codes that were coded during the SML phase. These variables were coded as 0 (not present)

**Figure 1.** Example of extraction of concepts from forum posts.

or 1 (present). Figure 1 shows a fictitious example of the extracted concepts from the forum opening posts.

## Phase 2: SML

We used SML to train classifiers for the references to mass or interpersonal communication (see [Multimedia Appendix 1](#) for a detailed description of the SML phase). A sample of 685 manually coded opening posts (in two rounds) was used as input for SML. This sample was split into a training set ( $n=548$ ) and a test set ( $n=137$ ) using an 80-20 split. Using Scikit-Learn [45], the data were preprocessed (see [Multimedia Appendix 1](#) for a detailed overview), and the classifiers were trained using different algorithms such as support vector classification, stochastic gradient descent, multinomial naïve Bayes, gradient boosting, and passive aggressive classifier. This was done to evaluate which algorithm would have the best performance for each concept it was trained to predict. In this process, we adopted a grid search strategy, which tests different combinations of parameters for each algorithm as well as different options for preprocessing the data.

The quality of the classifiers was assessed based on precision, recall, and F1 scores for their predictions of cases in which the category was present (ie, for cases in which the reference to mass or interpersonal communication was 1). Precision gives the proportion of the automatically assigned labels that correspond with the human-labeled data. Recall gives the proportion of the true labels that are found automatically. This often results in a tradeoff between the scores of precision and recall; for example, in cases of higher recall, the chance that some of the recalled data are false positives grows, and the precision score consequently goes down. F1 scores are the harmonic mean of the recall and precision. Stochastic gradient descent proved to have the best performance in predicting the classifiers convergence mass media and convergence interpersonal media (recall<sub>interpersonal</sub>=0.76, precision<sub>interpersonal</sub>=0.96, F1=0.85; recall<sub>massmedia</sub>=0.86, precision<sub>massmedia</sub>=0.92, F1=0.89). See [Multimedia Appendix 2](#) for the complete confusion matrix. These classifiers were applied to the complete dataset of opening posts ( $N=1708$ ) to create a subsample of opening posts that were likely to contain signs of convergence ( $n=771$ , 45.14%).

## Phase 3: Manual Coding Convergence Posts

To ensure the validity of the automatically assigned classifiers, the created subsample of posts was checked for correctness. Of the automatically labeled posts containing signs of convergence ( $n=771$ ), 245 posts (31.78%) did not contain signs of convergence after manually checking, and were coded as “no media.” Next, the subsample of posts was manually coded by the first author (MB) using the codebook for the remaining categories (see [Table 1](#)).

## Analysis

Before running the analysis, all independent variables were tested on possible issues due to multicollinearity. Only issues concerning time since diagnosis, stage of the disease, and type of disease ( $r_{\text{time-stage}}=0.82$ ,  $r_{\text{time-type}}=0.92$ ) were found; thus, these items were taken together as “disclosure of information about the disease” (eigenvalue=2.71;  $R^2=0.90$ ;  $\alpha=.95$ ).

RQ1 and RQ2 were answered using descriptive analyses. To compare the outcomes on the dependent variables between posts containing different signs of convergence (RQ3), two multinomial logistic regression analyses were conducted. In these two analyses, the referenced communication channels were the dependent variables (ie, no media, mass communication, or interpersonal communication). The first regression used no media as the reference category with information on poster and needs as the independent variables. Since the category “motive” was not applicable to the “no mass” or “interpersonal media” category, motive was omitted from this analysis. However, to pinpoint the differences between posts containing signs of convergence, motive, information on poster, and needs were included in the second regression analysis, in which mass communication was used as the reference category. The outcomes of these analyses are displayed as the odds ratios (OR). An  $R$  value of 1 indicates no differences in probability between the groups compared, whereas a value  $>1$  represents an increased probability and a value  $<1$  represents a decreased probability [46].

## Results

### Signs of Convergence in Forum Posts

The results showed that 30.80% of the complete sample of forum opening posts ( $n=526$ ) contained signs of convergence. These were divided into mass communication (324/526, 61.6%) and interpersonal communication (202/526, 38.4%). In the following sections, these categories will be described in more depth.

Of all mass communication references ( $n=324$ ), 274 (84.6%) referred to online sources (eg, other members’ profiles or blogs, news media articles concerning cancer [patients], and health information websites). Of all posts with mass communication references, 214/324 (66.1%) posts contained references to a website. Offline mass communication was referenced 49 (15.1%) times, which included references to printed newspapers, books, and television.

Of all interpersonal communication posts ( $n=202$ ), 162 references (80.2%) were made to medical experts (eg, oncologists, nurses, and general practitioners) and 14 references (6.9%) were made to family members. These include family members who either had personal experiences with the disease or provided information they received via other sources. Fellow patients who provided information offline were referenced 3 times (1.5%) and 22 references (10.9%) were made to communication events with other people. Often, these events consisted of work-related relationships (eg, employers, insurers, and rehabilitation agents).

With respect to RQ1, almost one-third of all forum opening posts contained signs of convergence and thus included references to either mass or interpersonal communication. When referencing mass communication, mainly online sources were mentioned, whereas for interpersonal communication, medical experts were most often mentioned.

## Posters' Needs

Of all 771 opening posts containing signs of convergence, 344 posts (44.6%) represented the need of asking for experiences regarding a specific treatment (eg, medicine, procedure) or experiences regarding (dealing with) the (emotional) effects of living with cancer (ie, dealing with side effects, reintegration into society, and body image). This was followed by forum opening posts related to community building (266/771, 34.5%). In these cases, the poster started a discussion about a particular topic such as developments in the medical sector, with or without a URL to a news story (online). The third-largest need to open a forum thread was to share one's personal experience. Overall, 143/771 (18.6%) of the posts featured this need. Finally, in 72 of the 771 opening posts (9.3%), the poster directly asked for sources to find more factual information on a particular topic such as (alternative) treatment options. Therefore, to answer RQ2, the main need for patients to be fulfilled, as reflected in forum openings post, is that of asking for information related to experiences. This need is followed by that of enhancing the community, sharing one's experiences, and asking for factual information.

## Differences in Posts for Different Kinds of Convergence

The first multinomial logistic regression model contained the variables from the categories information on poster and needs (adjusted  $R^2=0.30$ ,  $\chi^2_{12}=-563.27$ ,  $P<.001$ ; Table 3). In posts referencing mass communication, the disclosure of

disease-related information was 89% less likely to occur compared to posts that did not include a reference to mass communication. In contrast, posts including references to interpersonal communication had a 156% higher likelihood of featuring the disclosure of disease-related information (Table 3) compared to posts containing no references to media. These outcomes mean that the chance of disclosing disease-related information in forum posts in which interpersonal communication is mentioned is higher compared to that of forum posts with no signs of convergence and is lower for posts that include references to mass communication.

When considering the needs posters might have for opening a forum thread, differences in needs within different types of convergence were found. Higher likelihoods were found for posts including references to mass communication compared to posts containing no signs of convergence for the needs: community building, sharing experiences, and asking for information. This means that after mass communication exposure, posts have a 373% higher likelihood of containing the need to share the post for community building, a 291% higher likelihood of containing the need to share one's experience with others, and a 188% higher likelihood of asking for more information compared to no exposure to mass or interpersonal communication. Posts containing references to interpersonal communication had a 268% higher likelihood of displaying the need to ask fellow patients for their experiences compared to posts containing no signs of convergence.

**Table 3.** Differences between posts containing signs of convergence and posts without (reference category=no media).

Variable	Mass			Interpersonal		
	OR <sup>a</sup>	P value	95% CI	OR	P value	95% CI
<b>Information on poster</b>						
Disclosure of information about the disease	0.11	<.001	0.06-0.20	2.56	.009	1.26-5.20
Cancer in surrounding	1.08	.85	0.50-2.29	1.51	.10	0.92-2.49
<b>Needs</b>						
Community building	4.73	.002	0.92-2.49	0.21	.06	0.04-1.10
Sharing experiences	3.91	.003	1.59-9.59	1.71	.21	0.75-3.89
Asking for experiences	1.30	.58	0.52-3.30	3.68	.004	1.50-8.99
Asking for information	2.88	.04	1.04-7.98	1.91	.23	0.67-5.46

<sup>a</sup>OR: odds ratio.

The second multinomial logistic regression models contained the variables from the categories motive, information on poster, and needs (adjusted  $R^2=0.77$ ,  $\chi^2_{20}=-185.40$ ,  $P<.001$ ; Table 4). Within the category motive, in posts referencing interpersonal communication, a shortage of information was 81% less likely to be the reported outcome of the communication effort compared to posts referencing mass communication. Furthermore, within the category information on poster, posts referencing interpersonal communication were 2015% more likely to disclose information about the disease compared to posts referencing mass communication. Within the category needs, posts containing interpersonal communication were 93% less likely to display community building as a need of the post

compared to posts referencing mass communication. Furthermore, posts referencing interpersonal communication had a 227% higher likelihood of asking for other posters' experiences compared to posts referencing mass communication (OR 3.27,  $P=.04$ ).

To answer RQ3, compared to intramedium convergence, intermedium convergence posts are less likely to be motivated by a shortage of information and are more likely to contain information about the poster's condition. Furthermore, again compared to intramedium convergence, intermedium convergence is more likely to display the need for experiences and is less likely to exhibit a need for community building.



**Table 4.** Differences between posts containing signs of interpersonal convergence (reference category=mass communication).

Variable	OR <sup>a</sup>	P value	95% CI
<b>Motive</b>			
Conflict	2.77	.07	0.92-8.29
Shortage of information	0.19	.006	0.06-0.63
New questions	0.83	.69	0.30-2.23
Sharing of information	0.39	.11	0.12-1.24
<b>Information on poster</b>			
Disclosure of information about the disease	21.15	<.001	9.39-47.62
Cancer in surrounding	1.49	.36	0.64-3.47
<b>Needs</b>			
Community building	0.07	.004	0.01-0.44
Sharing experiences	0.65	.47	0.20-2.10
Asking for experience	3.27	.04	1.01-10.57
Asking for information	0.77	.70	0.20-2.90

<sup>a</sup>OR: odds ratio.

## Discussion

### Principal Findings

This study provides more insight into (the occurrence of) convergence using natural unsolicited data. Overall, intramedium and intermedium convergence resulted in posts containing different content and aiming to fulfill different needs. We found that nearly one-third of all forum opening posts in our sample contained signs of convergence by referencing either mass or interpersonal communication in the post. For intramedium convergence, online sources such as websites, forums, and online news articles were most often mentioned, frequently accompanied by a link to that source. In this way, posters seem to fulfill their need to help build the online community and initiate a discussion or to share experiences. Posts containing intermedium convergence often included references to a consultation with a medical expert. In these posts, users reported less shortage of information, disclosed more about themselves, and asked for more experiences from other users compared to posts containing intramedium convergence.

Our findings further emphasize the frequency of reported convergence and how intertwined these sources are. The main interpersonal communication source that was mentioned in the posts was that of a medical expert. This outcome is in line with previous research in which the medical expert, together with the internet, is named as the most important source of information for patients [19,23,47]. We found that one-third of the posts contained signs of convergence. The number of patients who use more than one medium is likely to be higher for two reasons. First, we only looked at specific types of convergence occurring in forum posts; however, based on previous research (eg, [35]), we know that signs of convergence also occur at the medical encounter and that different types of convergence exist. For example, during medical encounters, patients could discuss a forum they have read before the

consultation and thus engage in intermedium convergence (online forum-medical expert) or engage in intramedium convergence (ie, medical expert-medical expert) by referencing a medical expert during the consultation who provided a second opinion.

Second, we only coded explicit signs of convergences, whereas previous research also shows that patients implicitly mention different sources [48]. One aspect that is unique to this study is that although previous studies often examined both sources independently, the current results show how interdependent these sources are and how they are likely to continue to merge in the future. For example, a poster who recently had an appointment with a medical expert may have received a lot of information (convergence). After interpersonal communication, there is a lower likelihood that the patient experienced a shortage of information (motive). However, the patient might have missed information about how other patients experienced the situation, which motivates the patient to go online, write about their situation, and ask fellow patients for their experiences (need). According to the optimal matching theory [38], patients actively choose a medium that likely fulfills their needs.

In the context of support, some patients actively start participating in forums to find information that only fellow patients can provide—their experiences [14,49,50]. Our results also highlight the importance and added value of studying information sources in an interdependent context instead of independently. In light of the increased availability of different types of information on platforms, the internet seems to be a promising venue to fulfill needs that are not fulfilled during a consultation. Taking the notion of the optimal matching theory further, one could argue that it should not be a problem if patients report unmet needs based on their exposure to one medium, as another medium might be better able to fulfill these unmet needs. However, the medical expert and patient should

work together to make sure credible sources of information are known and available to the patient to fulfill their needs.

Based on our results, posters seem to require information provided by other patients combined with the information provided by the medical experts. Forums can be used to gain access to the experiences of fellow patients without the medical expert being an intermediary in this process. Users thereby benefit from both the expertise garnered during consultations with the medical expert and the experiences of fellow patients [51]. Eysenbach and colleagues [50] already highlighted that providing, receiving, and reading experiences from fellow patients is one of the main functions of social support communities. The current study shows how patients use health forums in a broader context of multiple available sources.

Because websites are easily shared and embedded in online tools such as online forums, the current study found many references to mass communication in general and online sources in particular. Mass communication is likely to be shared with members of the community to sustain and to inform the community through what is called “community building.” Community building creates a feeling of being part of a community and therefore fights the feeling of being alone, which in turn can emotionally support the patient [52].

### Limitations and Future Research

We posited that using a hybrid method on natural data could be a useful tool in meeting the challenges faced in studying convergence (ie, circular process, biased data when trusted on solicited recall data). Although we successfully analyzed indicators for convergence using forum data, some shortcomings must be acknowledged to advance future research. Despite the merits in using unsolicited data, not all aspects of convergence could be studied. First, we could only detect explicit signs of convergence. It would be a safe assumption to imagine convergence occurring in implicit ways as well, such as by simply posting a question without stating the events leading up to the post. Furthermore, convergence could only be measured when mass or interpersonal communication led to posting on a forum. However, posting online or reading posts and responding to these posts could lead to convergence elsewhere. By only studying online forum posts on one particular website, these types of convergence could not be measured. Although this would result in an underestimation of convergence instead of an overestimation, future research could address these types of convergence. Content analysis (on videotaped consultations) can, for instance, be combined with surveys to investigate patients’ (unmet) needs when they communicate and to gain insight into how patients use communication sources to cope with their needs. The online environment would be a logical place to administer these surveys since this environment does not require actual tracking; instead, log data and prompted surveys could minimize intrusion and reliance on recall. Finally, using natural data restricted the possibility to control for differences in personal characteristics of the poster because these variables are not known. Based on previous studies, we know that the way patients use online forums changes over time [53]. We did not account for these individual differences. Future studies could gather data from multiple forum messages and

profiles to extract information on the time of diagnosis, number of posts by the user, and type of disease to gain insight into these concepts.

SML was applied to create a subsample of posts containing signs of convergence. This approach resulted in a significantly smaller sample that had to be manually coded. If studies are interested in latent communication concepts such as the needs or motives of patients, researchers should take into account the time and effort needed to code a substantial part of their data as input for the SML, still without a guarantee that these latent construct can be reliably predicted. In an early phase of their study, researchers should decide on the role of SML in their project based on the number of positive cases per classifier and the initial SML results. Instead of coding a large portion of their data in the hope to obtain reliable classifiers for all constructs, reliable classifiers can be used in an early phase as a filter on the complete dataset to create a small subdataset that can be coded by hand.

The current study introduced two possible forms of biases. First, our sample consisted of posts from one forum on a highly trusted Dutch cancer website. Users on this forum might differ from the general cancer population in that they must have the skills to go online and register before using this forum. Furthermore, the fact that these patients opened a forum post could be an indication that they experienced a problem during a previous communication (eg, a shortage of information or conflicting information during the consultation with their medical expert). Therefore, the results might not be representative of all cancer patients, and the needs and motives found could be an overestimation of the unmet needs in this population. However, a complete export of all of the content of a platform with the informed consent of all users is still difficult to obtain, thus illustrating the uniqueness of our study. While the reported unmet needs might be an overestimation, these unmet needs still exist and will likely continue to exist. Therefore, scholars, medical experts, and (cancer) patient associations should work together to make convergence as easy as possible and try to incorporate alternative sources of information into the medical trajectory. For example, a leaflet or a website hosted by the hospital can provide patients with reliable sources but also well-known forums in which patients can exchange experiences and find support.

The second possible bias could have been created during the SML process. The SML algorithm that was used to create a sample of the posts used for the analysis might have caused a bias in the reference category. We manually created the reference category in which no signs of convergence were present. However, it is possible that the original algorithm marked these posts as false positives based on some shared content characteristics. This process might have led to differences between these false positives and the posts without signs of convergence in the corpus (ie, dataset) that were left out of the analysis. As a result, the reference sample might not completely be representative of the posts without signs of convergence. However, most of the main results are from a comparison between mass and interpersonal communication. These two samples were created by a combination of SML and manual checking; therefore, the above-described bias does not

play a role. To overcome this possible bias, future research could either randomly create a sample as the reference category or possibly compare the reference category that was created through machine learning to a random sample before running the analysis.

## Conclusions

To conclude, convergence is an important concept that represents the natural flow of patients' information-seeking behavior between and within interpersonal and mass communication. Understanding how patients use different

communication channels is essential to improving health care by providing guidance to patients who are trying to fulfill their needs. A better understanding of the conditions (ie, whether the information is discussed and in which way) under which the convergence of interpersonal and mass media results in positive patient outcomes might be the key to enhancing information provision to patients and in turn increasing patients' wellbeing. In doing so, providers should take a proactive role in discussing online information-seeking with patients and referring patients to the right sources that best meet their needs.

## Acknowledgments

We thank Mieke Burger, Integraal Kankercentrum Nederland (IKNL).

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Details of the supervised machine learning (SML) process.

[DOCX File, 14 KB - [jmir\\_v22i10e18303\\_app1.docx](#)]

### Multimedia Appendix 2

Confusion matrix classifiers per category.

[DOCX File, 15 KB - [jmir\\_v22i10e18303\\_app2.docx](#)]

## References

1. Bensing J, Verhaak P. Communication in medical encounters. In: Kaptein A, Weinman J, editors. Health Psychology. Malden: Blackwell Publishing; 2004:261-287.
2. Blandford A. HCI for health and wellbeing: Challenges and opportunities. Int J Hum-Comput Stud 2019 Nov;131:41-51. [doi: [10.1016/j.ijhcs.2019.06.007](#)]
3. Kreps GL. Online Information and Communication Systems to Enhance Health Outcomes Through Communication Convergence. Hum Commun Res 2017 Jun 30;43(4):518-530. [doi: [10.1111/hcre.12117](#)]
4. Voruganti T, Grunfeld E, Makuwaza T, Bender JL. Web-Based Tools for Text-Based Patient-Provider Communication in Chronic Conditions: Scoping Review. J Med Internet Res 2017 Oct 27;19(10):e366 [FREE Full text] [doi: [10.2196/jmir.7987](#)] [Medline: [29079552](#)]
5. Walther JB, Valkenburg PM. Merging Mass and Interpersonal Communication via Interactive Communication Technology: A Symposium. Hum Commun Res 2017 Jun 30;43(4):415-423. [doi: [10.1111/hcre.12120](#)]
6. van der Meij E, Huirne JA, Ten Cate AD, Stockmann HB, Scholten PC, Davids PH, et al. A Perioperative eHealth Program to Enhance Postoperative Recovery After Abdominal Surgery: Process Evaluation of a Randomized Controlled Trial. J Med Internet Res 2018 Jan 02;20(1):e1 [FREE Full text] [doi: [10.2196/jmir.8338](#)] [Medline: [29295808](#)]
7. van Eenbergen MC, van de Poll-Franse LV, Heine P, Mols F. The Impact of Participation in Online Cancer Communities on Patient Reported Outcomes: Systematic Review. JMIR Cancer 2017 Sep 28;3(2):e15 [FREE Full text] [doi: [10.2196/cancer.7312](#)] [Medline: [28958985](#)]
8. Lobchuk M, McClement S, Rigney M, Copeland A, Bayrampour H. A Qualitative Analysis of "Naturalistic" Conversations in a Peer-Led Online Support Community for Lung Cancer. Cancer Nurs 2015;38(5):E21-E31. [doi: [10.1097/NCC.0000000000000207](#)] [Medline: [25254411](#)]
9. Seale C, Ziebland S, Charteris-Black J. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. Soc Sci Med 2006 May;62(10):2577-2590. [doi: [10.1016/j.socscimed.2005.11.016](#)] [Medline: [16361029](#)]
10. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: [10.1001/jama.287.20.2691](#)] [Medline: [12020305](#)]
11. Batenburg A, Das E. Emotional approach coping and the effects of online peer-led support group participation among patients with breast cancer: a longitudinal study. J Med Internet Res 2014 Nov 28;16(11):e256 [FREE Full text] [doi: [10.2196/jmir.3517](#)] [Medline: [25474819](#)]

12. Nwosu CR, Cox BM. The impact of the internet on the doctor-patient relationship. *Health Informatics J* 2016 Jul 25;6(3):156-161. [doi: [10.1177/146045820000600308](https://doi.org/10.1177/146045820000600308)]
13. Caiata-Zufferey M, Abraham A, Sommerhalder K, Schulz PJ. Online health information seeking in the context of the medical consultation in Switzerland. *Qual Health Res* 2010 Aug;20(8):1050-1061. [doi: [10.1177/1049732310368404](https://doi.org/10.1177/1049732310368404)] [Medline: [20442347](https://pubmed.ncbi.nlm.nih.gov/20442347/)]
14. Rupert DJ, Gard Read J, Amoozegar JB, Moultrie RR, Taylor OM, O'Donoghue AC, et al. Peer-Generated Health Information: The Role of Online Communities in Patient and Caregiver Health Decisions. *J Health Commun* 2016 Nov;21(11):1187-1197 [FREE Full text] [doi: [10.1080/10810730.2016.1237592](https://doi.org/10.1080/10810730.2016.1237592)] [Medline: [27805496](https://pubmed.ncbi.nlm.nih.gov/27805496/)]
15. Shah DV, Cappella JN, Neuman WR. Big Data, Digital Media, and Computational Social Science. *Ann Am Acad Pol Soc Sci* 2015 Apr 09;659(1):6-13. [doi: [10.1177/0002716215572084](https://doi.org/10.1177/0002716215572084)]
16. Abdul-Muhsin H, Tyson M, Raghu S, Humphreys M. The Informed Patient: An Analysis of Information Seeking Behavior and Surgical Outcomes Among Men With Benign Prostatic Hyperplasia. *Am J Mens Health* 2017 Jan;11(1):147-153 [FREE Full text] [doi: [10.1177/1557988315617527](https://doi.org/10.1177/1557988315617527)] [Medline: [26614445](https://pubmed.ncbi.nlm.nih.gov/26614445/)]
17. Leydon G, Boulton M, Moynihan C, Jones A, Mossman J, Boudioni M, et al. Cancer patients' information needs and information seeking behaviour: in depth interview study. *BMJ* 2000 Apr 01;320(7239):909-913 [FREE Full text] [doi: [10.1136/bmj.320.7239.909](https://doi.org/10.1136/bmj.320.7239.909)] [Medline: [10742000](https://pubmed.ncbi.nlm.nih.gov/10742000/)]
18. Rahmani A, Ferguson C, Jabarzadeh F, Mohammadpoorasl A, Moradi N, Pakpour V. Supportive care needs of Iranian cancer patients. *Indian J Palliat Care* 2014 Sep;20(3):224-228 [FREE Full text] [doi: [10.4103/0973-1075.138400](https://doi.org/10.4103/0973-1075.138400)] [Medline: [25191012](https://pubmed.ncbi.nlm.nih.gov/25191012/)]
19. Hong Y, Peña-Purcell NC, Ory MG. Outcomes of online support and resources for cancer survivors: a systematic literature review. *Patient Educ Couns* 2012 Mar;86(3):288-296. [doi: [10.1016/j.pec.2011.06.014](https://doi.org/10.1016/j.pec.2011.06.014)] [Medline: [21798685](https://pubmed.ncbi.nlm.nih.gov/21798685/)]
20. Shea-Budgell M, Kostaras X, Myhill K, Hagen N. Information needs and sources of information for patients during cancer follow-up. *Curr Oncol* 2014 Aug;21(4):165-173 [FREE Full text] [doi: [10.3747/co.21.1932](https://doi.org/10.3747/co.21.1932)] [Medline: [25089098](https://pubmed.ncbi.nlm.nih.gov/25089098/)]
21. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. *Arch Intern Med* 2005;165(22):2618-2624. [doi: [10.1001/archinte.165.22.2618](https://doi.org/10.1001/archinte.165.22.2618)] [Medline: [16344419](https://pubmed.ncbi.nlm.nih.gov/16344419/)]
22. Prestin A, Vieux SN, Chou WS. Is Online Health Activity Alive and Well or Flatlining? Findings From 10 Years of the Health Information National Trends Survey. *J Health Commun* 2015;20(7):790-798. [doi: [10.1080/10810730.2015.1018590](https://doi.org/10.1080/10810730.2015.1018590)] [Medline: [26042588](https://pubmed.ncbi.nlm.nih.gov/26042588/)]
23. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. *Decis Supp Syst* 2014 Jan;57:417-427. [doi: [10.1016/j.dss.2012.10.047](https://doi.org/10.1016/j.dss.2012.10.047)]
24. Puts MTE, Papoutsis A, Springall E, Tourangeau AE. A systematic review of unmet needs of newly diagnosed older cancer patients undergoing active cancer treatment. *Support Care Cancer* 2012 Jul;20(7):1377-1394. [doi: [10.1007/s00520-012-1450-7](https://doi.org/10.1007/s00520-012-1450-7)] [Medline: [22476399](https://pubmed.ncbi.nlm.nih.gov/22476399/)]
25. Iconomou G, Vagenakis AG, Kalofonos HP. The informational needs, satisfaction with communication, and psychological status of primary caregivers of cancer patients receiving chemotherapy. *Support Care Cancer* 2001 Nov;9(8):591-596. [doi: [10.1007/s005200100259](https://doi.org/10.1007/s005200100259)] [Medline: [11762969](https://pubmed.ncbi.nlm.nih.gov/11762969/)]
26. Brandes K, van der Goot MJ, Smit EG, van Weert JC, Linn AJ. Understanding the interplay of cancer patients' instrumental concerns and emotions. *Patient Educ Couns* 2017 May;100(5):839-845. [doi: [10.1016/j.pec.2017.02.002](https://doi.org/10.1016/j.pec.2017.02.002)] [Medline: [28189470](https://pubmed.ncbi.nlm.nih.gov/28189470/)]
27. Hack TF, Degner LF, Parker PA, SCRIN Communication Team. The communication goals and needs of cancer patients: a review. *Psychooncology* 2005 Oct;14(10):831-845; discussion 846. [doi: [10.1002/pon.949](https://doi.org/10.1002/pon.949)] [Medline: [16200519](https://pubmed.ncbi.nlm.nih.gov/16200519/)]
28. Sommerhalder K, Abraham A, Zufferey MC, Barth J, Abel T. Internet information and medical consultations: experiences from patients' and physicians' perspectives. *Patient Educ Couns* 2009 Nov;77(2):266-271. [doi: [10.1016/j.pec.2009.03.028](https://doi.org/10.1016/j.pec.2009.03.028)] [Medline: [19411157](https://pubmed.ncbi.nlm.nih.gov/19411157/)]
29. Rheingold H. *The virtual community: Homesteading on the electronic frontier*. USA: Addison-Wesley Longman Publishing Co, Inc; 2000.
30. Gooden RJ, Winefield HR. Breast and prostate cancer online discussion boards: a thematic analysis of gender differences and similarities. *J Health Psychol* 2007 Jan;12(1):103-114. [doi: [10.1177/1359105307071744](https://doi.org/10.1177/1359105307071744)] [Medline: [17158844](https://pubmed.ncbi.nlm.nih.gov/17158844/)]
31. Sanders R, Linn AJ, Araujo TB, Vliegenthart R, van Eenbergen MC, van Weert JC. Different platforms for different patients' needs: Automatic content analysis of different online health information platforms. *Int J Hum-Comput Stud* 2020 May;137:102386. [doi: [10.1016/j.ijhcs.2019.102386](https://doi.org/10.1016/j.ijhcs.2019.102386)]
32. Sinha A, Porter T, Wilson A. The Use of Online Health Forums by Patients With Chronic Cough: Qualitative Study. *J Med Internet Res* 2018 Jan 24;20(1):e19 [FREE Full text] [doi: [10.2196/jmir.7975](https://doi.org/10.2196/jmir.7975)] [Medline: [29367181](https://pubmed.ncbi.nlm.nih.gov/29367181/)]
33. Bender JL, Feldman-Stewart D, Tong C, Lee K, Brundage M, Pai H, et al. Health-Related Internet Use Among Men With Prostate Cancer in Canada: Cancer Registry Survey Study. *J Med Internet Res* 2019 Nov 19;21(11):e14241 [FREE Full text] [doi: [10.2196/14241](https://doi.org/10.2196/14241)] [Medline: [31742561](https://pubmed.ncbi.nlm.nih.gov/31742561/)]



34. van Eenbergen MCHJ, Vromans RD, Boll D, Kil PJM, Vos CM, Krahmer EJ, et al. Changes in internet use and wishes of cancer survivors: A comparison between 2005 and 2017. *Cancer* 2020 Jan 15;126(2):408-415 [FREE Full text] [doi: [10.1002/cncr.32524](https://doi.org/10.1002/cncr.32524)] [Medline: [31580497](https://pubmed.ncbi.nlm.nih.gov/31580497/)]
35. Sanders R, Linn AJ. A Mixed Method Study Investigating the Impact of Talking about Patients' Internet Use on Patient-Reported Outcomes. *J Health Commun* 2018;23(9):815-823. [doi: [10.1080/10810730.2018.1514443](https://doi.org/10.1080/10810730.2018.1514443)] [Medline: [30351205](https://pubmed.ncbi.nlm.nih.gov/30351205/)]
36. Bylund CL, D'Agostino TA, Ostroff J, Heerdt A, Li Y, Dickler M. Exposure to and intention to discuss cancer-related internet information among patients with breast cancer. *J Oncol Pract* 2012 Jan;8(1):40-45 [FREE Full text] [doi: [10.1200/JOP.2011.000271](https://doi.org/10.1200/JOP.2011.000271)] [Medline: [22548010](https://pubmed.ncbi.nlm.nih.gov/22548010/)]
37. Linn AJ, Schouten BC, Sanders R, van Weert JC, Bylund CL. Talking about Dr. Google: Communication strategies used by nurse practitioners and patients with inflammatory bowel disease in the Netherlands to discuss online health information. *Patient Educ Couns* 2020 Jun;103(6):1216-1222. [doi: [10.1016/j.pec.2020.01.011](https://doi.org/10.1016/j.pec.2020.01.011)] [Medline: [32098743](https://pubmed.ncbi.nlm.nih.gov/32098743/)]
38. Cutrona CE. Stress and Social Support—in Search of Optimal Matching. *J Soc Clin Psychol* 1990 Mar;9(1):3-14. [doi: [10.1521/jscp.1990.9.1.3](https://doi.org/10.1521/jscp.1990.9.1.3)]
39. Chen Y, Li C, Liang J, Tsai C. Health Information Obtained From the Internet and Changes in Medical Decision Making: Questionnaire Development and Cross-Sectional Survey. *J Med Internet Res* 2018 Feb 12;20(2):e47 [FREE Full text] [doi: [10.2196/jmir.9370](https://doi.org/10.2196/jmir.9370)] [Medline: [29434017](https://pubmed.ncbi.nlm.nih.gov/29434017/)]
40. Stafford TF, Stafford MR, Schkade LL. Determining Uses and Gratifications for the Internet. *Decision Sci* 2004 May;35(2):259-288. [doi: [10.1111/j.00117315.2004.02524.x](https://doi.org/10.1111/j.00117315.2004.02524.x)]
41. Rubin A. The uses-and-gratifications perspective of media effects. In: Bryant J, Zillmann D, editors. *LEA's communication series. Media effects: Advances in theory and research*. New Jersey: Lawrence Erlbaum Associates Publishers; 2002:525-548.
42. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol* 2013 Sep 18;13:117 [FREE Full text] [doi: [10.1186/1471-2288-13-117](https://doi.org/10.1186/1471-2288-13-117)] [Medline: [24047204](https://pubmed.ncbi.nlm.nih.gov/24047204/)]
43. Kanker.nl. URL: [www.kanker.nl](http://www.kanker.nl) [accessed 2016-12-15]
44. World Health Organization. Cancer Today. International Agency for Research on Cancer. 2019. URL: <http://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf> [accessed 2019-10-19]
45. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Machine Learn Res* 2011 Jun;19(1):2825-2830.
46. Ranganathan P, Aggarwal R, Pramesh CS. Common pitfalls in statistical analysis: Odds versus risk. *Perspect Clin Res* 2015;6(4):222-224 [FREE Full text] [doi: [10.4103/2229-3485.167092](https://doi.org/10.4103/2229-3485.167092)] [Medline: [26623395](https://pubmed.ncbi.nlm.nih.gov/26623395/)]
47. Muusses LD, van Weert JCM, van Dulmen S, Jansen J. Chemotherapy and information-seeking behaviour: characteristics of patients using mass-media information sources. *Psychooncology* 2012 Sep;21(9):993-1002. [doi: [10.1002/pon.1997](https://doi.org/10.1002/pon.1997)] [Medline: [21671305](https://pubmed.ncbi.nlm.nih.gov/21671305/)]
48. Bylund CL, Makoul G. Examining empathy in medical encounters: an observational study using the empathic communication coding system. *Health Commun* 2005;18(2):123-140. [doi: [10.1207/s15327027hc1802\\_2](https://doi.org/10.1207/s15327027hc1802_2)] [Medline: [16083407](https://pubmed.ncbi.nlm.nih.gov/16083407/)]
49. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. *Soc Theory Health* 2013 Jun 19;11(3):256-270. [doi: [10.1057/sth.2013.10](https://doi.org/10.1057/sth.2013.10)]
50. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *BMJ* 2004 May 15;328(7449):1166 [FREE Full text] [doi: [10.1136/bmj.328.7449.1166](https://doi.org/10.1136/bmj.328.7449.1166)] [Medline: [15142921](https://pubmed.ncbi.nlm.nih.gov/15142921/)]
51. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. *J Med Internet Res* 2008 Aug 25;10(3):e22 [FREE Full text] [doi: [10.2196/jmir.1030](https://doi.org/10.2196/jmir.1030)] [Medline: [18725354](https://pubmed.ncbi.nlm.nih.gov/18725354/)]
52. Bender JL, Jimenez-Marroquin M, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. *J Med Internet Res* 2011 Feb 04;13(1):e16 [FREE Full text] [doi: [10.2196/jmir.1560](https://doi.org/10.2196/jmir.1560)] [Medline: [21371990](https://pubmed.ncbi.nlm.nih.gov/21371990/)]
53. Verberne S, Batenburg A, Sanders R, van Eenbergen M, Das E, Lambooi MS. Analyzing Empowerment Processes Among Cancer Patients in an Online Community: A Text Mining Approach. *JMIR Cancer* 2019 Apr 17;5(1):e9887 [FREE Full text] [doi: [10.2196/cancer.9887](https://doi.org/10.2196/cancer.9887)] [Medline: [30994468](https://pubmed.ncbi.nlm.nih.gov/30994468/)]

## Abbreviations

**OR:** odds ratio

**SML:** supervised machine learning



*Edited by G Eysenbach; submitted 24.02.20; peer-reviewed by PJ Schulz, YY Chen, D Carvalho; comments to author 20.04.20; revised version received 12.06.20; accepted 21.06.20; published 19.10.20.*

*Please cite as:*

Sanders R, Araujo TB, Vliegenthart R, van Eenbergen MC, van Weert JCM, Linn AJ

*Patients' Convergence of Mass and Interpersonal Communication on an Online Forum: Hybrid Methods Analysis*

*J Med Internet Res 2020;22(10):e18303*

URL: <https://www.jmir.org/2020/10/e18303>

doi: [10.2196/18303](https://doi.org/10.2196/18303)

PMID: [33074160](https://pubmed.ncbi.nlm.nih.gov/33074160/)

©Remco Sanders, Theo B Araujo, Rens Vliegenthart, Mies C van Eenbergen, Julia CM van Weert, Annemiek J Linn. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 19.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Responses to Concerning Posts on Social Media and Their Implications for Suicide Prevention Training for Military Veterans: Qualitative Study

Alan R Teo<sup>1,2,3</sup>, MD, MS; Wynn Strange<sup>1</sup>, MA; Ricky Bui<sup>1</sup>, BA; Steven K Dobscha<sup>1,2</sup>, MD; Sarah S Ono<sup>1,4,5</sup>, PhD

<sup>1</sup>Health Services Research and Development Center to Improve Veteran Involvement in Care, VA Portland Health Care System, Portland, OR, United States

<sup>2</sup>Department of Psychiatry, Oregon Health and Science University, Portland, OR, United States

<sup>3</sup>School of Public Health, Oregon Health and Science University and Portland State University, Portland, OR, United States

<sup>4</sup>Veterans Rural Health Resource Center–Portland, Veterans Health Administration Office of Rural Health, Portland, OR, United States

<sup>5</sup>Department of Family Medicine, Oregon Health and Science University, Portland, OR, United States

**Corresponding Author:**

Alan R Teo, MD, MS

Health Services Research and Development Center to Improve Veteran Involvement in Care

VA Portland Health Care System

3710 SW US Veterans Hospital Road

Portland, OR, 97239

United States

Phone: 1 5032208262 ext 52461

Email: [teoa@ohsu.edu](mailto:teoa@ohsu.edu)

## Abstract

**Background:** A “concerning post” is a display of a user’s emotional crisis on a social media platform. A better understanding of concerning posts is relevant to suicide prevention, but little is known about social media users’ attitudes and responses to concerning posts. Military veterans in the United States are disproportionately affected by suicide, often use social media, and may have exposure to individuals with elevated suicide risk via concerning posts.

**Objective:** The objective of the study was (1) to obtain insight into whether and how US military veterans respond to members of their social network on social media (ie, “friends”) who are experiencing substantial emotional distress, and (2) to identify potential interventions that could assist in users’ response to concerning posts.

**Methods:** We recruited veterans through Facebook and conducted semistructured interviews with 30 participants between June and December 2017. We used a summary template for rapid analysis of each interview, followed by double-coding using a codebook based on topic domains from the interview guide. Members of the research team met regularly to discuss emerging patterns in the data, generate themes, and select representative quotes for inclusion in the manuscript.

**Results:** Veterans were reluctant to disclose emotional and health issues on Facebook, but they were open to reaching out to others’ concerning posts. There was a complex calculus underlying whether and how veterans responded to a concerning post, which involved considering (1) physical proximity to the person posting, (2) relationship closeness, (3) existing responses to the post, and (4) ability to maintain contact with the person. Veterans desired additional training, backed by community-based veteran organizations, in how to respond to concerning posts from peers.

**Conclusions:** There is a need to incorporate features that will help veterans effectively respond to concerning posts from peers into suicide prevention training and to expand access for veterans to such training.

(*J Med Internet Res* 2020;22(10):e22076) doi:[10.2196/22076](https://doi.org/10.2196/22076)

**KEYWORDS**

concerning post; social media; suicide prevention; gatekeeper training; military veterans

## Introduction

As social media has become embedded into our daily lives, researchers have taken on the task of unpacking its myriad implications for our health. Studies of social media have often focused on examining psychosocial harms from exposure to social media, such as loneliness [1], negative social comparisons [2], and worsening of mental health status [3]. Perhaps because of negative public and scientific perceptions of social media, less attention has been paid to identifying ways that social media could be used to improve mental health [4]. In reality, both positive and negative consequences of social media coexist and are influenced by how individuals—and society collectively—interact using social media.

An especially delicate but important health concern in the social media landscape is suicide prevention. Suicidal individuals are more likely than nonsuicidal individuals to spend time online [5]. Expressions of suicidal ideation on social media have been widely reported and, at times, are followed by suicidal behavior [6,7]. Disclosures of emotional distress on social media reflect real-life concerns, with studies showing links between depressive language in Facebook posts and self-reported symptoms of depression and medical record–confirmed diagnoses of depression [8,9]. Even more visible to the public are media reports about suicide, which are relatively common on social media compared with reports of deaths from other causes [10]. Analysis of Twitter data has found that large reactions to this reporting, measured through increased mentions after the death of a celebrity by suicide, are linked to subsequent increases in deaths by suicide [11].

Amidst this contemporary reality, the notion of a “concerning post” has emerged, which we define as a display of a user’s emotional crisis or other indication of an acute mental health concern on a social media platform. Little is known about individuals’ attitudes and responses to concerning posts. Among the available studies, results have conflicted at times. For instance, one study found that youth lacked confidence in how to respond to a concerning post [12], while another study found that sharing emotional distress activated supportive responses [13].

While much of the research in this area has focused on youth, military veterans are another key target population for suicide interventions. Veterans in the United States are disproportionately affected by suicide. Their rate of suicide is 50% (1.5 times) higher than that of nonveterans, according to the most recent available data [14]. Suicide exposure (knowing someone who has died by suicide) is also very common and may occur via social media; an estimated 57% of veterans [15] and 65% of National Guard personnel [16] have been exposed to suicide. Social media plays an important role in maintaining veterans’ social networks, with one contributor being the physical separation from family and friends that occurs during deployment or active duty, as well as separation from their servicemember peers after leaving the military. Social media serves to sustain connection and communication over time and in particular through periods of transition [17].

Using data from interviews with veterans who use social media, our objectives in this study were (1) to obtain insight into whether and how veterans respond to “friends” and family members on social media who are experiencing substantial emotional distress, and (2) to identify potential interventions that could assist veterans in responding to concerning posts.

## Methods

### Setting and Sample

This research was conducted as part of a mixed-methods study focused on understanding opportunities to use social media platforms, particularly Facebook, to reach and provide support to veterans at risk of psychiatric problems. We interviewed Operation Enduring Freedom, Operation Iraqi Freedom, and Operation New Dawn veterans with a service separation date after October 1, 2001, who were in the metropolitan area of Portland, Oregon, United States. Thirty participants were interviewed—a number chosen based on the literature, feasibility, and our research team’s experience reaching thematic saturation with such a sample size [18,19].

### Data Collection

Participants who had responded to Facebook advertisements and completed an online survey were eligible for participation in the current study. The survey that was administered has been previously described [20,21]; in brief, it included items on sociodemographic characteristics, frequency of social contact, use of social media platforms, screening for psychiatric problems and health service utilization, and interest in potential online interventions for social support and suicide prevention. Among participants who used Facebook daily, we purposively sampled veterans that varied in the extent to which they used the Department of Veterans Affairs for their health care and varied in their interest in learning more about suicide prevention.

Fifty individuals were contacted to participate, and 30 of them completed an approximately 40- to 60-minute interview between June and December 2017. Twenty-three interviews were completed in person; the remaining seven interviews were conducted over the phone, which allowed for participation by individuals who could not present in person to participate in the study. Two research team members (HM and AT) conducted the interviews using a semistructured interview guide (Multimedia Appendix 1). Questions were organized into three primary domains: (1) participants’ social media use and preferences, (2) discussion of health issues on social media, and (3) suicide prevention in a social media context (including experiences with concerning posts and preferences related to suicide prevention training). Interviews were audio recorded and transcribed.

### Data Analysis

Survey responses were used for participants’ descriptive characteristics. For qualitative analysis, we first created a summary template for rapid analysis of each interview [22,23]. We then created a codebook based on topic domains from the semistructured interview guide and informed by notes from the summary templates. Two research assistants (RB and WS) coded the interviews using qualitative analysis software (Atlas.ti;

ATLAS.ti Scientific Software Development GmbH) for data management. As the primary coder, RB met weekly with the lead investigator (AT) to discuss general impressions and emergent themes. As the secondary coder, WS reviewed the primary coding, adding any codes that may not have been applied, with the goal of ensuring robust coding. AT adjudicated coding disputes as needed. Members of the research team (AT, RB, WS, SD, and SO) met regularly to discuss emerging patterns in the data, generate themes, and select representative quotes for inclusion in the manuscript. We assigned pseudonyms to the participants to maintain confidentiality.

## Results

### Principal Findings

A summary of participants' characteristics is presented in Table 1. The 30 participants (23 men, 7 women) came from all military branches, with 3 participants serving in two branches over their careers. The average age of participants was 41 years (range 24–64 years), 73% (22/30) were White, 60% (18/30) were married, and 57% (17/30) had a college or postgraduate degree. In terms of their mental health, 5 of 30 (17%) participants received a score of 3 or greater on the Patient Health Questionnaire-2, which is indicative of a positive screen for major depressive disorder [24]; 15 of 30 (50%) participants had considered suicide, and 5 of 30 (17%) participants had previously attempted suicide. While 90% (27/30) of participants used Facebook several times a day, the majority never or rarely used social media for sharing information about their health (23/30, 77%), symptoms such as mood swings, depression, anxiety, or sleep problems (21/30, 70%), or thoughts about suicide or self-harm (28/30, 93%).

Veterans were reluctant to disclose emotional and health issues on social media but were open to reaching out in response to others' concerning posts. Interviews revealed multiple reasons for veterans' reluctance to share negative emotional states or information about their health. First, Facebook was viewed as a venue for positive self-presentation; thus, they were disinclined to use it as a place to "air dirty laundry." Second, when veterans did need support, they usually preferred to obtain it from other avenues in the "real world" that involve "person-to-person interaction," or they preferred to be self-reliant. Also, veterans believed that sharing personal health information was "risky" to their privacy. Lastly, participants noted that Facebook friends were not necessarily "real friends."

While most veterans were cautious about disclosing their own health issues, a number of veterans were understanding of others who chose to post about health issues. One participant summarized this distinction:

*I think for some people they find that they can post about their bad day, and they get a bunch of people "like" it, a bunch of people put the sad face on, a*

*bunch of people say, "You're amazing. You're brilliant. You go girl." That kind of thing. I've definitely seen a ton of people do that, but at the same time I would never do that. And it's not like I don't want support, but I feel like I would be negatively judged for that. [Oscar]*

Further, many veterans were open to responding when others disclosed personal information in a post, particularly emotional struggles or suicidality.

*I've seen other friends do it, and I've come to the aid of other friends because they have shared it on Facebook out in the open just to their timeline or whatever. I've also helped friends who have reached out to me personally and [sent] Facebook messages and said, "I'm having this crisis. I can use some words of encouragement or some words of advice or some help just in general." [Bryan]*

Participants identified a post as concerning in several ways. Most obviously, this was direct language expressing emotional distress, particularly when the distressed person had a known mental health condition (eg, post-traumatic stress disorder or an anxiety disorder), when the person was acting out of character (eg, getting rid of their possessions), or if the person had one or more recent psychosocial stressor, such as divorce/breakup, death of a loved one, or loss of a job. Several veterans also expressed a feeling of concern without an identifiable cause, noting, for instance, "[The post] just seems off" [Seth]. Veterans occasionally did not interpret a post as concerning if they felt the post was "venting" [Natasha] or that the person was seeking attention.

There was a complex calculus behind whether and how individuals responded to a concerning post. Once veterans had determined a post to be concerning, they wanted to be there for a peer in distress, but they did not always know if and how they should respond. Before they reached out, veterans went through a series of considerations. Elements of this complex calculus included (1) physical proximity, (2) relationship closeness, (3) existing responses to the post, and (4) ability to follow up with the person in crisis. While the weight given to each consideration varied across situations and circumstances, each factor influenced decisions related to whether veterans responded, how they chose to reach out, and how urgently and directly they responded.

When a friend posted something on Facebook that appeared to indicate suicidal ideation, participants reported that the preferred response was to attempt to make direct, one-on-one contact to talk with the friend. This direct contact could be through Facebook itself (using a private message), email, a phone call or text, or meeting up in person. Multiple methods of contact could be combined, typically with the goal of having the online contacts lead to an in-person contact.

**Table 1.** Summary of participants' demographic, clinical, and social media use characteristics (N=30).

Characteristics	Mean (range)	n (%)
<b>Demographic characteristics</b>		
Age (years)	41 (24-64)	
Gender		
Male		23 (77)
Female		7 (23)
Race/ethnicity		
White		22 (73)
Hispanic		3 (10)
American Indian/Alaska Native		2 (7)
Asian		3 (10)
Educational attainment		
High school or high school equivalency certificate		1 (3)
Some college		9 (30)
Technical or vocational school		3 (10)
College graduate		11 (37)
Postgraduate		6 (20)
Marital status		
Single, never married		7 (18)
Divorced		4 (13)
Separated		1 (3)
Married		18 (60)
Military branch <sup>a</sup>		
Army		11 (37)
Marine Corps		9 (30)
Navy		6 (20)
Air Force		5 (17)
Coast Guard		2 (7)
<b>Clinical characteristics</b>		
Patient Health Questionnaire-2 score	1.5 (0-4)	
0		11 (37)
1		2 (7)
2		12 (40)
3		2 (7)
4		3 (10)
History of serious suicidal ideation <sup>b</sup>		
Yes		15 (50)
No		15 (50)
Number of prior suicide attempts <sup>c</sup>		
0		25 (83)
1		2 (7)
2		2 (7)



Characteristics	Mean (range)	n (%)
3		0 (0)
4		1 (3)
<b>Social media use</b>		
How often do you visit or use Facebook?		
Several times a day		27 (90)
Once a day		2 (7)
A few times a week		1 (3)
How often do you use any social media website to do the following:		
Share symptoms such as mood swings, depression, anxiety, or sleep problems?		
Never		14 (47)
Rarely		7 (18)
Sometimes		4 (13)
Usually		4 (13)
Always		1 (3)
Share information related to your health?		
Never		10 (33)
Rarely		13 (43)
Sometimes		5 (17)
Usually		2 (7)
Always		0 (0)
Share thoughts about suicide or hurting yourself in some way?		
Never		26 (87)
Rarely		2 (7)
Sometimes		2 (7)
Usually		0 (0)
Always		0 (0)

<sup>a</sup>Total is greater than 100% because three participants served in two branches.

<sup>b</sup>Survey item asked, "Have you ever seriously considered attempting suicide at some point in your life?"

<sup>c</sup>Survey item asked, "How many times in your life have you attempted suicide?"

## Physical Proximity

When veterans saw a concerning post on Facebook, one of the first considerations for how they chose to respond was their physical proximity to the individual in distress. If the friend was within a reasonable travel distance, veterans often indicated it was optimal to visit them or call immediately to arrange an in-person meetup as soon as possible. If a friend expressed loneliness or concerning thoughts, participants felt responding through Facebook alone was insufficient. By meeting up in person, veterans felt that they could "get them out and let them know that there are still people for them" [Jamie] and "read more in a person by watching their body language" [Robert]. If the friend was not in close proximity and the level of concern was high, some veterans noted that they might try to get a next-of-kin or other person who was physically closer to check in on the friend.

*Depends on physical proximity, if it's somebody within driving distance, within a few hours I'll be at their house banging on the fucking door. If not, there's going to be phone calls and emails and text messages and that's the way I would react. [Alyssa]*

## Relationship Closeness

Veterans varied in their responsiveness to concerning posts depending on the closeness of their relationship with the person in crisis. For instance, if the concerning post was from an acquaintance or someone that was on the periphery of their life, they were more likely to respond in a less direct or less immediate fashion.

*I'll send them the phone number for the veteran's suicide hotline, that's really about as far as I'm willing to go unless I know the person. [Andrea]*

*If I don't know them personally, I try to give an encouraging thought. [Natasha]*

For a close relationship, efforts to respond were more direct and more intensive. One veteran said that in the case of a close friend or family member, he would respond as follows:

*I probably wouldn't even be on social media in that case. I would've just gone to the person and, you know, physically gone to see them.* [Conrad]

For relationships that fall somewhere between acquaintance and close friend or family, participants tended to communicate their availability but ultimately left the onus of further connection on the person in crisis. These efforts were often made under the assumption that there was someone closer who could help the person in crisis but also acknowledged that they were unsure of this and were thus willing to provide support if needed.

*...you can reach out to them and just give them a little love and then you instant message him and say, "Here's my phone number and call me if you need to."* [Jerry]

### Existing Responses to the Post

A third consideration in the response calculus related to the interactions and activity in response to a concerning post, including "likes," comments, and shares that were visible to others. Often, the timing of the post relative to when it was being seen was a factor. Veterans stressed the importance of timing for two reasons: first, the circumstances surrounding the crisis may have changed, and second, others may have already intervened.

*I've seen it a couple times or I've been, for lack of better term, late to the conversation and you've seen other people comment before that and someone said, "Hey, I've talked to 'em.'"* [Neal]

Recognizing that timing of a post and actions by others might have resulted in resolution by the time they saw a concerning post, participants also described assessing if there had already been direct and supportive responses to the friend when deciding if they should respond.

*I'm usually not the first person to come along by it, so usually I'll check the comment section and see what's being done. Typically, someone will comment and then they'll say: I've got ahold of them, or got them talking.* [Richard]

*And after somebody posts, there's some likes and dislikes and maybe a "hey, what's up, call me." I imagine in those situations there's four or five personal conversations going on with that person. And then they decide who is understanding the situation, talk with them.* [Yolanda]

### Ability to Follow Up

Some veterans emphasized both the need to initially reach out and the importance of following up with a friend in crisis. Veterans who discussed the importance of follow-up considered whether they had the capacity to continue to maintain contact and actively provide support.

One veteran responded that she had a friend who posted on Facebook "reaching out about how lonely they were." She went on to say the following:

*...they're obviously looking for that public acknowledgment, so, I did that. And kept it... you know, more... general, I guess, and supportive, and then, messaged them more. You know, specifically relating to it and then scheduling a follow-up visit with them.* [Jamie]

She let the friend know they weren't alone and scheduled a time to meet up with them.

*...it was the next day or something. Just for a quick coffee date, you know, just to get them out and let them know that there are still people for them.* [Jamie]

*...you don't just listen and walk away. You still have to follow-up with them and keep talking with them. It might not even be someone that you actually really know, but if you come across someone in that state of mind, it's usually kind of obvious that something's not quite right so don't ignore them.* [Bryan]

### Suicide Prevention Training

Veterans desired suicide prevention training that would prepare them to assist others in emotional crisis. The suggestion for training on how to respond to concerning posts, specifically if suicidal ideation was suspected or indicated, was raised by participants as well as asked about by the interviewer. Veterans indicated a desire for training in suicide prevention skills to help a peer in crisis. A common reason they cited for wanting training was a sense of being unprepared to assist others in crisis or even worry about exacerbating the situation.

*I have no idea what to say but just trying to make myself available, trying to say as little damaging stuff as I can... It's mostly that I have no idea what should I do to help? ...I got pretty much nothing else.* [Alyssa]

*I don't know how to sometimes get the point across to them without sounding judgmental.* [Lynn]

### Training as an Ongoing Process

Even veterans who indicated having prior suicide prevention training, often during military service, felt that the training was inadequate or insufficient. This seemed especially vital for veterans who were interested in helping others but worried about how to initiate conversations with more socially distant individuals.

*Would I want more? Yes. Just to let me know I've got the basics down. I'm not a licensed mental health professional so I can't go about helping them solve their problems, but I can get [someone in crisis] in the right direction.* [Edward]

*I'd like to recognize more warning signs, and I would like to have some conversation items that I could take as far as how to talk about it, or how to bring it up.* [Lewis]

### Preferred Features for Suicide Prevention Training

Veterans identified two features of suicide prevention training that would make it especially appealing. First, training sponsored or delivered by community-based veteran organizations was believed to “carry more weight” [Neal] and be “a good source of credibility” [Edward]. Many participants who followed at least one veteran-centric Facebook page or group (eg, Dysfunctional Veterans, Lift for the 22, Gulf War Veterans, etc) felt these groups could act as trusted sponsors of suicide prevention training. One veteran described the kind of support he’s encountered in one such private Facebook group:

*...every 22nd of every month, we do a check in. We'll post a thing, check in with our veteran friends because it comes out that there's 22 vets that commit suicide every day. And we started a program called "22 to None," so we support each other. I've made a couple friends with some other vets, and other Navy vets. And, one guy... he told me, gave me his number, he says you ever start feeling really bad, give me a call. He was a trained counselor as well. [Keith]*

Second, veterans were most willing to receive training targeted at helping their veteran peers. Their shared experiences, military background, and camaraderie as a result of service made a difference in their level of comfort and trust when talking about the topic of suicide with other veterans.

*It's very easy for me to have a relationship or an exchange of information between another person who's in the service. It's easier than it is for me to sometimes talk with my wife, or members of my family because they don't have that background like we have—those who've served. [Meghan]*

*Have you ever had two different people tell you the same piece of information? One is Bob who you don't really know particularly well, but the other one is somebody who is close to you—you know and respect—they repeat that same information, you place a lot more weight on it. With somebody who you've had a lot of those shared experiences with, I may not know them particularly well, but they've been in the same places I have, they've faced the same problems I've had, and if they say that this is helpful, I place a lot more weight on it. [Alyssa]*

## Discussion

### Principal Findings

In the daily course of their lives, veterans are likely to encounter others in emotional crisis, and this can manifest through concerning posts on social media platforms such as Facebook. When presented with a concerning post, veterans go through a complex calculus—involving an assessment of physical proximity, relationship closeness, activity on the post, and ability to follow-up—which informs the extent and manner in which they reach out to provide support. The order of the complex calculus and the weight of each element varies in each situation. Veterans are often inclined to help their peers in some way, but they often feel unprepared as to how to respond to a concerning

post. Given these circumstances, veterans desire suicide prevention training that extends beyond what they received during military service. They seek additional training that is germane to responding to concerning posts on social media, backed by community-based veteran organizations, and tailored to providing assistance to their veteran peers.

### Links to Existing Research on Social Media and Suicide Prevention

Our finding that veterans are thoughtful and potentially motivated to respond positively to concerning posts suggests a key exception to prior research on displays of negative emotions on social media. Experimental and observational studies have shown that emotions are transmitted through social media posts [25,26]. Individuals who frequently post on social media are more likely to post about health issues and to have a diagnosis of depression [27]. Moreover, early research on Facebook posting suggested that depressed individuals see the platform as an appealing venue for self-disclosure, but their posts are often met with undesirable responses due to the negativity of the posts [28]. It may be that military culture around altruism and stewardship toward veteran peers (eg, taking care of “our own,” leaving no one behind, etc) [29] contribute to the positive intentions we found toward responding to concerning posts. Our findings also comport with research showing that individuals often fear personal expressions of vulnerability but tend to react more positively to those who express it [30].

### Training Models Applicable to Social Media–Based Suicide Prevention Training

Our findings complement the use of two education and training intervention models in the literature: gatekeeper training and bystander education. Both gatekeeper training and bystander education train individuals to recognize the signs of crisis and develop skills to triage such a situation and assist with accessing subsequent professional assistance [31,32].

Gatekeeper training has been specifically designed to support “upstream” suicide prevention. Reviews have found evidence of gatekeeper training’s ability to increase knowledge and foster self-efficacy and other beliefs adaptive to suicide prevention [32]. In a recently published pilot intervention, a gatekeeper training approach helped prepare adults to support American Indian and Alaska Native youth who post or view concerning messages on social media [33]. Bystander interventions, also called bystander education, have been primarily used for prevention of sexual violence, and they emphasize a five-step model: (1) awareness of an issue that requires intervention; (2) identifying when the bystander should intervene; (3) deciding to take responsibility; (4) deciding how to help; and (5) taking action [31]. Education about all five steps is meant to increase the likelihood of a bystander taking action at multiple time points. Both of these educational approaches would seem highly applicable to craft training for veterans and others in how to identify and respond to concerning posts. Gatekeeper training and bystander education might be strengthened if future research identifies how expressions of vulnerability and peer support could normalize help-seeking and dialogue about suicide.

## Limitations

Our findings are constrained by several study limitations. Participants were recruited from a sample of locally recruited veterans from Facebook, and thus our findings may not be entirely generalizable to veterans nationally, individuals less inclined to participate in research or to use Facebook, or individuals who have experiences on other social media platforms. The last constraint is particularly relevant given rapid shifts in the social media landscape. There are other facets of Facebook posts, such as posts within private Facebook groups or automatic identification of concerning posts using machine learning models [34], that were beyond the scope of this study but merit closer investigation as part of a larger social media

suicide prevention strategy. Finally, as with all qualitative work, we took participants on their word that they accurately represented themselves and their experiences.

## Conclusions

Social media is an important venue through which many connect with their social network. One consequence of this, particularly among veterans who are disproportionately affected by and exposed to suicide, is encountering concerning posts. Our findings call for expanding access to ongoing suicide prevention training opportunities—in partnership with community-based veteran organizations—and incorporating features into existing suicide prevention training programs that will help veterans in their response to concerning posts from their veteran peers.

## Acknowledgments

This project was supported, in part, by a Career Development Award to Dr Teo from the Veterans Health Administration Health Service Research and Development (CDA 14-428) and the Health Service Research and Development Center to Improve Veteran Involvement in Care (CIVIC) (CIN 13-404). This project was also supported by awards received by Dr Teo from the Collins Medical Trust Research Grant and the Medical Research Foundation of Oregon New Investigator Grant (1603). The authors wish to thank Ms Heather Marsh, at CIVIC, for her assistance.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview Guide.

[DOCX File, 29 KB - [jmir\\_v22i10e22076\\_app1.docx](#)]

## References

1. Primack BA, Shensa A, Sidani JE, Whaitte EO, Lin LY, Rosen D, et al. Social Media Use and Perceived Social Isolation Among Young Adults in the U.S. *Am J Prev Med* 2017;53(1):1-8. [doi: [10.1016/j.amepre.2017.01.010](#)] [Medline: [28279545](#)]
2. Verduyn P, Ybarra O, Résibois M, Jonides J, Kross E. Do Social Network Sites Enhance or Undermine Subjective Well-Being? A Critical Review. *Social Issues and Policy Review* 2017 Jan 13;11(1):274-302. [doi: [10.1111/sipr.12033](#)]
3. Shakya HB, Christakis NA. Association of Facebook Use With Compromised Well-Being: A Longitudinal Study. *Am J Epidemiol* 2017 Feb 01;185(3):203-211. [doi: [10.1093/aje/kww189](#)] [Medline: [28093386](#)]
4. Pagoto S, Waring ME, May CN, Ding EY, Kunz WH, Hayes R, et al. Adapting Behavioral Interventions for Social Media Delivery. *J Med Internet Res* 2016 Jan 29;18(1):e24 [FREE Full text] [doi: [10.2196/jmir.5086](#)] [Medline: [26825969](#)]
5. Harris KM, Mclean JP, Sheffield J. Suicidal and Online: How Do Online Behaviors Inform Us of This High-Risk Population? *Death Studies* 2014;38(6):387-394. [doi: [10.1080/07481187.2013.768313](#)] [Medline: [24666145](#)]
6. Fu K, Cheng Q, Wong PWC, Yip PSF. Responses to a self-presented suicide attempt in social media: a social network analysis. *Crisis* 2013 Jan 01;34(6):406-412 [FREE Full text] [doi: [10.1027/0227-5910/a000221](#)] [Medline: [23871954](#)]
7. Ruder TD, Hatch GM, Ampanozi G, Thali MJ, Fischer N. Suicide announcement on Facebook. *Crisis* 2011;32(5):280-282. [doi: [10.1027/0227-5910/a000086](#)] [Medline: [21940257](#)]
8. Moreno MA, Christakis DA, Egan KG, Jelenchick LA, Cox E, Young H, et al. A pilot evaluation of associations between displayed depression references on Facebook and self-reported depression using a clinical scale. *J Behav Health Serv Res* 2012 Jul;39(3):295-304 [FREE Full text] [doi: [10.1007/s11414-011-9258-7](#)] [Medline: [21863354](#)]
9. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preotăciuc-Pietro D, et al. Facebook language predicts depression in medical records. *Proc Natl Acad Sci U S A* 2018 Oct 30;115(44):11203-11208. [doi: [10.1073/pnas.1802331115](#)] [Medline: [30322910](#)]
10. Scourfield J, Colombo G, Burnap P, Evans R, Jacob N, Williams M, et al. The Number and Characteristics of Newspaper and Twitter Reports on Suicides and Road Traffic Deaths in Young People. *Arch Suicide Res* 2019 Sep 05;23(3):507-522. [doi: [10.1080/13811118.2018.1479321](#)] [Medline: [29856679](#)]
11. Ueda M, Mori K, Matsubayashi T, Sawada Y. Tweeting celebrity suicides: Users' reaction to prominent suicide deaths on Twitter and subsequent increases in actual suicides. *Soc Sci Med* 2017 Sep;189:158-166 [FREE Full text] [doi: [10.1016/j.socscimed.2017.06.032](#)] [Medline: [28705550](#)]



12. Gritton J, Rushing S, Stephens D, Ghost Dog T, Kerr B, Moreno M. Responding to concerning posts on social media: Insights and solutions from American Indian and Alaska Native youth. *Am Indian Alsk Native Ment Health Res* 2017;24(3):63-87. [doi: [10.5820/aian.2403.2017.63](https://doi.org/10.5820/aian.2403.2017.63)] [Medline: [29161455](https://pubmed.ncbi.nlm.nih.gov/29161455/)]
13. Park J, Lee DS, Shablack H, Verduyn P, Deldin P, Ybarra O, et al. When perceptions defy reality: The relationships between depression and actual and perceived Facebook social support. *J Affect Disord* 2016 Aug;200:37-44. [doi: [10.1016/j.jad.2016.01.048](https://doi.org/10.1016/j.jad.2016.01.048)] [Medline: [27126138](https://pubmed.ncbi.nlm.nih.gov/27126138/)]
14. Office of Mental Health and Suicide Prevention. 2019 National Veteran Suicide Prevention Annual Report Internet. US Department of Veterans Affairs. 2019. URL: [https://www.mentalhealth.va.gov/docs/data-sheets/2019/2019\\_National\\_Veteran\\_Suicide\\_Prevention\\_Annual\\_Report\\_508.pdf](https://www.mentalhealth.va.gov/docs/data-sheets/2019/2019_National_Veteran_Suicide_Prevention_Annual_Report_508.pdf) [accessed 2020-10-20]
15. Hom MA, Podlogar MC, Stanley IH, Joiner TE. Ethical Issues and Practical Challenges in Suicide Research. *Crisis* 2017 Mar;38(2):107-114. [doi: [10.1027/0227-5910/a000415](https://doi.org/10.1027/0227-5910/a000415)] [Medline: [27561221](https://pubmed.ncbi.nlm.nih.gov/27561221/)]
16. Bryan CJ, Cerel J, Bryan AO. Exposure to suicide is associated with increased risk for suicidal thoughts and behaviors among National Guard military personnel. *Comprehensive Psychiatry* 2017 Aug;77:12-19. [doi: [10.1016/j.comppsy.2017.05.006](https://doi.org/10.1016/j.comppsy.2017.05.006)]
17. Mikal JP, Rice RE, Abeyta A, DeVilbiss J. Transition, stress and computer-mediated social support. *Computers in Human Behavior* 2013 Sep;29(5):A40-A53. [doi: [10.1016/j.chb.2012.12.012](https://doi.org/10.1016/j.chb.2012.12.012)]
18. Mason M. Sample Size and Saturation in PhD Studies Using Qualitative Interviews. *Forum: Qualitative Social Research* 2010 Aug;11(3):Art. 8 [FREE Full text]
19. Charmaz K. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. London: Sage Publications; 2006.
20. Teo AR, Chan BK, Saha S, Nicolaidis C. Frequency of social contact in-person vs. on Facebook: An examination of associations with psychiatric symptoms in military veterans. *J Affect Disord* 2019 Jan 15;243:375-380. [doi: [10.1016/j.jad.2018.09.043](https://doi.org/10.1016/j.jad.2018.09.043)] [Medline: [30266029](https://pubmed.ncbi.nlm.nih.gov/30266029/)]
21. Teo AR, Liebow SB, Chan B, Dobscha SK, Graham AL. Reaching Those At Risk for Psychiatric Disorders and Suicidal Ideation: Facebook Advertisements to Recruit Military Veterans. *JMIR Ment Health* 2019 Jan 09;6(1):e10078 [FREE Full text] [doi: [10.2196/13035](https://doi.org/10.2196/13035)] [Medline: [30626563](https://pubmed.ncbi.nlm.nih.gov/30626563/)]
22. Beebe J. *Rapid Assessment Process: An Introduction*. Walnut Creek: AltaMira Press; 2001.
23. Hamilton A. Qualitative methods in rapid turn-around health services research Internet. US Department of Veterans Affairs. 2013. URL: [http://www.hsrd.research.va.gov/for\\_researchers/cyber\\_seminars/archives/video\\_archive.cfm?SessionID=780](http://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=780) [accessed 2016-11-03]
24. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003 Nov;41(11):1284-1292. [doi: [10.1097/01.MLR.0000093487.78664.3C](https://doi.org/10.1097/01.MLR.0000093487.78664.3C)] [Medline: [14583691](https://pubmed.ncbi.nlm.nih.gov/14583691/)]
25. Ferrara E, Yang Z. Measuring Emotional Contagion in Social Media. *PLoS ONE* 2015 Nov 6;10(11):e0142390. [doi: [10.1371/journal.pone.0142390](https://doi.org/10.1371/journal.pone.0142390)]
26. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci U S A* 2014 Jun 17;111(24):8788-8790. [doi: [10.1073/pnas.1320040111](https://doi.org/10.1073/pnas.1320040111)] [Medline: [24889601](https://pubmed.ncbi.nlm.nih.gov/24889601/)]
27. Smith RJ, Crutchley P, Schwartz HA, Ungar L, Shofer F, Padrez KA, et al. Variations in Facebook Posting Patterns Across Validated Patient Health Conditions: A Prospective Cohort Study. *J Med Internet Res* 2017 Dec 06;19(1):e7 [FREE Full text] [doi: [10.2196/jmir.6486](https://doi.org/10.2196/jmir.6486)] [Medline: [28062392](https://pubmed.ncbi.nlm.nih.gov/28062392/)]
28. Forest AL, Wood JV. When social networking is not working: individuals with low self-esteem recognize but do not reap the benefits of self-disclosure on Facebook. *Psychol Sci* 2012 Mar 07;23(3):295-302. [doi: [10.1177/0956797611429709](https://doi.org/10.1177/0956797611429709)] [Medline: [22318997](https://pubmed.ncbi.nlm.nih.gov/22318997/)]
29. Greden J, Valenstein M, Spinner J, Blow A, Gorman L, Dalack G, et al. Buddy-to-Buddy, a citizen soldier peer support program to counteract stigma, PTSD, depression, and suicide. *Ann N Y Acad Sci* 2010 Oct;90-97. [doi: [10.1111/j.1749-6632.2010.05719.x](https://doi.org/10.1111/j.1749-6632.2010.05719.x)] [Medline: [20955330](https://pubmed.ncbi.nlm.nih.gov/20955330/)]
30. Bruk A, Scholl SG, Bless H. Beautiful mess effect: Self-other differences in evaluation of showing vulnerability. *Journal of Personality and Social Psychology* 2018 Aug;115(2):192-205. [doi: [10.1037/pspa0000120](https://doi.org/10.1037/pspa0000120)]
31. Amar AF, Sutherland M, Kesler E. Evaluation of a Bystander Education Program. *Issues in Mental Health Nursing* 2012 Dec 10;33(12):851-857. [doi: [10.3109/01612840.2012.709915](https://doi.org/10.3109/01612840.2012.709915)]
32. Burnette C, Ramchand R, Ayer L. Gatekeeper Training for Suicide Prevention: A Theoretical Model and Review of the Empirical Literature. *Rand Health Q* 2015 Jul 15;5(1):16 [FREE Full text] [Medline: [28083369](https://pubmed.ncbi.nlm.nih.gov/28083369/)]
33. Kerr B, Stephens D, Pham D, Ghost Dog T, McCray C, Caughlan C, et al. Assessing the Usability, Appeal, and Impact of a Web-Based Training for Adults Responding to Concerning Posts on Social Media: Pilot Suicide Prevention Study. *JMIR Ment Health* 2020 Jan 20;7(1):e14949 [FREE Full text] [doi: [10.2196/14949](https://doi.org/10.2196/14949)] [Medline: [31958066](https://pubmed.ncbi.nlm.nih.gov/31958066/)]
34. Liu X, Liu X, Sun J, Yu NX, Sun B, Li Q, et al. Proactive Suicide Prevention Online (PSPO): Machine Identification and Crisis Management for Chinese Social Media Users With Suicidal Thoughts and Behaviors. *J Med Internet Res* 2019 May 08;21(5):e11705 [FREE Full text] [doi: [10.2196/11705](https://doi.org/10.2196/11705)] [Medline: [31344675](https://pubmed.ncbi.nlm.nih.gov/31344675/)]



*Edited by G Eysenbach; submitted 02.07.20; peer-reviewed by E Ding, J Richards; comments to author 22.08.20; revised version received 08.09.20; accepted 21.09.20; published 30.10.20.*

*Please cite as:*

Teo AR, Strange W, Bui R, Dobscha SK, Ono SS

*Responses to Concerning Posts on Social Media and Their Implications for Suicide Prevention Training for Military Veterans: Qualitative Study*

*J Med Internet Res* 2020;22(10):e22076

URL: <https://www.jmir.org/2020/10/e22076>

doi: [10.2196/22076](https://doi.org/10.2196/22076)

PMID: [33124990](https://pubmed.ncbi.nlm.nih.gov/33124990/)

©Alan R Teo, Wynn Strange, Ricky Bui, Steven K Dobscha, Sarah S Ono. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Health Outcomes from Home Hospitalization: Multisource Predictive Modeling

Mireia Calvo<sup>1</sup>, PhD; Rubèn González<sup>2</sup>, MSc; Núria Seijas<sup>2</sup>, MSc, RN; Emili Vela<sup>3</sup>, MSc; Carme Hernández<sup>2</sup>, PhD; Guillem Batiste<sup>2</sup>, MD; Felip Miralles<sup>4</sup>, PhD; Josep Roca<sup>2</sup>, MD, PhD; Isaac Cano<sup>2\*</sup>, PhD; Raimon Jané<sup>1\*</sup>, PhD

<sup>1</sup>Institute for Bioengineering of Catalonia (IBEC), Barcelona Institute of Science and Technology (BIST), Universitat Politècnica de Catalunya (UPC), CIBER-BBN, Barcelona, Spain

<sup>2</sup>Hospital Clínic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona (UB), Barcelona, Spain

<sup>3</sup>Àrea de sistemes d'informació, Servei Català de la Salut, Barcelona, Spain

<sup>4</sup>Eurecat, Technology Center of Catalonia, Barcelona, Spain

\*these authors contributed equally

**Corresponding Author:**

Isaac Cano, PhD

Hospital Clínic de Barcelona

Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS)

Universitat de Barcelona (UB)

Villarroel 170

Barcelona, 08036

Spain

Phone: 34 932275540

Email: [ISCANO@clinic.cat](mailto:ISCANO@clinic.cat)

## Abstract

**Background:** Home hospitalization is widely accepted as a cost-effective alternative to conventional hospitalization for selected patients. A recent analysis of the home hospitalization and early discharge (HH/ED) program at Hospital Clínic de Barcelona over a 10-year period demonstrated high levels of acceptance by patients and professionals, as well as health value-based generation at the provider and health-system levels. However, health risk assessment was identified as an unmet need with the potential to enhance clinical decision making.

**Objective:** The objective of this study is to generate and assess predictive models of mortality and in-hospital admission at entry and at HH/ED discharge.

**Methods:** Predictive modeling of mortality and in-hospital admission was done in 2 different scenarios: at entry into the HH/ED program and at discharge, from January 2009 to December 2015. Multisource predictive variables, including standard clinical data, patients' functional features, and population health risk assessment, were considered.

**Results:** We studied 1925 HH/ED patients by applying a random forest classifier, as it showed the best performance. Average results of the area under the receiver operating characteristic curve (AUROC; sensitivity/specificity) for the prediction of mortality were 0.88 (0.81/0.76) and 0.89 (0.81/0.81) at entry and at home hospitalization discharge, respectively; the AUROC (sensitivity/specificity) values for in-hospital admission were 0.71 (0.67/0.64) and 0.70 (0.71/0.61) at entry and at home hospitalization discharge, respectively.

**Conclusions:** The results showed potential for feeding clinical decision support systems aimed at supporting health professionals for inclusion of candidates into the HH/ED program, and have the capacity to guide transitions toward community-based care at HH discharge.

(*J Med Internet Res* 2020;22(10):e21367) doi:[10.2196/21367](https://doi.org/10.2196/21367)

**KEYWORDS**

home hospitalization; health risk assessment; predictive modeling; chronic care; integrated care; modeling; hospitalization; health risk; prediction; mortality; clinical decision support

## Introduction

### Home Hospitalization and Early Discharge at the Hospital Clinic of Barcelona

Home hospitalization (HH)/early discharge (ED) programs [1-6] show substantial site heterogeneities in terms of service workflows and organizational aspects. However, overall, they have demonstrated maturity and health care value generation [7] such that it is well accepted that HH/ED constitutes an effective alternative to inpatient care for a select group of patients requiring hospital admission.

The characteristics of the deployment and adoption of the HH/ED program at Hospital Clinic of Barcelona (HCB) were recently described in a report [8]. In this report, HH/ED is defined as a service providing acute, home-based, short-term, complex interventions aimed at substituting conventional hospitalization fully with HH [7,9] or partially with ED [10]. The service at HCB is delivered by trained hospital personnel, and it is provided for a period of time that is not longer than the expected length of hospital stay for the patients' diagnostic related groups involved [11]. The Hospital retains the entire clinical, fiscal, and legal responsibilities. Virtual beds are used to support the required administrative and clinical processes. The report concluded that HH/ED for acute medical and surgical patients in a real-world setting was safe, generated healthcare efficiencies, and was well accepted by 98% of patients and professionals [8]. Moreover, the study stressed the potential of HH/ED to strengthen care coordination between highly specialized hospital-based care and home-based services involving different levels of complexity [8].

Currently, the HH/ED program at HCB is a mainstream, mature service that is offered 24 hours a day, 7 days a week, all year round, with 48 virtual beds available per day. It is the first choice for eligible patients requiring hospital admission when attended in the Emergency Department, and it serves the entire Health district of Barcelona Eixample-Esquerra, which has 540,000 inhabitants.

It is well accepted that the key health outcomes that define the success of hospitalization at home [8] are mortality and unplanned emergency room consultations that lead to in-hospital admissions, either during the home hospitalization episode or during the 30-day period after discharge. This study relies on the assumption that multisource predictive modeling facilitating clinical decision support at 2 key time points—(1) at entry, and (2) at HH/ED discharge—could be useful to enhance service outcomes. Risk assessment at entry may contribute to reducing undesirable events during the episode of HH/ED, whereas the assessment of unexpected events after discharge will likely contribute to improving transitional care [12,13] and better definition of personalized care pathways within a care continuum scenario [14].

### The Use of Multisource Predictive Modeling for Enhanced Risk Assessment

This study was designed to elaborate and assess the potential of a machine learning approach to the prediction of mortality and hospital admission at entry and at discharge from HH/ED.

A key specificity of the study is the use of various data sources to estimate the 2 outcomes, mortality and hospital re-admission, as conventional inpatient care. In addition to classical clinical and biological information obtained from electronic medical records (EMR), we have also considered the inclusion of Catalan population–health risk assessment scoring, known as Adjusted Morbidity Groups (GMA) [15,16], and purposely collected data on patients' performance and frailty.

The GMA is an open, publicly owned algorithm that does not rely on expert-based fixed coefficients. Such characteristics provide a high degree of flexibility for multisource predictive modeling and good potential for transferability to other sites, as demonstrated through its validation and current use in 13 of the 17 health regions in Spain, encompassing approximately 38,000,000 citizens [15]. It is fully operational since 2015 for health policy purposes and for clinicians in primary care workstations, providing yearly updated risk stratification with a population health orientation. It takes into account multimorbidity and complexity, that is, impact on health care, using data across health care tiers stored in the Catalan Health Surveillance System.

The approach adopted in this study was based on the hypothesis that the application of holistic strategies for subject-specific risk prediction and stratification, which consider multisource covariates influencing patient health, could increase predictive accuracy and facilitate clinical decision-making based on sound estimates of individual prognosis [17]. Developed predictive models were evaluated on a real-world database, which included all cases admitted to HH/ED at HCB from January 2009 to December 2015.

## Methods

### Dataset

Retrospective data from 1936 patients admitted to the HH/ED program at HCB from January 2009 to December 2015 (Table 1S in [Multimedia Appendix 1](#)) were considered in the analyses carried out to elaborate the predictive modeling of mortality and hospital re-admission at 2 time points: (1) at entry into HH/ED, and (2) at discharge from the HH/ED program. HH/ED at HCB is run as a transversal program, under the responsibility of the medical and nurse directors of the Hospital, serving the different clinical specialties. Patients included in the HH/ED show a broad spectrum of primary diagnoses, as displayed in Table 1S in [Multimedia Appendix 1](#).

The potential covariates considered for predictive modeling purposes (Table 2S in [Multimedia Appendix 1](#)) encompassed 3 dimensions: (1) standard clinical and biological information obtained from EMRs; (2) patients' functional performance and frailty data, specifically collected to characterize these patients; and (3) GMA scoring indicating multimorbidity, complexity, and patients' allocation into the population–health risk stratification pyramid.

### Ethical Approval

The Ethical Committee for Human Research at HCB approved the study, and all participants signed an informed consent prior

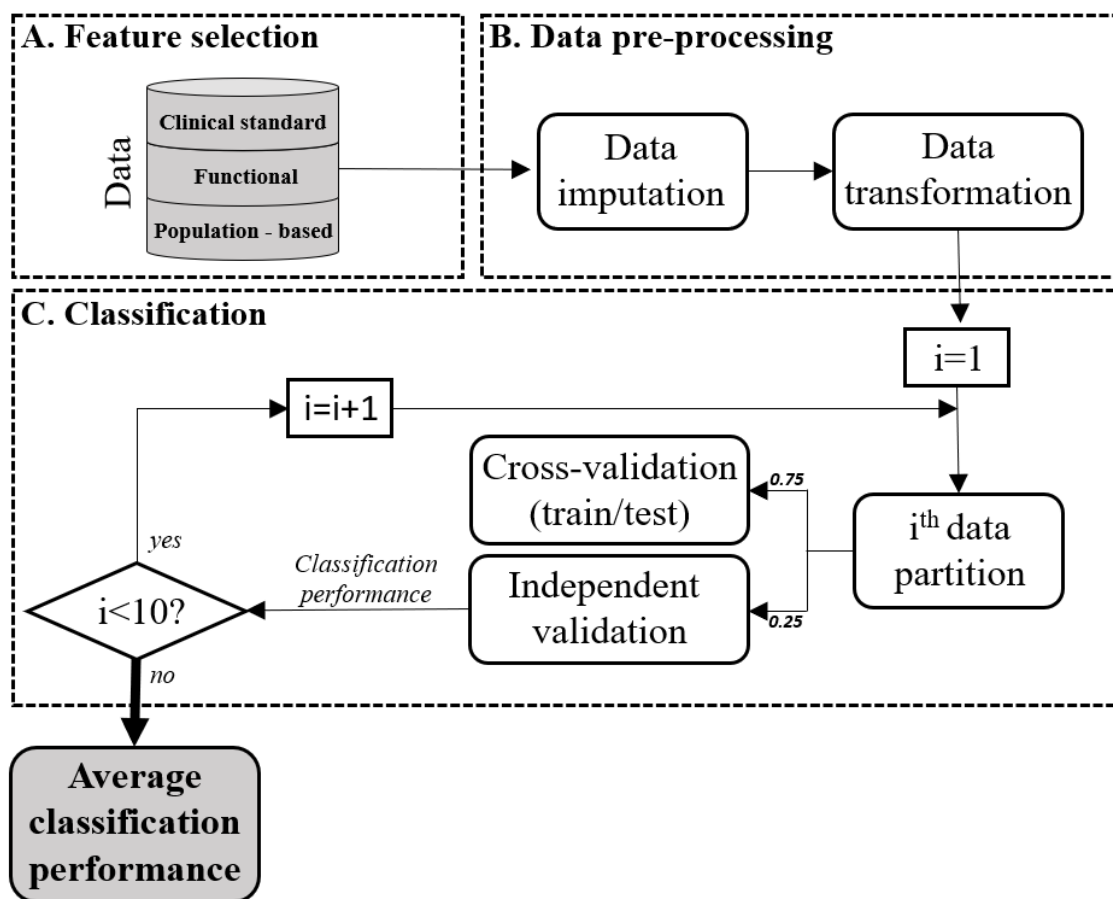
to any procedure. The program was registered at ClinicalTrials.gov: NCT03130283.

### Predictive Analytics Workflow

Figure 1 illustrates the global methodology proposed to identify patients at risk of re-admission or death after HH discharge; the

elaboration of predictive modeling followed 3 successive steps: (1) feature selection, (2) data preprocessing, and (3) classification.

**Figure 1.** Predictive analytics workflow, composed of 3 main steps: (A) feature selection, (B) data preprocessing, and (C) classification.



### Feature Selection

Feature selection refers to different processes involving data cleaning, selection of variables to be considered for predictive modelling, as well as selection of the final set of patients included in the analyses.

### Data Preprocessing

In order to handle the impact of missing values, a robust method was designed for mixed-type data imputation. To this end, the missForest algorithm was applied to the whole dataset [18]. Moreover, we applied a rediscretization of some categorical variables to avoid under-represented categories.

### Classification

Different strategies were considered for the elaboration of predictive models in this study. Specifically, 3 of them were explored in detail (Multimedia Appendix 1); that is, logistic regression and 2 machine learning approaches: a decision tree and random forest classifiers.

For model training, the dataset was 10-times divided in (1) a training subset, taking 75% of randomly selected cases, and (2) a validation subset with the remaining 25% of cases. For each

data partition, the model was trained using 4-fold cross-validation on the training subset. As successful cases (ie, survivors not requiring hospital admission) were far superior in number, the effect of class imbalance was reduced by applying a random stratified-sampling strategy [19].

Model performance was assessed by computing the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, and score metrics in the validation subset. Score is a measure of prediction accuracy and is defined as the weighted harmonic mean of the sensitivity and specificity of the model. The final performance of the models was assessed as the average performance of all independent validations.

As indicated above, the methodology was applied to predict 2 types of events: (1) mortality, and (2) in-hospital admission until 30-days after HH/ED discharge. Risk assessment was conducted in 2 different scenarios: (1) at entry into the HH/ED program, and (2) at discharge. Accordingly, the analyses led to 4 different risk models (RM): (1) RM1 accounts for predicting the need for conventional hospitalization at entry into the HH/ED program; (2) RM2 predicts mortality during the study period assessed at entry; (3) RM3 refers to predictive modeling of conventional hospital admissions assessed at HH/ED

discharge; and (4) RM4 predicts mortality during the study period assessed at HH/ED discharge. The risk of mortality and re-admission during HH at entry was not assessed due to the scarcity of unsuccessful cases during HH/ED.

## Results

### Study Population

All 1936 patients admitted to the HH/ED program at HCB during the study period were included in the research. However, the analyses conducted in the study were based on 1925 cases; 4 cases were discarded for having unrecoverable wrong data and 7 for having missing mandatory data. The mean age of the study group was 70.85 (SD 14.88) years; 1201 (62.4%) were men and 724 (37.6%) were women. The list of main diagnoses is depicted in Table 1S in [Multimedia Appendix 1](#). Up to 64 variables, grouped into the 3 categories indicated above, were considered in the analyses (Table 2S in [Multimedia Appendix 1](#)).

To characterize different subpopulations of risk, patients were classified as undergoing successful and unsuccessful home

hospitalization stays based on their re-admission and mortality during the study period and 30 days after hospital discharge ([Tables 1-2](#)). Of the 1925 patients admitted to the HH/ED program, 3 (0.2%) patients died and 96 (5.0%) cases were eventually readmitted to the hospital due to complications of heterogeneous origin during HH/ED. Of the remaining 1922 patients, within 30 days after HH/ED discharge, 37 (1.9%) patients died and 210 (10.9%) cases were identified as falling into the unsuccessful groups when analyzing re-admission risk. [Tables 1](#) and [2](#) summarize the baseline characteristics of both study groups, according to mortality and re-admission, respectively.

Mortality was higher in elderly ( $P<.001$ ) and comorbid patients, GMA ( $P=.02$ ), and the Charlson Comorbidity Index ( $P=.001$ ), especially in those with cardiovascular ( $P<.001$ ) and oncologic disorders ( $P=.019$ ). Mortality was lower in postoperative patients ( $P<.001$ ) and in those with respiratory diseases ( $P=.005$ ). Interestingly, in-hospital re-admission was only slightly associated with higher age ( $P=.003$ ) and a major complexity of comorbid conditions, GMA ( $P<.001$ ), and the Charlson Comorbidity Index ( $P<.001$ ), without well-defined associations with the characteristics of the main diagnosis.

**Table 1.** Clinical characteristics of successful and unsuccessful home hospitalization (HH) cases (n=1925) based on mortality.

Patient characteristics	Successful cases (n=1885)	Unsuccessful cases during HH (n=3)	Unsuccessful cases 30 days after HH discharge (n=37)	P value <sup>a</sup>
Age, mean (SD)	70.7 (14.9)	89.3 (15.1)	77.9 (10.6)	<.001
<b>Sex, n (%)</b>				
Male	1181 (62.7)	1 (33.3)	19 (51.3)	.145
Female	704 (37.3)	2 (66.6)	18 (48.7)	.145
GMA, mean (SD)	21.3 (13.5)	21.4 (3.1)	27.0 (14.2)	.020
Charlson Comorbidity Index, mean (SD)	4.3 (2.8)	5.7 (4.9)	5.8 (2.7)	.001
<b>Diagnostic group, n (%)</b>				
Cardiology	202 (10.7)	1 (33.3)	16 (43.2)	<.001
Respiratory	583 (30.9)	0 (0.0)	5 (13.6)	.005
Oncology	145 (7.7)	0 (0.0)	8 (21.6)	.019
Surgery	375 (19.9)	0 (0.0)	0 (0.0)	<.001
Other medical acute conditions	580 (30.8)	2 (66.7)	8 (21.6)	.440

<sup>a</sup>P values were calculated comparing successful and unsuccessful groups during the full period.



**Table 2.** Clinical characteristics of successful and unsuccessful home hospitalization (HH) cases (n=1925) based on re-admission.

Patient characteristics	Successful cases (n=1638)	Unsuccessful cases during HH (n=96)	Unsuccessful cases 30 days after HH discharge (n=210)	P value <sup>a</sup>
Age, mean (SD)	70.5 (15.2)	72.9 (14.8)	73.2 (11.9)	.003
<b>Sex, n (%)</b>				
Male	1007 (61.6)	63 (65.6)	142 (67.6)	.056
Female	631 (38.4)	33 (34.4)	68 (32.4)	.056
GMA, mean (SD)	20.3 (13.1)	26.8 (15.0)	28.7 (14.7)	<.001
Charlson Comorbidity Index, mean (SD)	4.1 (2.8)	5.3 (2.6)	5.6 (2.6)	<.001
<b>Diagnostic group, n (%)</b>				
Cardiology	162 (9.9)	24 (25.0)	38 (18.1)	.068
Respiratory	507 (30.9)	24 (25.0)	62 (29.5)	.722
Oncology	113 (6.9)	8 (8.3)	32 (15.2)	.123
Surgery	340 (20.8)	14 (14.6)	23 (11.0)	.136
Other medical acute conditions	516 (31.5)	26 (27.1)	55 (26.2)	.450

<sup>a</sup>P values were calculated comparing successful and unsuccessful groups during the full period.

### Predictive Modeling

Different modeling approaches were considered for this purpose, including logistic regression, decision trees, and random forests. The averaged AUROC of each modeling approach that was considered is presented in [Table 3](#).

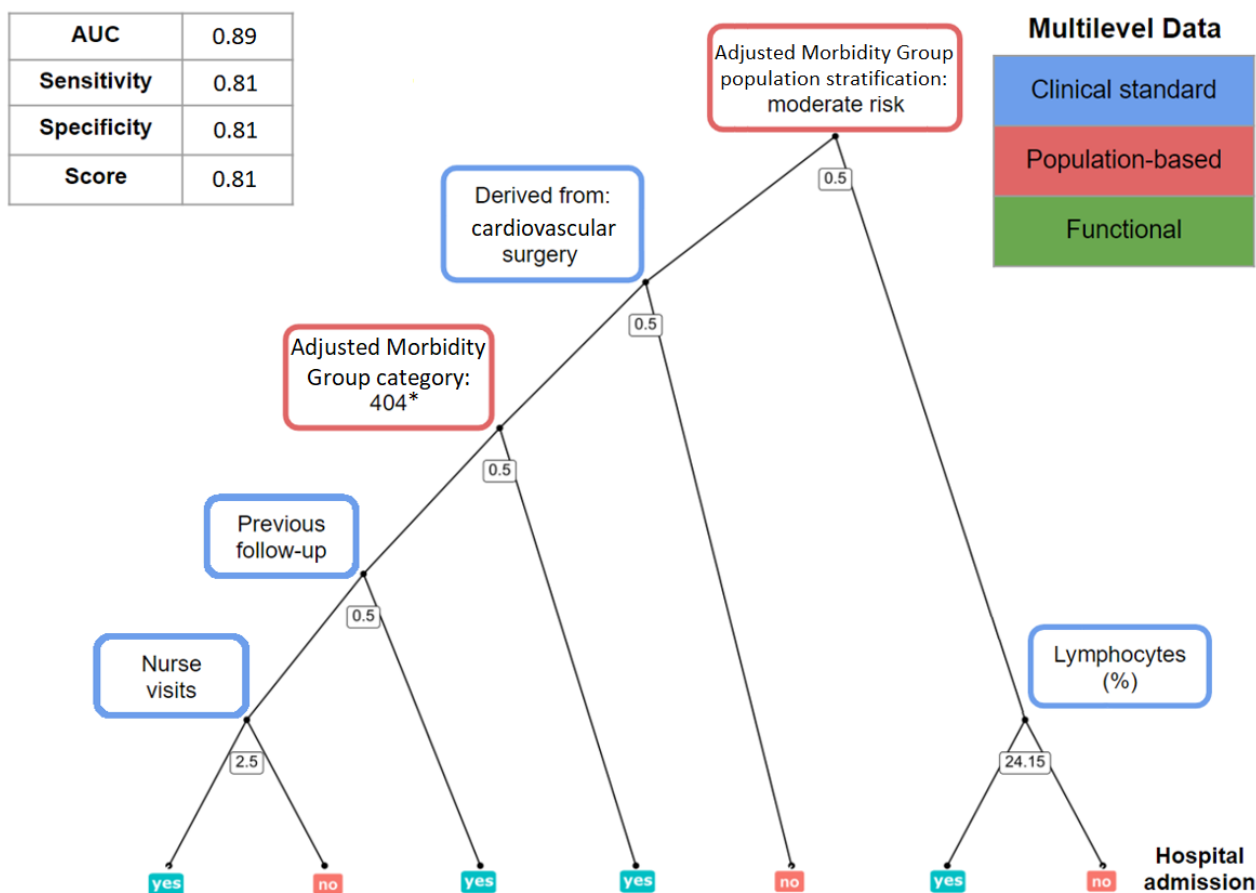
Among the different modeling strategies developed, random forest classifier ([Figure 2](#)) showed the best performance averaged over the 4 risk scenarios.

[Table 4](#) summarizes the performance of the 4 predictive models proposed in the study for in-hospital admission (RM1 and RM3) and for mortality (RM2 and RM4); [Multimedia Appendix 2](#) depicts the relative weight, expressed as the mean decrease in accuracy (MDA) [20], of the 10 most relevant variables for each of the 4 predictive models.

**Table 3.** Area under the receiver operating characteristic curve (AUROC; sensitivity/specificity) performance of the modeling strategies explored.

Model	Mean AUROC (sensitivity/ specificity)	RM1 AUROC (sensitivity/ specificity)	RM2 AUROC (sensitivity/ specificity)	RM3 AUROC (sensitivity/ specificity)	RM4 AUROC (sensitivity/ specificity)
Logistic regression	0.58 (0.54/0.57)	0.65 (0.68/0.58)	0.54 (0.50/0.59)	0.59 (0.61/0.52)	0.54 (0.38/0.58)
Decision tree	0.59 (0.81/0.47)	0.62 (0.82/0.43)	0.64 (0.88/0.51)	0.57 (0.64/0.42)	0.64 (0.88/0.52)
Random forest	0.80 (0.75/0.71)	0.71 (0.67/0.64)	0.88 (0.81/0.76)	0.70 (0.71/0.61)	0.89 (0.81/0.81)

**Figure 2.** Overview of the predictive modeling strategy taking, as an example, prediction of re-admission at home hospitalization discharge. Upper-left table: metrics used for model performance assessment; AUC: area under the receiver operating characteristic curve. Center figure: representation of 1 decision tree using a random subset of features; on the nodes, threshold values for each variable determine the path from the root to the leaves (0.5 for Boolean variables), moving toward the left when the decision rule is met; on a random forest model, final predictions are averaged over multiple decision trees. Upper-right table: 3 categories of data that are included in the models. \*GMA category 404; 40: patient with active neoplasms; 4: high complexity conditions (percentile between 0.85 and 0.95).



**Table 4.** Average results of the performance of the 4 home-hospitalization/early discharge (HH/ED) predictive risk models (RM).

Model	AUROC <sup>a</sup> , mean (SD)	Sensitivity, mean (SD)	Specificity, mean (SD)	Score, mean (SD)
Readmission risk at HH/ED admission (RM1)	0.71 (0.03)	0.67 (0.06)	0.64 (0.05)	0.66 (0.03)
Readmission risk at HH/ED discharge (RM3)	0.70 (0.02)	0.71 (0.06)	0.61 (0.05)	0.66 (0.03)
Mortality risk at HH/ED admission (RM2)	0.88 (0.04)	0.81 (0.09)	0.76 (0.04)	0.78 (0.06)
Mortality risk at HH/ED discharge (RM4)	0.89 (0.04)	0.81 (0.12)	0.81 (0.05)	0.81 (0.06)

<sup>a</sup>AUROC: area under the receiver operating characteristic curve.

For risk of in-hospital admissions (Multimedia Appendix 2, panels A and C), multimorbidity (expressed as GMA scoring) showed the highest predictive impact, followed by red cell distribution width (RDW). Other top predictors were polypharmacy, body mass index (BMI), a few biological variables (blood cells characteristics and glucose), and physical and mental status.

For risk of mortality (Multimedia Appendix 2, panels B and D), RDW and physical status at entry (assessed using the SF-36 questionnaire [21]) showed the highest impact in the models.

Notably, enriching the model with information acquired during HH/ED (Multimedia Appendix 2, panels C and D), several

variables gained importance, such as hospital admissions during HH/ED, length of current hospitalization period, and nursing home visits.

## Discussion

### Principal Findings

The current research has developed and internally validated 4 machine learning algorithms predicting the risk of in-hospital admission and mortality for patients undergoing home-based hospitalization until 30-days after discharge from the service at HCB, from 2009 to 2015. Predictions of the 2 undesirable

events were performed at 2 specific time points: at entry and at discharge from home-based hospitalization.

The study design was formulated and adopted under the hypothesis that robust predictions could be useful for clinical decision making: (1) to decide patients' admission into the HH/ED service (RM1 and RM2); and (2) to personalize care paths for transitional care, as well as for enhanced vertical integration between specialized care and community-based services, both at patients' discharge from HH.

A unique aspect of this research is that predictors considered in the analyses encompass 3 different categories of variables (Table 2S in [Multimedia Appendix 1](#)): (1) clinical data and biological information [22-24] extracted from patients' electronic medical records; (2) additional variables often not considered in the clinical records specifically collected in the research protocol to reflect patients' functional capacities and health care resources; and (3) information from GMA, the population-based, health-risk assessment tool developed and implemented in Catalonia (ES) [15,16,25].

We understand that the multisource approach adopted in this research was the most appropriate to elaborate predictive modeling in a highly heterogeneous group of patients undergoing HH/ED, in terms of clinical diagnosis and frailty status [8]. The results depicted in [Table 4](#), in terms of AUROC and score values, indicate the reasonably good performance of the predictive models as compared to recent studies on similar scenarios [26], demonstrating the feasibility of the proposed approach and leveraging the advantages of applying machine learning in clinical risk prediction contexts in front of more traditional approaches based on standard multiple regression analyses [27]. Moreover, [Multimedia Appendix 2](#) (panels A-D) shows a high relative contribution of variables usually not considered to be of clinical standard or relevant biological information recorded in the EMR. Overall, our results indicate that our multisource approach significantly contributes to enhanced health risk assessment with a potentially high impact on clinical decision support.

### Limitations of the Study and Lessons Learned for Clinical Application

We have not been able to identify literature on predictive modeling specifically addressing HH/ED. It may partly be due to the heterogeneity of orientations and characteristics of the ongoing HH/ED programs among sites. This fact constitutes a limitation regarding the potential for generalization of the results of this research to other sites. However, we understand that the multisource approach undertaken in this study shows enormous potential for risk assessment regarding mortality and early re-admissions of hospitalizations in general, and may show high applicability beyond the field of HH/ED. The predictive modeling undertaken in the study should be useful for defining the characteristics of personalized care paths of transitional care after hospital discharge. As indicated above, the results can have a high impact on shaping the interactions between specialized and community-based care in patients with high risk for hospital re-admissions.

A major general limitation of machine learning approaches such as the one proposed here is the fact that they can be considered "black-box" solutions, difficult to interpret by clinicians. Our work, however, is based on random forest models that provide interpretable information regarding variable importance ([Multimedia Appendix 2](#), panels A-D) and even model visualization, thus facilitating the understanding of their predictions. We believe that the clinical interpretation of the predictors may require different approaches; for example, variables like age and diagnosis should be individually assessed for clinical judgment, while others, like the different GMA parametrization (including the Charlson Index), should be assessed by taking the category as a whole (and likewise, abnormalities in some blood test variables). On the other hand, this study indicates that the impact of patients' functional status on outcomes is high. However, some of the measurements included in this category are not scalable in the clinical scenario (ie, SF36). Therefore, our results clearly indicate that surrogates with higher applicability [28,29] should likely be considered for inclusion in real-life clinical settings. This could be achieved through patients' self-tracking equipment (ie, apps) that provides information on different dimensions characterizing the functional performance of the patient, namely physical and psychological status, wellbeing, activation, etc.

It is acknowledged that the generalization of the use of new clinical scores generated from predictive modeling needs external validation on other patient cohorts or in different timeframes, and even on the development of impact studies in real-world settings [30]. Apart from being costly, such a validation process can show limitations partly due to rapidly evolving clinical environments, as is the case for HH/ED at HCB, expanded to the entire health district of Eixample-Esquerra during 2018. The new scenario implies great changes in the clinical environment, patients' characteristics, and data sources prompting the need for designing dynamic models in the context of learning health systems (LHS) [31,32]. It is of note that within a mature digital health scenario, the multisource predictive modeling approach could be enriched with other sources of data, such as patient self-reported data and data from social care. The lack of digital maturity of the current ecosystems constitutes a limiting factor for now, but in the near future, risk assessment tools are expected to improve in terms of robustness, potential for generalization of the results, and incorporation of a dynamic predictive approach.

### Steps Toward Dynamic Learning Health Systems

There is little doubt about the high potential shown by the digital transformation of health as part of a large-scale adoption of integrated care. It is acknowledged, however, that practical applications of this vision face major limitations when it comes to accessing and mining health data stored in distributed silos of information. However, it seems clear that integrating and analyzing highly complex data would open new avenues for digital health in the clinical arena.

The integration of biomedical research information systems with in-place electronic health records in hospitals and in primary care centers having interoperability with patients' self-tracking information would enable the development of

innovative, dynamic predictive modeling approaches, opening up entirely new and fascinating scenarios for an interplay between clinical practice and biomedical research [33,34]. We have identified 4 main interrelated enablers of this scenario [15,17,35]: (1) cloud-based tools and services allowing secure analysis of patient-centric distributed and multi-disciplinary health-related information; (2) systems medicine approaches to generate clinical predictive modeling to feed clinical decision support systems and patient decision support systems; (3) implementation and evaluation strategies for real-world implementation and assessment of cloud-based services, and (4) governance, regulatory aspects, and service adoption throughout the health care systems; these are all key to harnessing the strengths and opportunities of LHS.

Combined actions involving organizational changes with the engagement of all stakeholders, selective adoption of novel biomedical and digital tools, and the achievement of financial sustainability through enhanced accountability and entrepreneurial actions should pave the way toward the transition to LHS.

## Conclusions

This study proves the potential of the proposed multisource machine-learning models for the prediction of risk of re-admissions and deaths in patients undergoing home-based hospitalization in a real-world setting. Further steps beyond this study include the development of dynamic clinical decision support systems allowing progression towards sustainable patient-centered health care services.

## Acknowledgments

This project was funded by Fondo de Investigaciones Sanitarias (FIS), Instituto Carlos III, Proyecto: Smart PITES PI18/00841.

This work was supported by NEXTCARE (COMRD15-1-0016-2016), the CERCA Programme, the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 01770), MCIU/AEI/FEDER (RTI2018-098472-B-I00), the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement (no. 712754), and the Spanish Ministry of Economy and Competitiveness under the Severo Ochoa grant (SEV-2014-0425; 2015-2019).

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

On-line supplementary material: Multisource Predictive Modelling of Health Outcomes from Home Hospitalization. [DOCX File, 333 KB - [jmir\\_v22i10e21367\\_app1.docx](#)]

### Multimedia Appendix 2

Relative importance measures based on the mean decrease in accuracy (MDA) of the 10 most relevant variables for each model. GMA: morbidity adjusted group; BMI: body mass index; RDW: red cell distribution width; HH: home hospitalization. [DOCX File, 17 KB - [jmir\\_v22i10e21367\\_app2.docx](#)]

## References

1. Leff B, Burton L, Mader SL, Naughton B, Burl J, Inouye SK, et al. Hospital at Home: Feasibility and Outcomes of a Program To Provide Hospital-Level Care at Home for Acutely Ill Older Patients. *Ann Intern Med* 2005 Dec 06;143(11):798. [doi: [10.7326/0003-4819-143-11-200512060-00008](#)]
2. Caplan GA, Sulaiman NS, Mangin DA, Aimonino Ricauda N, Wilson AD, Barclay L. A meta - analysis of "hospital in the home". *Medical Journal of Australia* 2012 Nov 05;197(9):512-519. [doi: [10.5694/mja12.10480](#)]
3. Cryer L, Shannon SB, Van Amsterdam M, Leff B. Costs for 'hospital at home' patients were 19 percent lower, with equal or better outcomes compared to similar inpatients. *Health Affairs (Millwood)* 2012 Jun;31(6):1237-1243. [doi: [10.1377/hlthaff.2011.1132](#)] [Medline: [22665835](#)]
4. Jacobs JM, Hammerman-Rozenberg R, Stessman J. Home Hospitalization: 15 Years of Experience. *Annals of Internal Medicine* 2006 Mar 21;144(6):456. [doi: [10.7326/0003-4819-144-6-200603210-00023](#)]
5. Jacobs JM, Cohen A, Rozengarten O, Meiller L, Azoulay D, Hammerman-Rozenberg R, et al. Closure of a home hospital program: Impact on hospitalization rates. *Archives of Gerontology and Geriatrics* 2007 Sep;45(2):179-189. [doi: [10.1016/j.archger.2006.10.009](#)]
6. Montalto M. The 500 - bed hospital that isn't there: the Victorian Department of Health review of the Hospital in the Home program. *Medical Journal of Australia* 2010 Nov 15;193(10):598-601. [doi: [10.5694/j.1326-5377.2010.tb04070.x](#)]
7. Shepperd S, Iliffe S, Doll HA, Clarke MJ, Kalra L, Wilson AD, et al. Admission avoidance hospital at home. *Cochrane Database Systematic Reviews* 2016;4:---. [doi: [10.1002/14651858.cd007491.pub2](#)]

8. Hernández C, Aibar J, Seijas N, Puig I, Alonso A, Garcia-Aymerich J, et al. Implementation of Home Hospitalization and Early Discharge as an Integrated Care Service: A Ten Years Pragmatic Assessment. *International Journal of Integrated Care* 2018 May 16;18:12. [doi: [10.5334/ijic.3431](https://doi.org/10.5334/ijic.3431)]
9. Leff B, Montalto M. Home hospital-toward a tighter definition. *Journal of the American Geriatrics Society* 2004 Nov 18;52(12):2141. [doi: [10.1111/j.1532-5415.2004.52579.1.x](https://doi.org/10.1111/j.1532-5415.2004.52579.1.x)]
10. Gonçalves-Bradley D, Iliffe S, Doll H, Broad J, Gladman J, Langhorne P, et al. Early discharge hospital at home. *Cochrane Database of Systematic Reviews* 2017;2017(6):1465. [doi: [10.1002/14651858.cd000356.pub4](https://doi.org/10.1002/14651858.cd000356.pub4)]
11. Fetter R, Shin Y, Freeman J, Averill R, Thompson J. Case mix definition by diagnosis-related groups. *Med Care* 1980 Feb;18(2 Suppl):iii, 1-iii,53. [Medline: [7188781](https://pubmed.ncbi.nlm.nih.gov/7188781/)]
12. Yang H, Dervin G, Madden S, Beaulé PE, Gagné S, Crossan ML, et al. Postoperative Home Monitoring After Joint Replacement: Feasibility Study. *JMIR Perioperative Medicine* 2018 Sep 05;1(2):e10168. [doi: [10.2196/10168](https://doi.org/10.2196/10168)]
13. Lopez KN, O'Connor M, King J, Alexander J, Challman M, Lovick DK, et al. Improving Transitions of Care for Young Adults With Congenital Heart Disease: Mobile App Development Using Formative Research. *JMIR Formative Research* 2018 Sep 11;2(2):e16. [doi: [10.2196/formative.9963](https://doi.org/10.2196/formative.9963)]
14. Taylor A, Broadbent M, Wallis M, Marsden E. The predictive validity of the interRAI ED screener for predicting re-presentation within 28 days for older adults at a regional hospital emergency department. *Australasian Emergency Care* 2019 Sep;22(3):149-155. [doi: [10.1016/j.auec.2019.04.005](https://doi.org/10.1016/j.auec.2019.04.005)]
15. Dueñas-Espín I, Vela E, Pauws S, Bescos C, Cano I, Cléries M, et al. Proposals for enhanced health risk assessment and stratification in an integrated care scenario. *BMJ Open* 2016 Apr;6(4):e010301. [doi: [10.1136/bmjopen-2015-010301](https://doi.org/10.1136/bmjopen-2015-010301)] [Medline: [27084274](https://pubmed.ncbi.nlm.nih.gov/27084274/)]
16. Monterde D, Vela E, Cléries M, Garcia-Eroles L, Roca J, Pérez-Sust P. Multimorbidity as a predictor of health service utilization in primary care: a registry-based study of the Catalan population. *BMC Family Practice* 2020 Feb 17;21(1):39. [doi: [10.1186/s12875-020-01104-1](https://doi.org/10.1186/s12875-020-01104-1)]
17. Cano I, Tenyi A, Vela E, Miralles F, Roca J. Perspectives on Big Data applications of health information. *Current Opinion in Systems Biology* 2017 Jun;3:36-42. [doi: [10.1016/j.coisb.2017.04.012](https://doi.org/10.1016/j.coisb.2017.04.012)]
18. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2011 Oct 28;28(1):112-118. [doi: [10.1093/bioinformatics/btr597](https://doi.org/10.1093/bioinformatics/btr597)]
19. Calvo M, Cano I, Hernandez C. Class Imbalance Impact on the Prediction of Complications during Home Hospitalization: A Comparative Study. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. Institute of Electrical and Electronics Engineers Inc; 2019 Oct 07 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 23-27 July 2019; Berlin, Germany p. A.* [doi: [10.1109/embc.2019.8857746](https://doi.org/10.1109/embc.2019.8857746)]
20. Breiman L. Random forests. *Machine Learning* 2001 Oct;45:5-32. [doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)]
21. Alonso J, Prieto L, Antó J, Alonso J, Antó J. La versión española del SF-36 HealthSurvey (Cuestionario de salud SF36): un instrumento para la medida de los resultados clínicos. *Medicina Clínica (Barcelona)* 1995;104 [FREE Full text]
22. Ryckwaert F, Boccara G, Frappier JM, Colson PH. Incidence, risk factors, and prognosis of a moderate increase in plasma creatinine early after cardiac surgery\*. *Critical Care Medicine* 2002;30(7):1495-1498. [doi: [10.1097/00003246-200207000-00016](https://doi.org/10.1097/00003246-200207000-00016)]
23. Yamanaka T, Matsumoto S, Teramukai S, Ishiwata R, Nagai Y, Fukushima M. The Baseline Ratio of Neutrophils to Lymphocytes Is Associated with Patient Prognosis in Advanced Gastric Cancer. *Oncology* 2007;73(3-4):215-220. [doi: [10.1159/000127412](https://doi.org/10.1159/000127412)]
24. González-Ferrer JJ, García-Rubira JC, Balcones DV, Gil IN, Barrio RC, Fuentes-Ferrer M, et al. Influence of Hemoglobin Level on In-Hospital Prognosis in Patients With Acute Coronary Syndrome. *Revista Española de Cardiología (English Edition)* 2008 Jan;61(9):945-952. [doi: [10.1016/s1885-5857\(08\)60255-8](https://doi.org/10.1016/s1885-5857(08)60255-8)]
25. Monterde D, Vela E, Cléries M, grupo colaborativo GMA. Adjusted morbidity groups: A new multiple morbidity measurement of use in Primary Care. *Atención Primaria* 2016 Dec;48(10):674-682 [FREE Full text] [doi: [10.1016/j.aprim.2016.06.003](https://doi.org/10.1016/j.aprim.2016.06.003)] [Medline: [27495004](https://pubmed.ncbi.nlm.nih.gov/27495004/)]
26. Min X, Yu B, Wang F. Predictive Modeling of the Hospital Readmission Risk from Patients' Claims Data Using Machine Learning: A Case Study on COPD. *Scientific Reports* 2019 Feb 20;9(1):2362. [doi: [10.1038/s41598-019-39071-y](https://doi.org/10.1038/s41598-019-39071-y)]
27. Reilly BM, Evans AT. Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions. *Annals of Internal Medicine* 2006 Feb 07;144(3):201. [doi: [10.7326/0003-4819-144-3-200602070-00009](https://doi.org/10.7326/0003-4819-144-3-200602070-00009)]
28. Foebel AD, Hirdes JP, Heckman GA, Kergoat M, Patten S, Marrie RA. Diagnostic data for neurological conditions in interRAI assessments in home care, nursing home and mental health care settings: a validity study. *BMC Health Services Research* 2013 Nov 1;13(1):457. [doi: [10.1186/1472-6963-13-457](https://doi.org/10.1186/1472-6963-13-457)]
29. Berwick DM, Nolan TW, Whittington J. The Triple Aim: Care, Health, And Cost. *Health Affairs* 2008 May 01;27(3):759-769. [doi: [10.1377/hlthaff.27.3.759](https://doi.org/10.1377/hlthaff.27.3.759)]
30. Bleeker S, Moll H, Steyerberg E, Donders A, Derksen-Lubsen G, Grobbee D, et al. External validation is necessary in prediction research: A clinical example. *Journal of Clinical Epidemiology* 2003 Sep;56(9):826-832. [doi: [10.1016/s0895-4356\(03\)00207-5](https://doi.org/10.1016/s0895-4356(03)00207-5)]



31. Smith M, Saunders R, Stuckhardt L, McGinnis J. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, D.C: National Academies Press; 2013.
32. Maddox TM, Albert NM, Borden WB, Curtis LH, Ferguson TB, Kao DP, et al. The Learning Healthcare System and Cardiovascular Care: A Scientific Statement From the American Heart Association. *Circulation* 2017 Apr 04;135(14):e826-e857. [doi: [10.1161/CIR.0000000000000480](https://doi.org/10.1161/CIR.0000000000000480)] [Medline: [28254835](https://pubmed.ncbi.nlm.nih.gov/28254835/)]
33. Auffray C, Charron D, Hood L. Predictive, preventive, personalized and participatory medicine: back to the future. *Genome Medicine* 2010;2(8):57. [doi: [10.1186/gm178](https://doi.org/10.1186/gm178)]
34. Aarestrup FM, Albeyatti A, Armitage WJ, Auffray C, Augello L, Balling R, et al. Towards a European health research and innovation cloud (HRIC). *Genome Medicine* 2020 Feb 19;12(1):18. [doi: [10.1186/s13073-020-0713-z](https://doi.org/10.1186/s13073-020-0713-z)]
35. Habl C, Renner AT, Bobek J, Laschkolnig A. Study on Big Data in Public Health, Telemedicine and Healthcare. European Commission. 2016 Dec. URL: [https://ec.europa.eu/health/sites/health/files/ehealth/docs/bigdata\\_report\\_en.pdf](https://ec.europa.eu/health/sites/health/files/ehealth/docs/bigdata_report_en.pdf) [accessed 2020-09-25]

## Abbreviations

**AUROC:** area under the receiver operating characteristic curve

**ED:** early discharge

**EMR:** electronic medical records

**GMA:** Adjusted Morbidity Groups

**HCB:** Hospital Clínic de Barcelona

**HH:** Home Hospitalization

**LHS:** Learning Health Systems

**MDA:** mean decrease in accuracy

**RDW:** red cell distribution width

**RM:** risk model

*Edited by CL Parra-Calderón; submitted 12.06.20; peer-reviewed by J Piera-Jiménez, J de Batlle; comments to author 08.08.20; revised version received 20.08.20; accepted 08.09.20; published 07.10.20.*

### *Please cite as:*

Calvo M, González R, Seijas N, Vela E, Hernández C, Batiste G, Miralles F, Roca J, Cano I, Jané R

Health Outcomes from Home Hospitalization: Multisource Predictive Modeling

*J Med Internet Res* 2020;22(10):e21367

URL: <http://www.jmir.org/2020/10/e21367/>

doi: [10.2196/21367](https://doi.org/10.2196/21367)

PMID: [33026357](https://pubmed.ncbi.nlm.nih.gov/33026357/)

©Mireia Calvo, Rubèn González, Núria Seijas, Emili Vela, Carme Hernández, Guillem Batiste, Felip Miralles, Josep Roca, Isaac Cano, Raimon Jané. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 07.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Impact of Electronic Health Record Interface Design on Unsafe Prescribing of Ciclosporin, Tacrolimus, and Diltiazem: Cohort Study in English National Health Service Primary Care

Brian MacKenna<sup>1</sup>, MPH, MPharm; Sebastian Bacon<sup>1</sup>, BA; Alex J Walker<sup>1</sup>, PhD; Helen J Curtis<sup>1</sup>, DPhil; Richard Croker<sup>1</sup>, MSc; Ben Goldacre<sup>1</sup>, BA, MA, MSc, MBBS, MRCPsych

The DataLab, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

**Corresponding Author:**

Ben Goldacre, BA, MA, MSc, MBBS, MRCPsych

The DataLab

Nuffield Department of Primary Care Health Sciences

University of Oxford

Radcliffe Observatory Quarter

Woodstock Road

Oxford, OX2 6GG

United Kingdom

Phone: 44 1865289313

Email: [ben.goldacre@phc.ox.ac.uk](mailto:ben.goldacre@phc.ox.ac.uk)

## Abstract

**Background:** In England, national safety guidance recommends that ciclosporin, tacrolimus, and diltiazem are prescribed by brand name due to their narrow therapeutic windows and, in the case of tacrolimus, to reduce the chance of organ transplantation rejection. Various small studies have shown that changes to electronic health record (EHR) system interfaces can affect prescribing choices.

**Objective:** Our objectives were to assess variation by EHR systems in breach of safety guidance around prescribing of ciclosporin, tacrolimus, and diltiazem, and to conduct user-interface research into the causes of such breaches.

**Methods:** We carried out a retrospective cohort study using prescribing data in English primary care. Participants were English general practices and their respective EHR systems. The main outcome measures were (1) the variation in ratio of safety breaches to adherent prescribing in all practices and (2) the description of observations of EHR system usage.

**Results:** A total of 2,575,411 prescriptions were issued in 2018 for ciclosporin, tacrolimus, and diltiazem (over 60 mg); of these, 316,119 prescriptions breached NHS guidance (12.27%). Breaches were most common among users of the EMIS EHR system (breaches in 18.81% of ciclosporin and tacrolimus prescriptions and in 17.99% of diltiazem prescriptions), but breaches were observed in all EHR systems.

**Conclusions:** Design choices in EHR systems strongly influence safe prescribing of ciclosporin, tacrolimus, and diltiazem, and breaches are prevalent in general practices in England. We recommend that all EHR vendors review their systems to increase safe prescribing of these medicines in line with national guidance. Almost all clinical practice is now mediated through an EHR system; further quantitative research into the effect of EHR system design on clinical practice is long overdue.

(*J Med Internet Res* 2020;22(10):e17003) doi:[10.2196/17003](https://doi.org/10.2196/17003)

**KEYWORDS**

prescribing; primary care; electronic health records; clinical software; branded prescribing; diltiazem; tacrolimus; ciclosporin

## Introduction

Over 1.1 billion prescriptions are issued through primary care in England each year, at a cost of £8.8 billion in 2018 [1]. The overwhelming majority of these prescriptions are generated

through an electronic health record (EHR) system [2], where the clinician selects a medicine from a “picking list,” specifies directions (eg, “once daily”), and then signs the prescription either electronically or on a printed copy. All National Health Service (NHS) practices use one of four EHRs made available through NHS Digital [3]. We have previously described a small

design choice in an EHR medicines selection screen that costs the NHS approximately £9.5 million in one year across a wide range of drugs [4,5].

Tacrolimus and ciclosporin are used in organ transplantation and other conditions, such as rheumatoid arthritis, psoriasis, and severe atopic dermatitis. Both medicines have a narrow therapeutic window: minor differences in blood levels have the potential to cause graft rejection reactions, and switching between tacrolimus products has been associated with reports of toxicity and graft rejection. Therefore, the Medicines and Healthcare products Regulatory Agency (MHRA) recommends that tacrolimus and ciclosporin are prescribed and dispensed by brand name [6,7]. Similarly, diltiazem—a calcium channel blocker commonly used for conditions such as angina and mild-to-moderate hypertension—should be prescribed and dispensed by brand for preparations containing greater than 60 mg, as different brands of modified-release formulations with over 60 mg diltiazem may not have the same clinical effect [8].

Our team delivers OpenPrescribing.net, a publicly funded and openly accessible data explorer for NHS primary care prescribing, with 135,000 unique users in the past year. OpenPrescribing.net supports bespoke data queries alongside numerous predefined standard measures for safety, cost, and effectiveness, with data shared for every practice in England. Two of these standard measures assess compliance with the guidance to prescribe a branded preparation of ciclosporin, tacrolimus, and diltiazem (over 60 mg). Through our prior knowledge around EHR design, as clinicians and researchers engaged with software development, we suspected there may be a relationship between choice of EHR system and breaches of this safety guidance.

We therefore set out to describe and map variation between practices and their parent clinical commissioning groups (CCGs) in breaches of safe prescribing guidance for ciclosporin, tacrolimus, and diltiazem (over 60 mg), and assess the impact of EHR choice on the proportion of safety breaches. Having identified one EHR system as being strongly associated with safety breaches, we conducted user testing to explore the causes of this association. We aimed to produce a rapid report to

support further investigations and any required software modifications to the EHRs, in light of the patient safety risk.

Methods

Study Design

This study is a retrospective cohort study in prescribing data from all English NHS general practices and CCGs, complemented with data on EHR deployment from NHS Digital, as well as in user testing of two commonly used EHRs by a senior pharmacist.

Data Sources

We extracted data from the OpenPrescribing.net database. This imports openly accessible prescribing data from the large monthly files published by the NHS Business Services Authority, which contain data on cost and items prescribed for each month, for every typical general practice and CCG in England since mid-2010 until May 2019 [9]. The monthly prescribing datasets contain one row for each different medication and dose, in each prescribing organization in NHS primary care in England, describing the number of items (ie, prescriptions issued) and the total cost. These data are sourced from community pharmacy claims data and, therefore, contain all items that were dispensed. We extracted all available data for typical general practices, excluding other organizations such as prisons and hospitals, according to the NHS Digital dataset of practice characteristics and excluded practices that had not prescribed at least one item per measure. We excluded all other organizations such as prisons and hospitals. Data on which EHR was deployed in each practice were extracted from a monthly file that is circulated by NHS Digital to interested parties and is available on request [9].

Prevalence of Safety Breaches by General Practices in Implementation of Brand Prescribing of Ciclosporin, Tacrolimus, and Diltiazem

We measured the number of safety breaches and created practice-level deciles at each month for the measure [10] (see Table 1) of proportion of brand prescribing of ciclosporin, tacrolimus, and diltiazem (over 60 mg).

Table 1. OpenPrescribing measures of generic ciclosporin, tacrolimus, and diltiazem prescribing.

Measure	Definition (full technical definitions are available on GitHub [10])
Ciclosporin and tacrolimus	Total items of generic ciclosporin and tacrolimus preparations, as a proportion of total items of all ciclosporin and tacrolimus items. These medicines are grouped due to them having similar uses and covered by similar safety alerts by the Medicines and Healthcare products Regulatory Agency (MHRA).
Diltiazem	Total items of generic diltiazem modified-release preparations, as a proportion of total items of all diltiazem modified-release items (over 60 mg).

Variation in Safety Breaches by Electronic Health Record

We measured the proportion of safety breaches for prescribing of ciclosporin, tacrolimus, and diltiazem for each of the four principal EHRs.

Influence of Principal Electronic Health Record on Breaches of Safety Guidance

We conducted mixed-effects logistic regression to determine the effect size of the EHR used and the extent to which CCG membership and other factors affected these estimates. The independent variable was each prescription as a binary choice between generic and branded, while the main fixed-effect variable was EHR vendor, with CCG membership as a random

effect. Other practice factors were selected *a priori* based on their reasonable availability, as well as clinical judgement. These were as follows: percentage of patients over 65 years of age, percentage of patients under 18 years of age, percentage of patients with a long-term health condition, dispensing practice status, single-handed practice status, ruralness or urbanness of practice location, and index of multiple deprivation. Where data for predictor variables were missing, these practices were excluded from the regression.

### Electronic Health Record System User-Interface Evaluation

One senior pharmacist issued prescriptions in the EMIS and SystmOne computer systems to a test patient and observed the prompts. These two systems hold 95% of the EHR-use market share in general practices in England.

### Software and Reproducibility

Data management was performed using Python, version 3.7 (Python Software Foundation), and Google BigQuery, with analysis carried out using Stata 13.1 (StataCorp) and Python. Data, as well as all code for data management and analysis, is archived online in a public, open access repository on GitHub [11].

### Patient and Public Involvement

Our website, OpenPrescribing.net, is an openly accessible data explorer for all NHS England primary care prescribing data, which receives a large volume of user feedback from professionals, patients, and the public. This feedback is used to refine and prioritize our informatics tools and research activities. Patients were not formally involved in developing this specific study design.

### Ethical Approval

This study uses open, publicly available data and data that are publicly available from NHS Digital on request; therefore, no ethical approval was required.

## Results

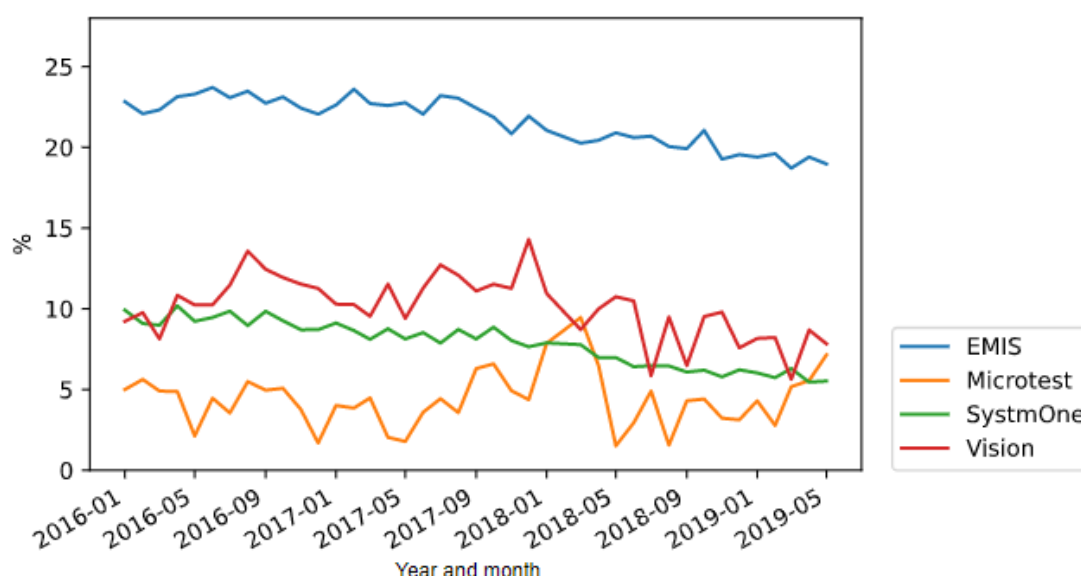
### Prevalence of Safety Breaches

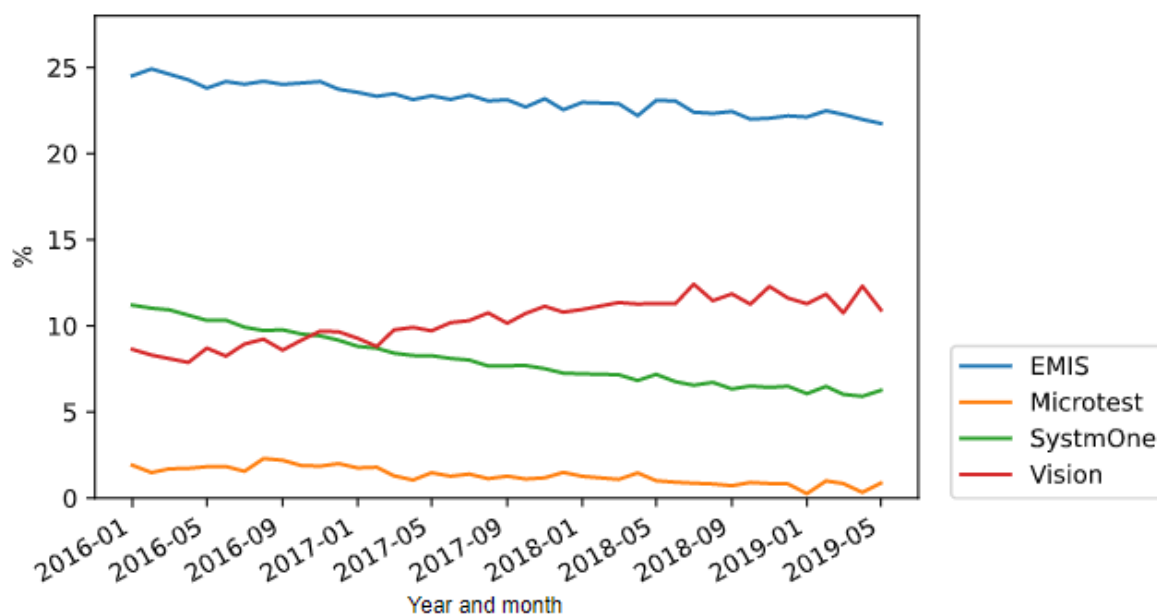
In 2018, 316,119 of the 2,575,411 (12.27%) prescriptions for ciclosporin, tacrolimus, and diltiazem (over 60 mg) breached prescribing safety guidance. Of practices that prescribed ciclosporin or tacrolimus at least once, 2241 out of 5439 (41.20%) breached safety guidance, while 5777 practices out of 7186 (80.39%) breached safety guidance with regard to diltiazem. The ratio of safety breaches as a proportion of all prescribing for these items for practices was 12.28% (292,331/2,380,128) (10th-90th percentile range = 0%-47%) for diltiazem and 12.18% (23,788/195,283) (10th-90th percentile range = 0%-67%) for ciclosporin and tacrolimus.

### Variation in Safety Breaches by Electronic Health Record

The mean values for the ciclosporin and tacrolimus safety breaches and for the diltiazem safety breaches as a proportion of all prescribing for the respective items from January 2016 to May 2019 are shown in Figures 1 and 2, respectively. The rate of breaches was consistently highest in EMIS practices, and in 2018, 18.81% (16,521/87,828) of all ciclosporin and tacrolimus prescriptions issued from the EMIS EHR system breached safety prescribing safety guidance, while 17.99% (209,371/1,163,578) of all diltiazem (over 60 mg) prescriptions breached safety guidance.

**Figure 1.** Mean values for the ciclosporin and tacrolimus safety breaches as a proportion of all prescribing by electronic health record (EHR).



**Figure 2.** Mean values for the diltiazem safety breaches as a proportion of all prescribing by electronic health record (EHR).

The results of the logistic regression are shown in [Table 2](#). When generating a diltiazem (over 60 mg) prescription, practices using the EMIS EHR system were around 4 times more likely to prescribe in breach of guidance with regard to diltiazem than practices using SystmOne, with a similar size of effect for ciclosporin and tacrolimus breaches. Practices using Microtest and Vision EHR systems were substantially less likely to breach safety guidance, although these practices make up a relatively

small proportion of practices. Adjusting for CCG membership as a random effect had little effect on these estimates, although CCG membership was still responsible for a high proportion of the variation in propensity to prescribe generically: 16.4% for diltiazem and 8.4% for ciclosporin. Other factors had even less of an effect on the estimates, though some factors were weakly associated with the propensity to prescribe in breach of guidance (see [Multimedia Appendix 1](#)).

**Table 2.** Ciclosporin, tacrolimus, and diltiazem prescribing measures stratified by electronic health record (EHR) system along with odds ratios (ORs) from a univariable and multivariable logistic regression model.

Medicine and EHR systems	Breaches of safety guidance (%), mean	Univariable logistic regression, OR (95% CI)	Mixed-effects logistic regression <sup>a</sup> , OR (95% CI)
<b>Diltiazem</b>			
EMIS	22.7	Reference	Reference
Microtest	1.0	0.045 (0.039-0.052)	0.051 (0.044-0.060)
SystmOne	7.0	0.252 (0.249-0.254)	0.239 (0.235-0.243)
Vision	12.0	0.462 (0.451-0.473)	0.300 (0.290-0.309)
<b>Ciclosporin and tacrolimus</b>			
EMIS	23.2	Reference	Reference
Microtest	6.1	0.103 (0.072-0.148)	0.161 (0.107-0.242)
SystmOne	7.8	0.238 (0.230-0.247)	0.271 (0.256-0.287)
Vision	10.8	0.378 (0.346-0.412)	0.219 (0.195-0.247)

<sup>a</sup>Adjusted for clinical commissioning group (CCG) membership as a random effect, percentage of patients over 65 years of age, percentage of patients under 18 years of age, percentage of patients with a long-term health condition, dispensing practice status, single-handed practice status, ruralness or urbanness of location, and index of multiple deprivation.



## Electronic Health Record System User-Interface Evaluation

### EMIS

On reviewing the medicines-selection screen, it was observed that when searching for a medicine by its brand name, EMIS always presents the generic version of the medicine as priority in the medicines-picking list; this represents a breach of safety guidance for ciclosporin, tacrolimus, and diltiazem (over 60 mg). Warnings are presented to prescribers to prescribe by brand when a generic version of these medicines is selected; however, these warnings appear alongside multiple other warnings such as interactions with other medications. These warnings take the form of a *pop-up* box, which can be easily overridden allowing the generic version of all three medicines to be issued.

### SystmOne

When the search terms “ciclosporin,” “tacrolimus,” or “diltiazem” were entered, no results were returned, encouraging a user to search by brand names in line with safety guidance. However, in another part of the user-interface screen there is a separate tick-box, “non-prescribable.” When ticked by a user, input of the search terms “ciclosporin,” “tacrolimus,” or “diltiazem” does return results, allowing selection of the generic medicine in breach of safety guidance. A single, stand-alone *pop-up* is presented if the user proceeds, warning of the importance of brand prescribing. This is in contrast to EMIS, where it is presented alongside other clinical warnings.

## Discussion

### Principal Findings

A total of 2,575,411 prescriptions were issued in 2018 for ciclosporin, tacrolimus, and diltiazem (over 60 mg); of these, 316,119 prescriptions breached NHS safety guidance (12.27%). Breaches were most common among users of the EMIS EHR system ( $n=3888$ ; December 2018), but breaches were observed in all EHR systems.

### Strengths and Weaknesses

We included all typical practices in England, thus minimizing the potential for obtaining a biased sample. We used prescribing data derived from pharmacy claims data used to calculate the transfer of funds from CCGs to dispensing pharmacies: all parties are motivated to ensure the accuracy of this data. We excluded a small number of settings such as walk-in centers, which typically do not issue repeat prescriptions for medicines, and where no data on EHR usage is available. The data does not distinguish which prescriptions may have been written by hand without EHR generation; however, as 74% of prescriptions are transmitted electronically [2] and many more are generated electronically and then printed on paper, we do not expect this to substantially affect our results. Our data do not include hospital medicines data, but we do not expect the same issue associated with EHR system design to occur in hospitals, as medicines are procured differently and the use of electronic prescribing and EHRs in secondary care in England is limited [12]. Observation of EHR systems was limited to one 30-minute session on each EHR system with limited scenarios in a single

location. EHRs can be manually adapted and defaults set locally; compliance with safety guidance may, therefore, differ between individual practices even with the same EHR systems. Additionally, practices may have secondary decision support software systems providing information and prompts, which may affect breaches.

One caveat regarding clinical impact is that the available data reflect prescribing and not dispensing: in England, when medicines are prescribed generically, pharmacies are entitled to dispense a brand, and some pharmacists may try to ensure they dispense the same brand that the patient previously received; however, pharmacists are likely to be blocked in this by lack of access to the patient record and, commonly, by the absence of information on specific brands in the patient's record [13]. Additionally in England, CCGs employ teams of pharmacy professionals who in some areas may have made concerted manual effort to amend repeat prescriptions following audit. In any case, occasional remedial interventions do not affect the key observation that EHR systems strongly predict safety breaches among prescribing clinicians.

### Findings in Context

We are aware of no prior work on the prevalence of safety breaches around use of ciclosporin, tacrolimus, and diltiazem; however, incomplete implementation of this important national prescribing safety recommendation is consistent with extensive prior work showing incomplete or slow adoption for other national prescribing guidance [14–16]. To our knowledge, we are the first group to use natural variation in prescribing behavior between EHR systems to identify, explain, and address suboptimal prescribing using national data. One small study involving 90 general practices in the Netherlands did find an association between prescribing safety and EHR systems with regard to prescribing of gastroprotective agents to prevent side effects of anti-inflammatory medicines [17]. A 2017 systematic review [18] identified 34 relevant studies exploring the role of computerized systems in suboptimal prescribing. However, none used quantitative methods to compare different systems, instead relying on questionnaires to interrogate clinicians about their experiences of EHRs, qualitative research observing or interviewing clinicians, and descriptions of clinicians' spontaneous reports on errors and safety issues in EHR systems.

Various small studies have aimed to evaluate the impact of a single specific new change to a computerized prescribing system, as a behavioral intervention to increase the probability of a desired choice being made by clinicians using the system [19–23]. Despite evidence for effectiveness from *pop-ups*, *alert fatigue* has also been described, whereby large numbers of pop-ups can result in salient messages being ignored or disabled by prescribers [24]. One small mixed methods study involving the EMIS EHR system found that only three pop-ups from 117 alerts resulted in the general practitioner checking, but not altering, the prescription [25].

### Policy Implications and Interpretation

Breaches of safety guidance around diltiazem, ciclosporin, and tacrolimus can expose patients to avoidable clinical risk. This is especially important in the case of organ transplantation,

where failure to adhere to the safety guidance assessed in our analysis has resulted in organ rejection [7]. The current World Health Organization challenge on medication safety encourages countries to reduce medication errors by 50% by 2022; our finding represents a clear example of an error amenable to change [26]. We strongly recommend that picking-lists be configured in line with best practice guidelines, and that compliance with this be audited by national organizations such as NHS Digital and NHSX. The US National Coordinator for Health Information Technology has made similar recommendations [18,27,28]. For the specific safety issue raised by our analysis, the mandated medicines data standard for the NHS, the Dictionary of Medicines and Devices (dm+d), already includes a field that can be used to mandate branded prescribing where this is required by safety guidance; indeed, it is likely that this field triggers the pop-up warnings described above.

### Future Research

We are concerned by the relative absence of applied practical research around the EHR systems used by clinicians to store information, retrieve relevant information rapidly when assessing a patient, and implement specific clinical actions, such as ordering a test or prescribing a treatment. Health care activity is increasingly computerized, and EHR software is likely to exert a very substantial influence on the way that modern medicine is practiced, in the same way that the rapid explosion in social media usage has changed the ways that people interact socially [29]. We can find no previous attempts to evaluate the impact of EHR system design on clinical practice by analyzing variations between organizations using different EHR systems. In our view, questions of how best to represent, retrieve, and present knowledge about patients to clinicians—and the impact this can have on patient care—should

be a key priority for funders and researchers in “digital health.” The NHS and health care systems around the world spend large sums of money on EHR systems [30] and there is substantial room for collaborative improvement to make clinical care safer, cheaper, and more effective. Since sharing our preprint version of this paper, EMIS have implemented a change to their EHR system to address our findings and we will evaluate the effectiveness of this change in 12 months’ time.

There is also a role for open standards in this work. We are aware that in most general practices, complementary decision support systems are deployed alongside the core EHR system, to support medicines-optimization activity and other aspects of quality improvement [31-33]. However, we have been repeatedly blocked from researching the impact of these pop-ups on routine prescribing, as there is no national framework or data detailing which pop-ups are implemented in each setting, and we have been unable to access the data by private negotiation. NHS commissioners and leaders are also blocked from routinely monitoring which pop-ups are implemented across the NHS and from defining or deploying pop-ups nationally. In our view, this represents a failure to set standards and ensure the proportionate and secure data sharing necessary to evaluate and improve patient safety.

### Summary

National guidance on safe prescribing for ciclosporin, tacrolimus, and diltiazem is commonly breached, and the prevalence of safety breaches is strongly influenced by the brand of EHR system. We recommend that EHR vendors immediately review their systems to mandate safe prescribing of these medicines in line with national guidance. We also recommend that more data and funding be made available to support research into the impact of EHR system design on clinical practice.

### Acknowledgments

We are grateful to Peter Inglesby and Dave Evans for their contributions to maintaining databases and the OpenPrescribing website. We are grateful to wider NHS colleagues for discussions that have informed our work on this topic. No specific funding was sought for this work. BG’s work on clinical informatics has been supported by the National Institute for Health Research (NIHR) Biomedical Research Centre, Oxford; NHS England; the Health Foundation (Award Reference Number 7599); the NIHR School of Primary Care Research (SPCR) (Award Reference Number 327); the NIHR Applied Research Collaboration Oxford and Thames Valley; Health Data Research UK; and policy work by the Mohn-Westlake Foundation. Funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

### Authors' Contributions

All authors conceived the study and designed the methods. BMK, SB, and AW analyzed the data. BMK, HC, and BG drafted the manuscript. All authors contributed to and approved the final manuscript. BG supervised the project and is guarantor.

### Conflicts of Interest

All authors have completed the International Committee of Medical Journal Editors (ICMJE) uniform disclosure form available at the ICMJE website [34]. BG has received research funding from the Laura and John Arnold Foundation, the Wellcome Trust, the Oxford Biomedical Research Centre, the NIHR Applied Research Collaboration Oxford and Thames Valley, the NHS NIHR SPCR, the Health Foundation, Health Data Research UK, the Mohn-Westlake Foundation, and the World Health Organization; he also receives personal income from speaking and writing for lay audiences on the misuse of science. RC, AW, HC, and SB are employed on BG’s grants for OpenPrescribing. BMK is seconded to the DataLab from NHS England. The views expressed

in this publication are those of the authors and not necessarily those of the NIHR, NHS England, or the Department of Health and Social Care.

## Multimedia Appendix 1

Full regression models.

[DOCX File, 115 KB - [jmir\\_v22i10e17003\\_app1.docx](#)]

## References

1. Prescription Cost Analysis - England, 2018. NHS Digital. 2019 Mar 28. URL: <https://digital.nhs.uk/data-and-information/publications/statistical/prescription-cost-analysis/2018> [accessed 2019-04-08]
2. Deployment and utilisation progress data. NHS Digital. URL: <https://digital.nhs.uk/data-and-information/data-tools-and-services/tools-for-accessing-data/deployment-and-utilisation-progress-data> [accessed 2019-07-02]
3. GPSoc services. NHS Digital. URL: <https://digital.nhs.uk/services/gp-systems-of-choice/gpsoc-services> [accessed 2019-04-08]
4. MacKenna B. Ghost branded generics: Why does the cost of generic atorvastatin vary? EBM DataLab. 2018 Dec 20. URL: <https://ebmdatalab.net/ghost-branded-generics-why-does-the-cost-of-generic-atorvastatin-vary%ef%bb%bf/> [accessed 2019-04-08]
5. MacKenna B, Curtis HJ, Walker AJ, Bacon S, Croker R, Goldacre B. Suboptimal prescribing behaviour associated with clinical software design features: A retrospective cohort study in English NHS primary care. *Br J Gen Pract* 2020 Sep;70(698):e636-e643 [FREE Full text] [doi: [10.3399/bjgp20X712313](https://doi.org/10.3399/bjgp20X712313)] [Medline: [32784218](#)]
6. Ciclosporin: Must be prescribed and dispensed by brand name. Drug Safety Update. Volume 3, Issue 5. London, UK: Medicines and Healthcare products Regulatory Agency (MHRA); 2009 Dec. URL: <https://webarchive.nationalarchives.gov.uk/20101116205950/http://www.mhra.gov.uk/home/groups/pl-p/documents/publication/con065445.pdf> [accessed 2020-09-15]
7. Medicines and Healthcare products Regulatory Agency (MHRA). Oral tacrolimus products: Reminder to prescribe and dispense by brand name only. GOV.UK. 2017 Nov 24. URL: <https://www.gov.uk/drug-safety-update/oral-tacrolimus-products-reminder-to-prescribe-and-dispense-by-brand-name-only> [accessed 2019-04-08]
8. Joint Formulary Committee. Diltiazem hydrochloride. British National Formulary (<https://www.bnf.org/>). London, UK: BMJ Group and Pharmaceutical Press URL: <https://bnf.nice.org.uk/drug/diltiazem-hydrochloride.html#indicationsAndDoses> [accessed 2018-07-13]
9. Freedom of information request NIC-142425-M5S5D. NHS Digital. URL: <https://webarchive.nationalarchives.gov.uk/20180307204908/https://digital.nhs.uk/article/9028/Freedom-of-information-request-NIC-142425-M5S5D> [accessed 2020-09-16]
10. Open prescribing. GitHub. URL: <https://github.com/ebmdatalab/openprescribing> [accessed 2019-08-06]
11. Software differences paper. GitHub. URL: [https://github.com/ebmdatalab/software\\_differences\\_paper](https://github.com/ebmdatalab/software_differences_paper) [accessed 2019-10-25]
12. Department of Health and Social Care. Regional funding announced for electronic prescribing. GOV.UK. 2018 Nov 28. URL: <https://www.gov.uk/government/news/regional-funding-announced-for-electronic-prescribing> [accessed 2019-07-02]
13. Pharmacy access to full patient records is more critical than ever. The Pharmaceutical Journal. Vol 303, No 7927. London, UK: Royal Pharmaceutical Society; 2019 Jul 23. URL: <https://www.pharmaceutical-journal.com/opinion/editorial/pharmacy-access-to-full-patient-records-is-more-critical-than-ever/20206820.article> [accessed 2019-08-06]
14. Walker AJ, Bacon S, Curtis H, Croker R, MacKenna B, Goldacre B. Six months on: NHS England needs to focus on dissemination, implementation and audit of its low-priority initiative. *J R Soc Med* 2019 Jan;112(1):4-5. [doi: [10.1177/0141076818808429](https://doi.org/10.1177/0141076818808429)] [Medline: [30507285](#)]
15. Croker R, Walker AJ, Goldacre B. Why did some practices not implement new antibiotic prescribing guidelines on urinary tract infection? A cohort study and survey in NHS England primary care. *J Antimicrob Chemother* 2019 Apr 01;74(4):1125-1132. [doi: [10.1093/jac/dky509](https://doi.org/10.1093/jac/dky509)] [Medline: [30590552](#)]
16. MacKenna B, Curtis HJ, Walker AJ, Croker R, Bacon S, Goldacre B. Trends and variation in unsafe prescribing of methotrexate: A cohort study in English NHS primary care. *Br J Gen Pract* 2020 Jul;70(696):e481-e488 [FREE Full text] [doi: [10.3399/bjgp20X710993](https://doi.org/10.3399/bjgp20X710993)] [Medline: [32571771](#)]
17. Opondo D, Visscher S, Eslami S, Verheij RA, Korevaar JC, Abu-Hanna A. Quality of co-prescribing NSAID and gastroprotective medications for elders in the Netherlands and its association with the electronic medical record. *PLoS One* 2015;10(6):e0129515 [FREE Full text] [doi: [10.1371/journal.pone.0129515](https://doi.org/10.1371/journal.pone.0129515)] [Medline: [26110650](#)]
18. Brown CL, Mulcaster HL, Triffitt KL, Sittig DF, Ash JS, Reygate K, et al. A systematic review of the types and causes of prescribing errors generated from using computerized provider order entry systems in primary and secondary care. *J Am Med Inform Assoc* 2017 Mar 01;24(2):432-440. [doi: [10.1093/jamia/ocw119](https://doi.org/10.1093/jamia/ocw119)] [Medline: [27582471](#)]
19. Patel MS, Day S, Small DS, Howell JT, Lautenbach GL, Nierman EH, et al. Using default options within the electronic health record to increase the prescribing of generic-equivalent medications: A quasi-experimental study. *Ann Intern Med* 2014 Nov 18;161(10 Suppl):S44-S52. [doi: [10.7326/M13-3001](https://doi.org/10.7326/M13-3001)] [Medline: [25402402](#)]

20. Patel MS, Day SC, Halpern SD, Hanson CW, Martinez JR, Honeywell S, et al. Generic medication prescription rates after health system-wide redesign of default options within the electronic health record. *JAMA Intern Med* 2016 Jun 01;176(6):847-848 [FREE Full text] [doi: [10.1001/jamainternmed.2016.1691](https://doi.org/10.1001/jamainternmed.2016.1691)] [Medline: [27159011](https://pubmed.ncbi.nlm.nih.gov/27159011/)]
21. Halpern SD, Ubel PA, Asch DA. Harnessing the power of default options to improve health care. *N Engl J Med* 2007 Sep 27;357(13):1340-1344. [doi: [10.1056/NEJMs071595](https://doi.org/10.1056/NEJMs071595)] [Medline: [17898105](https://pubmed.ncbi.nlm.nih.gov/17898105/)]
22. Patel MS, Volpp KG, Asch DA. Nudge units to improve the delivery of health care. *N Engl J Med* 2018 Jan 18;378(3):214-216 [FREE Full text] [doi: [10.1056/NEJMp1712984](https://doi.org/10.1056/NEJMp1712984)] [Medline: [29342387](https://pubmed.ncbi.nlm.nih.gov/29342387/)]
23. Vaughn VM, Linder JA. Thoughtless design of the electronic health record drives overuse, but purposeful design can nudge improved patient care. *BMJ Qual Saf* 2018 Aug;27(8):583-586. [doi: [10.1136/bmjqs-2017-007578](https://doi.org/10.1136/bmjqs-2017-007578)] [Medline: [29574463](https://pubmed.ncbi.nlm.nih.gov/29574463/)]
24. Backman R, Bayliss S, Moore D, Litchfield I. Clinical reminder alert fatigue in healthcare: A systematic literature review protocol using qualitative evidence. *Syst Rev* 2017 Dec 13;6(1):255 [FREE Full text] [doi: [10.1186/s13643-017-0627-z](https://doi.org/10.1186/s13643-017-0627-z)] [Medline: [29237488](https://pubmed.ncbi.nlm.nih.gov/29237488/)]
25. Hayward J, Thomson F, Milne H, Buckingham S, Sheikh A, Fernando B, et al. 'Too much, too late': Mixed methods multi-channel video recording study of computerized decision support systems and GP prescribing. *J Am Med Inform Assoc* 2013 Jun;20(e1):e76-e84 [FREE Full text] [doi: [10.1136/amiajnl-2012-001484](https://doi.org/10.1136/amiajnl-2012-001484)] [Medline: [23470696](https://pubmed.ncbi.nlm.nih.gov/23470696/)]
26. Short life working group report on medication errors Internet. Department of Health and Social Care. URL: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/683430/short-life-working-group-report-on-medication-errors.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/683430/short-life-working-group-report-on-medication-errors.pdf) [accessed 2019-05-20]
27. Brigham and Women's Hospital, Harvard Medical School, Partners HealthCare. Computerized Prescriber Order Entry Medication Safety (CPOEMS): Uncovering and Learning From Issues and Errors. Silver Spring, MD: US Food and Drug Administration; 2015. URL: <https://www.fda.gov/media/95234/download> [accessed 2020-09-14]
28. Rizk S, Oguntebi G, Graber ML, Johnston D, RTI International. Report on the Safe Use of Pick Lists in Ambulatory Care Settings: Issues and Recommended Solutions for Improved Usability in Patient Selection and Medication Ordering. Washington, DC: US Department of Health and Human Services; 2016 Sep. URL: <https://www.healthit.gov/sites/default/files/report-on-the-safe-use-of-pick-lists-in-ambulatory-care-settings.pdf> [accessed 2020-09-14]
29. Shirky C. Here Comes Everybody: The Power of Organizing Without Organizations. New York, NY: The Penguin Press; 2008.
30. Hughes O. GP IT futures: First stage of procurement process now complete. *Digital Health*. 2019 Aug 14. URL: <https://www.digitalhealth.net/2019/08/gp-it-futures-first-stage-of-procurement-process-now-complete/> [accessed 2019-10-23]
31. ScriptSwitch. Optum. URL: <http://www.optum.co.uk/how-we-help/scriptswitch.html> [accessed 2019-06-26]
32. FDB OptimiseRx®. First Databank. URL: <https://www.fdbhealth.co.uk/solutions/optimiserx> [accessed 2019-06-26]
33. Prescribing Services. URL: <https://www.prescribingservices.org/> [accessed 2019-10-25]
34. Conflicts of interest. International Committee of Medical Journal Editors (ICMJE). URL: <http://www.icmje.org/conflicts-of-interest/> [accessed 2020-09-15]

## Abbreviations

**CCG:** clinical commissioning group  
**dm+d:** Dictionary of Medicines and Devices  
**EHR:** electronic health record  
**ICMJE:** International Committee of Medical Journal Editors  
**MHRA:** Medicines and Healthcare products Regulatory Agency  
**NHS:** National Health Service  
**NIHR:** National Institute for Health Research  
**SPCR:** School of Primary Care Research

*Edited by G Eysenbach; submitted 11.11.19; peer-reviewed by D Wong, T Cornford; comments to author 16.01.20; revised version received 11.02.20; accepted 29.02.20; published 16.10.20.*

### *Please cite as:*

MacKenna B, Bacon S, Walker AJ, Curtis HJ, Croker R, Goldacre B  
Impact of Electronic Health Record Interface Design on Unsafe Prescribing of Ciclosporin, Tacrolimus, and Diltiazem: Cohort Study  
in English National Health Service Primary Care  
*J Med Internet Res* 2020;22(10):e17003  
URL: <https://www.jmir.org/2020/10/e17003>  
doi:[10.2196/17003](https://doi.org/10.2196/17003)  
PMID:[33064085](https://pubmed.ncbi.nlm.nih.gov/33064085/)



©Brian MacKenna, Sebastian Bacon, Alex J Walker, Helen J Curtis, Richard Croker, Ben Goldacre. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 16.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Deep Learning With Electronic Health Records for Short-Term Fracture Risk Identification: Crystal Bone Algorithm Development and Validation

Yasmeen Adar Almog<sup>1</sup>, BSE; Angshu Rai<sup>1</sup>, BTECH, MTECH; Patrick Zhang<sup>1</sup>, BSc, MS; Amanda Moulaison<sup>1</sup>, BSc; Ross Powell<sup>1</sup>, BA; Anirban Mishra<sup>1</sup>, BSc, MSc; Kerry Weinberg<sup>1</sup>, BSc, MBA, MSc; Celeste Hamilton<sup>2</sup>, BHK, MSc, PhD; Mary Oates<sup>3</sup>, BSc, MD; Eugene McCloskey<sup>4</sup>, MD, FRCP; Steven R Cummings<sup>5</sup>, BA, MD, FACP

<sup>1</sup>Digital Health & Innovation, Amgen Inc, Thousand Oaks, CA, United States

<sup>2</sup>Global Medical Operations, Amgen Inc, Thousand Oaks, CA, United States

<sup>3</sup>US Medical, Amgen Inc, Thousand Oaks, CA, United States

<sup>4</sup>Department of Oncology & Metabolism, The University of Sheffield, Sheffield, United Kingdom

<sup>5</sup>Department of Medicine, University of California San Francisco, San Francisco, CA, United States

**Corresponding Author:**

Yasmeen Adar Almog, BSE  
Digital Health & Innovation  
Amgen Inc  
1 Amgen Center Drive  
MS 38-3B  
Thousand Oaks, CA, 91320  
United States  
Phone: 1 4243463036  
Email: [yalmog@amgen.com](mailto:yalmog@amgen.com)

## Abstract

**Background:** Fractures as a result of osteoporosis and low bone mass are common and give rise to significant clinical, personal, and economic burden. Even after a fracture occurs, high fracture risk remains widely underdiagnosed and undertreated. Common fracture risk assessment tools utilize a subset of clinical risk factors for prediction, and often require manual data entry. Furthermore, these tools predict risk over the long term and do not explicitly provide short-term risk estimates necessary to identify patients likely to experience a fracture in the next 1-2 years.

**Objective:** The goal of this study was to develop and evaluate an algorithm for the identification of patients at risk of fracture in a subsequent 1- to 2-year period. In order to address the aforementioned limitations of current prediction tools, this approach focused on a short-term timeframe, automated data entry, and the use of longitudinal data to inform the predictions.

**Methods:** Using retrospective electronic health record data from over 1,000,000 patients, we developed Crystal Bone, an algorithm that applies machine learning techniques from natural language processing to the temporal nature of patient histories to generate short-term fracture risk predictions. Similar to how language models predict the next word in a given sentence or the topic of a document, Crystal Bone predicts whether a patient's future trajectory might contain a fracture event, or whether the signature of the patient's journey is similar to that of a typical future fracture patient. A holdout set with 192,590 patients was used to validate accuracy. Experimental baseline models and human-level performance were used for comparison.

**Results:** The model accurately predicted 1- to 2-year fracture risk for patients aged over 50 years (area under the receiver operating characteristics curve [AUROC] 0.81). These algorithms outperformed the experimental baselines (AUROC 0.67) and showed meaningful improvements when compared to retrospective approximation of human-level performance by correctly identifying 9649 of 13,765 (70%) at-risk patients who did not receive any preventative bone-health-related medical interventions from their physicians.

**Conclusions:** These findings indicate that it is possible to use a patient's unique medical history as it changes over time to predict the risk of short-term fracture. Validating and applying such a tool within the health care system could enable automated and widespread prediction of this risk and may help with identification of patients at very high risk of fracture.

**KEYWORDS**

fracture; bone; osteoporosis; low bone mass; prediction; natural language processing; NLP; machine learning; deep learning; artificial intelligence; AI; electronic health record; EHR

**Introduction**

Fractures due to osteoporosis and low bone mass are associated with a significant personal, clinical, and economic burden. These fractures are common; the risk of sustaining such a fracture increases with age, and their incidence is expected to increase worldwide as the population ages [1-11]. In the United States, an estimated 1 in 2 women and 1 in 4 men over 50 years of age will experience such a fracture [12-14]. However, there remains a significant diagnosis and treatment gap for osteoporosis [1,2,4,12]. When these fractures occur, they often result in a loss of independence for patients and can lead to functional disability, lower quality of life, and increased mortality [5,15-38]. Given this substantial burden and unmet need for interventions, it is critical to identify patients at risk of fracture, as effective management of risk can prevent these deleterious outcomes.

Several fracture risk prediction tools have been developed for clinical use. The most commonly used tools are the University of Sheffield Fracture Risk Assessment Tool, known as FRAX [39], and the Garvan Institute of Health Bone Fracture Risk Calculator (GIH-BFRC) [40]. Both tools use a set of cross-sectional clinical risk factors to evaluate fracture likelihood, and typically require manual data entry to perform the predictions. The performance of both methods varies greatly in real-world analyses; this variance is partially explained by study population and design and predicted fracture outcome (hip vs other osteoporotic fractures). In a review [41], 12 studies of FRAX showed an average area under the receiver operating characteristics curve (AUROC) of 0.65 (SD 0.038) when predicting major osteoporotic fractures without including bone mineral density in the model, and similar results were shown for GIH-BFRC [41]. These commonly used risk assessment tools estimate 5- and 10-year fracture risk but do not provide estimates of 1- to 2-year risk [42-45].

Increased risk of fracture in the next 1-2 years is not routinely assessed in clinical practice, despite the existence of rapid-acting preventative therapeutics [8,46,47]. Although methods for predicting short-term risk have been explored [48-50], they have not yet been widely clinically accepted. Furthermore, these models are limited to a specific set of cross-sectional information, some of which may not readily be available. Thus, there remains a need to further develop a fracture risk prediction tool that predicts on a short-term time frame in order to facilitate identification of patients at high risk. While there are published examples [51-53] applying artificial intelligence to fracture and osteoporosis risk, these approaches focus either on imaging data [51] or on cross-sectional data for long-term predictions [52,53]. To our knowledge, there is no existing method that applies deep learning to sequential patient data for predicting fracture risk.

To address these unmet needs, we developed *Crystal Bone*, a machine learning approach that leverages techniques typically applied in natural language processing. However, rather than applying these methods to text-based data, we applied them to longitudinal data contained in electronic health records. Specifically, we focused on diagnosis codes (International Classification of Diseases; ICD), treating each code as a word and sequences of codes as stories. The goal of this study was to evaluate the ability of these natural language processing-based models to learn patterns associated with increased short-term (ie, 2-year) fracture risk. The results of our analyses suggest that not only does this unique longitudinal method produce accurate short-term fracture risk predictions, but also that the approach can help fulfill the unmet need that exists in fracture-risk identification.

**Methods****Data Background**

We used subsets of the Optum deidentified electronic health record data set, which contains comprehensive longitudinal electronic health record data for 91 million patients from over 140,000 providers (as of March 2018) from the United States. The subsets, which contain bone health and pan-therapeutic populations respectively, cover the time from January 1, 2007, through December 31, 2018 (Optum, email communication, August 2019).

The bone health subset was obtained by filtering for patients with osteoporosis, fractures, or bone-related medications ( $n=6,329,986$ ). In the period covered by the data set, the fracture incidence rate (ie, the proportion of fractures among all events detected, which may include multiple fractures per person) was 39% in the population over 50 years of age. The bone health data set was primarily used for training the model.

The pan-therapeutic data set represented a random sample of 5% of the overall Optum electronic health record data set and contained patient data ( $n=3,476,219$ ) with no filtering for any specific comorbidities or treatments; this dataset had a fracture incidence rate of 8.5% in the population over 50 years of age. Because the sample was drawn from such a large population, the pan-therapeutic data set was assumed to be broadly representative of the US population. As such, we performed all model evaluations on a testing sample from this data set (a holdout data set), to better understand the generalizability of the model in a real-world setting.

**Ethical Approval**

Since this was a retrospective study using deidentified data, patients were not required to actively participate in the study. Therefore, neither informed consent of patients nor institutional review board approval was required.

## Data Engineering and Cohort Selection

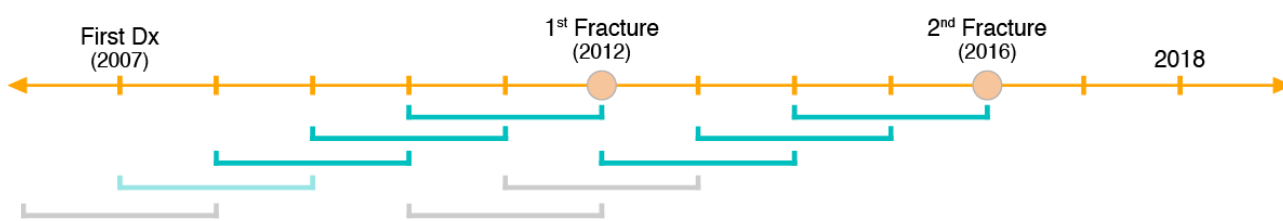
The cohort consisted of patients who were at least 50 years of age at the time of their event; this criterion was chosen to reduce the data to a population that is more susceptible to fractures associated with osteoporosis and low bone mass. For fracture patients, an event is the date of occurrence of any qualifying fracture. Qualifying fractures are defined by a set of rules based on those used by Wright et al [54] for identifying novel and relevant fracture events in claims data. For nonfracture patients, an event is the date of the last recorded diagnosis of any kind in the data set. We describe further details of the fracture identification process in [Multimedia Appendix 1](#).

We further filtered our cohorts for patients with at least 2 years of medical history leading up to their respective events. Applying these parameters limited the bone health cohort to 3,408,494 patients and the pan-therapeutic cohort to 700,315 patients.

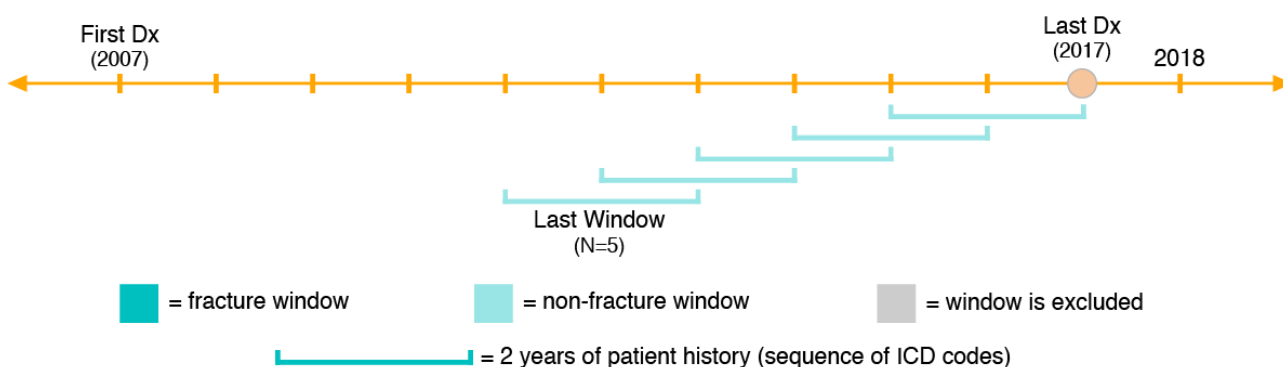
We applied sliding windows to the data ([Figure 1](#)), where each event could have up to 5 windows, and each window was a historical sequence defined as the list of chronologically ordered ICD codes in the 2 years leading up to an event. These historical sequences were then used to predict risk of fracture within a 2-year horizon (a 1-year horizon was also explored, see [Multimedia Appendix 1](#)). As shown in [Figure 1](#), some windows were dropped from the analysis due to incomplete or potentially overlapping coverage. Additionally, windows that occur more than 2 years before a fracture event were labeled as nonfracture windows. The motivation for this approach was to provide the algorithm with multiple unique code sequences leading up to the same event that may reflect changes in risk at various times within the given time horizon. Furthermore, the fixed window size provided a consistent timeframe for prediction as opposed to varying lengths of time for each patient, which would have occurred if patients' complete code histories were used. Further details regarding the motivation and methodology of this approach are in [Multimedia Appendix 1](#).

**Figure 1.** Sliding window algorithm schematic. This schematic depicts the sliding window algorithm for a multifracture and nonfracture patient. Dx: diagnosis; ICD: International Classification of Diseases.

### Sample Fracture Patient Windows:



### Sample Non-Fracture Patient Windows:



There was no additional filtering based on specific diagnoses or comorbidities. For each qualifying patient, the algorithms utilized all available ICD codes in the historical sequences described above. Only the codes that occurred fewer than 5 times in the full cohort were excluded, as these codes were too rare to be included in the diagnosis code vocabulary.

## Data Sampling

Before model training, we generated a 70:30 random split of the pan-therapeutic data, representing training and holdout subsets. Since the pan-therapeutic data set is highly imbalanced, with a fracture event incidence of only 6.5% after applying the sliding window algorithm, we oversampled additional fracture windows from the bone health data set to achieve a balanced

(50:50) training set for modeling. This oversampling training paradigm was replicated for all models. The holdout set remained untouched, with the original distribution of fractures.

## Modeling Approaches

### Overview

Crystal Bone was inspired by techniques that are typically applied in natural language processing. However, instead of applying these techniques to text-based data, we applied them to sequences of ICD codes. Correspondingly, each ICD code was analogous to a word, and each sequence of ICD codes was analogous to a document. To this end, we implemented 2 distinct frameworks: (1) ICD code vectorization and long short-term

memory networks, and (2) patient-level vectorization and extreme gradient boosting decision trees. Both approaches utilize sequences of ICD codes as inputs. The ICD code vectorization and long short-term memory framework undertakes this task by first learning semantic definitions for the codes, then evaluating the sequence of definitions through a deep learning network.. The patient-level vectorization and extreme gradient boosting modeling framework employs a similar approach; however, rather than embedding individual ICD codes, it embeds the entire ICD code sequence for each patient, thereby learning “summaries” of patient sequences. This framework produces a prediction by feeding these summaries through a decision tree classifier. The model parameters were tuned to optimize AUROC; details of this process are provided in [Multimedia Appendix 1](#).

### ***Framework 1: ICD Code Vectorization + Long Short-Term Memory***

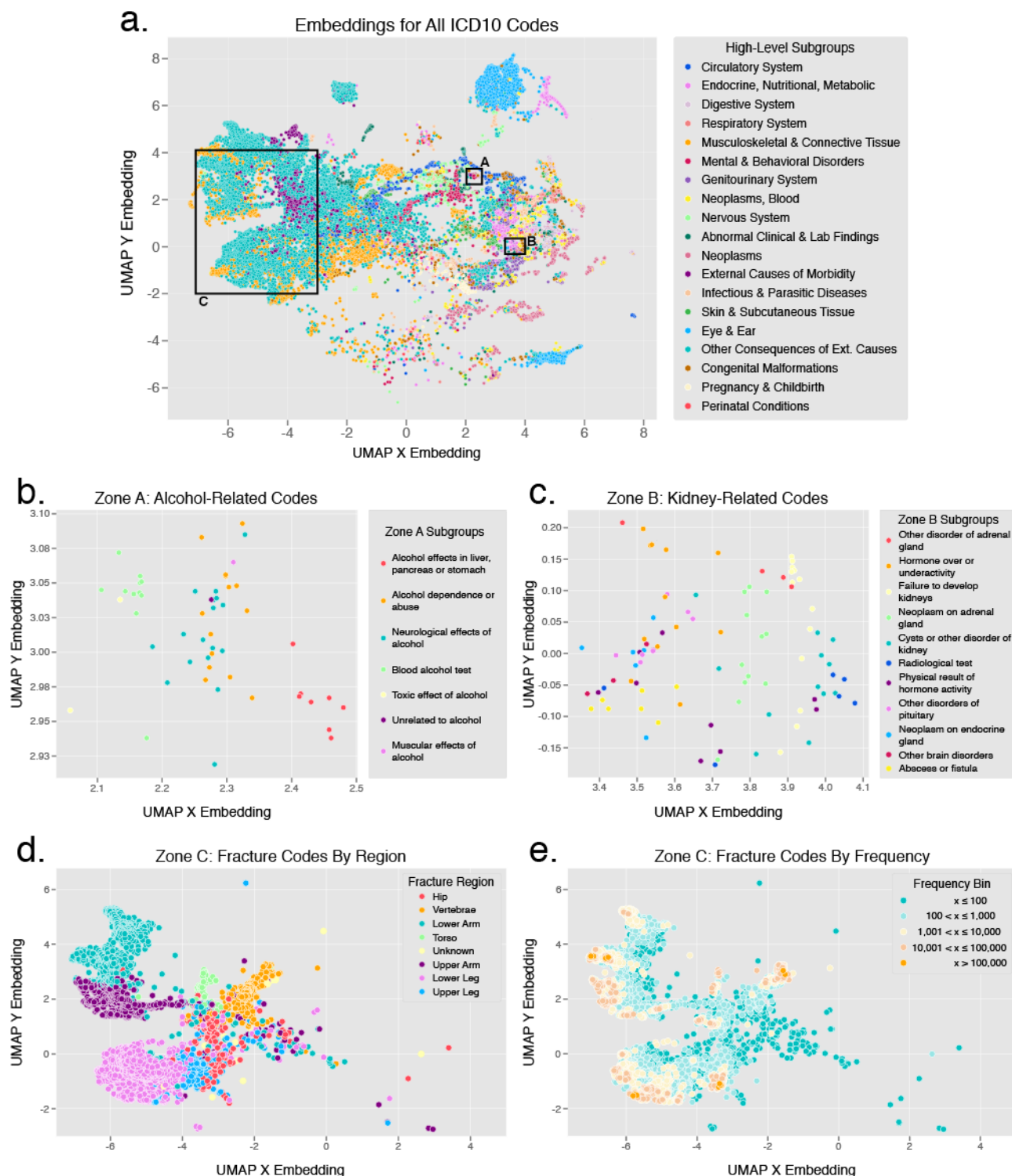
The first framework consisted of 2 primary components. The ICD code vectorization component was responsible for learning a “definition” for each ICD code based on skip-gram architecture word embedding (word2vec) [55], an unsupervised learning approach that mapped each code in the vocabulary to a 100-dimensional vector. To generate these embeddings, we utilized sequences from the pan-therapeutic training set alone (without oversampling), to avoid bias toward bone-health related codes. In our implementation, the vocabulary consisted of all diagnosis codes that occurred at least 5 times in this data set, amounting to more than 40,000 unique codes. The method generated a vector for each code based on the context in which it appeared; in electronic health records, similar ICD codes

appear in similar contexts, and as a result have similar vector representations. These embeddings reduced the dimensionality and sparsity of the feature space, and helped the neural network recognize related ICD codes. [Figure 2](#) illustrates the encoded vectors projected onto a 2D space using uniform manifold approximation and projection (UMAP) for dimension reduction [56]. The collocation of related diagnosis codes in this coordinate space provided qualitative evidence that the ICD code vectorization had encoded meaningful latent information.

The long short-term memory component consisted of a neural network with long short-term memory layers, a deep learning architecture that enables the evaluation of recurrent data, such as sequences of embedded ICD codes. We trained this network with the complete training set (including oversampling from the bone health data set). The long short-term memory network predicted the likelihood of a fracture event within 2 years as a classification problem. Long short-term memory networks are a common approach for solving such problems [57].

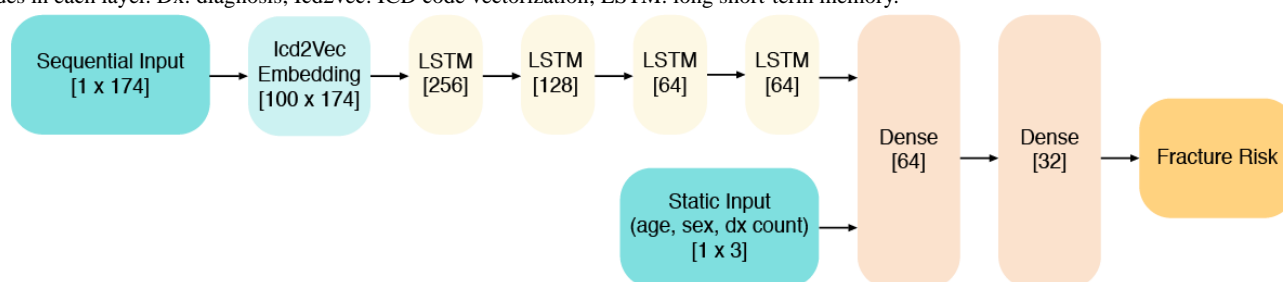
Additionally, given the ubiquitous use of nonsequential features such as age and sex for predicting fracture risk, we supplied age and sex to the neural network as static features through concatenation of long short-term memory and dense layers. Furthermore, because the long short-term memory framework required all input sequences to have uniform length, we also included total diagnosis count as a static feature to account for the effects of truncating or padding the sequences. The schematic in [Figure 3](#) provides an overview of the model architecture and inputs to the algorithm, namely age, sex, diagnosis count, and the patient’s unique sequence of ICD codes.

**Figure 2.** 2D projection of ICD-10 code embeddings from the ICD code vectorization model: (a) All ICD-10 codes by the first letter (high-level category) of the code, (b) a cluster of codes related to alcohol near coordinates (2.3, 3) by code subgroups, (c) a cluster of codes related to kidney function near coordinates (3.75, 0.025) by code subgroups, and all ICD-10 fracture codes in region C (d) by region of the body, and (e) by frequency of occurrence. ICD: International Classification of Diseases; UMAP: uniform manifold approximation and projection.





**Figure 3.** High-level architecture of the long short-term memory neural network including the dimensionality of the inputs, as well as the number of nodes in each layer. Dx: diagnosis; Icd2vec: ICD code vectorization; LSTM: long short-term memory.



### Framework 2: Patient-Level Vectorization and Extreme Gradient Boosting

Similar to the ICD code vectorization + long short-term memory modeling framework, the patient-level vectorization and extreme gradient boosting decision trees framework consists of 2 components. First, the patient-level vectorization embeds entire ICD code sequences to a 128-dimensional semantic space using the distributed bag of words framework [58]. Much as the ICD code vectorization learned definitions of individual ICD codes, the patient-level vectorization instead learned summaries of patient sequences. The method for doing so is the same; patients with similar sequential contexts will have similar summary vectors. We trained the patient-level vectorization with the sliding window ICD code sequences, again only utilizing the pan-therapeutic data to avoid bias toward the bone health therapeutic area. This created embeddings that represented 2-year episodes of patient histories; a detailed exploration of these embeddings is in [Multimedia Appendix 1](#).

The extreme gradient boosting decision trees component utilized the embeddings from the patient-level vectorization, as well as the static features of age, sex, and total diagnosis count that were incorporated in Framework 1, to predict fracture risk. This type of algorithm, also referred to as XGBoost, is a scalable tree-based modeling approach that improves the generalizability, speed, and efficacy of prediction [59]. We trained this algorithm with the full training set (including bone health data set oversampling) to learn a classification model that predicted the likelihood of fracture within 2 years.

### Ensemble Model

An ensemble model was also evaluated. This algorithm combined the outputs of both the aforementioned frameworks with a logistic regression metaclassifier.

### Baseline Models

We compared these modeling frameworks to 2 baseline models. The first baseline model utilized the age and sex of each patient at each window. These were 2 of the only features shared by the FRAX tool and the GIH-BFRC models. The other shared feature is prior fracture; however, because neither the FRAX tool nor GIH-BFRC's method of measuring this value was possible for our data set without censoring, we did not include it in the model. The second baseline incorporated age, sex, and total diagnosis count (number of ICD codes) in each sample; these represent all of the static features used by both modelling frameworks, enabling evaluation of the relative benefit of including sequential ICD code data. Both baseline models utilized extreme gradient boosting decision tree algorithms, the same classification approach that was used in Framework 2.

### Human-Level Performance Approximation

In addition to these baselines, we approximated human-level performance by isolating a set of retrospective physician-prescribed interventions that were identifiable in the electronic health record data set. These interventions consisted of diagnostic tests as well as pharmacologic treatments. The list of interventions was based on treatment guidelines provided by the National Osteoporosis Foundation [60] and the Journal of Clinical Endocrinology and Metabolism [61] and was further validated by the physician coauthors of this manuscript, who confirmed that the interventions aligned with their understanding of osteoporosis treatment guidelines ([Table 1](#)). If a patient received one of these interventions in a 2-year historical window, that window was flagged as "physician-identified risk, worthy of intervention." A full description of the limitations of this approach is described in [Multimedia Appendix 1](#).

**Table 1.** List of physician interventions for human-level performance analysis.

Type and name	Pharmacologic
<b>Procedure</b>	
Dual-energy x-ray absorptiometry	No
Vertebral fracture assessment	No
Quantitative computed tomography	No
Other bone density measurements (single energy x-ray absorptiometry, radiographic absorptiometry, ultrasound, single-photon absorptiometry)	No
Bone turnover markers	No
Administration of any medications referenced below	Yes
<b>Treatment</b>	
Bisphosphonates (alendronate, alendronate-cholecalciferol, ibandronate, risedronate, zoledronic acid)	Yes
Abaloparatide	Yes
Denosumab	Yes
Raloxifene	Yes
Bazedoxifene	Yes
Romosozumab	Yes
Teriparatide	Yes
Calcitonin	Yes
<b>Diagnosis</b>	
Osteoporosis (M80, M81, 733.0)	No

We defined the cohort of patients who did not receive any form of intervention (diagnoses, tests, or treatments) as *no intervention* and assessed how well the algorithm was able to correctly identify which patients had a fracture within 2 years, as well as how frequently the algorithm mistakenly flagged patients with no imminent fracture. We also evaluated the patients who received interventions (the *intervention cohort*) with this method, referred to as the *cohort analysis*. However, since an intervention can directly modulate fracture risk, we performed a separate analysis in order to mitigate some of the uncertainty due to the effects of interventions. For this analysis, we identified each patient's first pharmacologic intervention and used the diagnosis history leading up to this date as input. This analysis allowed us to gauge the extent to which the algorithm flags agreed with human-level performance interventions (without needing to adjust for their effects). We termed this the *overlap analysis*. The cohort analysis utilized

the full list of interventions, while the overlap analysis utilized the pharmacological subset of the list of interventions.

### Model Performance

We report model performance on a set of 5 primary metrics: AUROC, recall (sensitivity), specificity, precision, and area under the precision-recall curve (AUPRC).

## Results

### Model Performance

The overall performance of the algorithms is shown through comparison of the 2 frameworks with the 2 baseline models to demonstrate the quality of each algorithm's predictions. [Table 2](#) shows a summary of key model performance metrics on the same holdout data set. The Crystal Bone models, including the ensemble model that combined the 2 approaches, outperformed the baseline models for nearly all performance metrics.

**Table 2.** Comparison of model performance metrics.

Model	AUROC <sup>a</sup>	Recall	Specificity	Precision	AUPRC <sup>b</sup>
ICD code vectorization + LSTM <sup>c</sup>	0.812	0.646	0.812	0.192	0.462
Patient level vectorization + XGBoost <sup>d</sup>	0.790	0.670	0.758	0.161	0.358
Ensemble	0.818	0.693	0.777	0.177	0.463
Baseline (age, sex)	0.667	0.787	0.416	0.0855	0.119
Baseline (age, sex, diagnosis count)	0.668	0.547	0.707	0.114	0.130

<sup>a</sup>AUROC: area under the receiver operating characteristics curve.

<sup>b</sup>AUPRC: area under the precision-recall curve.

<sup>c</sup>LSTM: long short-term memory.

<sup>d</sup>XGBoost: extreme gradient boosting.

### ICD Code Vectorization + Long Short-Term Memory Model

To further characterize this performance, we evaluated the ICD code vectorization and long short-term memory model on primary and subsequent fracture events. While the model performs best on subsequent fractures, both primary and subsequent fracture analyses (AUROC 0.742 and 0.910, respectively) show a marked improvement against corresponding baseline models (AUROC 0.591 and 0.747, respectively). We report detailed results of this experiment and additional evaluations of sensitivity and robustness of this model in [Multimedia Appendix 1](#).

### Human-Level Performance Comparison

[Table 3](#) contains the results of the cohort analysis. For windows with no interventions, Crystal Bone Framework 1 correctly flagged 16,127 of the 28,626 windows that resulted in fracture

(56.3%); this corresponds to 9649 out of 13,765 (70.1%) of the unique fracture events. Crystal Bone Framework 1 incorrectly flagged 91,717 of the 532,621 windows with no fractures as at-risk (17.2%); however, 1053 of the windows in this cohort (3%) sustained a fracture in >2 years.

For windows with interventions, only 11,833 of 69,198 (17.1%) of the detected interventions included treatments; thus, the remaining 57,365 (82.9%) interventions were either diagnoses or diagnostic tests. In the intervention cohort, Crystal Bone Framework 1 correctly captured 10,277 out of 12,244 windows for which fracture occurred within 2 years (83.9%). For the windows with interventions and no fracture event, 19,235 out of 56,954 (33.8%) are incorrectly flagged by our algorithm as at risk. These results suggest Crystal Bone's ability to recognize interventions through their associated ICD codes and adjust the predicted fracture risk accordingly. However, a deeper exploration of specific interventions is required to verify this.

**Table 3.** Human-level performance results.

Cohort	Windows, n (%)	Flag, n (%)	No flag, n (%)
<b>Total</b>	630,445 (100)	— <sup>a</sup>	—
<b>No intervention</b>	561,247 (89.0)	—	—
Fracture	28,626 (5.1)	16,127 (56.3)	12,449 (43.7)
Nonfracture	532,621 (94.9)	91,717 (17.2)	440,904 (82.8)
<b>Intervention</b>	69,198 (11.0)	—	—
Fracture	12,244 (17.7)	10,277 (83.9)	1967 (16.1)
Nonfracture	56,954 (82.3)	19,235 (33.8)	37,719 (66.2)

<sup>a</sup>Not reported.

The overlap analysis enabled us to better understand how well Crystal Bone Framework 1 correlated with observed physician interventions through exploration of the first pharmacological treatment in the holdout set. Of the 7127 patients who received treatment, 6071 had enough medical history leading up to this treatment for Crystal Bone Framework 1. When evaluating these patients, 3017 out of those 6071 (49.7%) were considered at risk of fracture in 2 years.

We evaluated the incidence of fracture within 2 years for this subgroup. Of the cohort deemed at risk by the algorithm, 684 out of 3017 (22.7%) experienced a fracture within 2 years of

the first intervention date. This precision is a slight improvement over that of the algorithm on the overall holdout set, at 19.2%. Furthermore, of all 570 patients in this pharmacological intervention cohort who ultimately suffered from a fracture within 2 years, Crystal Bone Framework 1 correctly flagged 469 (82.3%).

## Discussion

### General

In this study, we evaluated the performance of 2 natural language processing–inspired fracture prediction models: (1) ICD code vectorization and long short-term memory (AUROC 0.812) and (2) patient-level vectorization and extreme gradient boosting (AUROC 0.790). The performance of these models reflected a substantial improvement over 2 baseline models: (1) with age and sex (AUROC 0.670) and (2) with age, sex, total diagnosis count (AUROC 0.670). Furthermore, these short-term prediction metrics were an improvement over cross-sectional tools for long-term time frames, such as FRAX and GIH-BFRC, which have been widely clinically accepted [41]. Although fundamental differences in study design make it impossible to compare these metrics directly, sensitivity analyses of Crystal Bone across fracture types, prediction time frames, and fracture definitions suggest robust predictive performance and generalizability. To our knowledge, this is the first study that has experimented with separate models for primary and subsequent fracture types; further discussion of this analysis, as well as the additional sensitivity analyses, is in [Multimedia Appendix 1](#).

The human-level performance comparison provides deeper insight to the benefits of Crystal Bone. The retrospective labeling utilized in both the cohort and overlap analyses enabled a scalable, data-driven comparison of physician action and Crystal Bone and avoided bias that may occur through alternative methods of human-level performance evaluation [62]. To our knowledge, this is the first fracture risk prediction study which includes such a human-level performance comparison in the analysis.

Through the cohort analysis we learned that only a small proportion of patients received preventative interventions, including basic diagnostic tests, showcasing the extent of unmet need in the health care system [1,2,4,12]. In the subset of patient windows with no interventions, Crystal Bone was able to flag 70.1% of the unique fracture events. Given the existence of rapid-acting preventative therapeutics [8,46,47], as well as the demonstrated efficacy of bone-forming agents in reduction of 1- to 2-year fracture risk [63–69], these results suggest that, had appropriate preventative measures been taken, the risk of these fractures may have been reduced, thus mitigating a significant burden to both the patient and the health care system.

The findings of the overlap analysis further support the merits of Crystal Bone, through demonstration of alignment with observable interventions made by physicians. Because it is impossible to confirm whether these treatment interventions were taken in response to a perceived short-term risk of fracture, we cannot expect 100% overlap between Crystal Bone and these observed interventions. We saw that Crystal Bone was aligned with these physician interventions 49.7% of the time. While this overlap is not complete, it captured 82.3% of the patients who ultimately experienced a fracture, reflecting the algorithm's increased sensitivity for the cohort deemed at-risk by physicians.

This suggests a meaningful alignment with both physician evaluation and actual observed fracture risk. Ultimately, these human-level performance comparisons, coupled with performance against baseline models and alternative risk prediction methods, suggest that Crystal Bone can fulfill a critical unmet need through identification of patients at high risk of fracture.

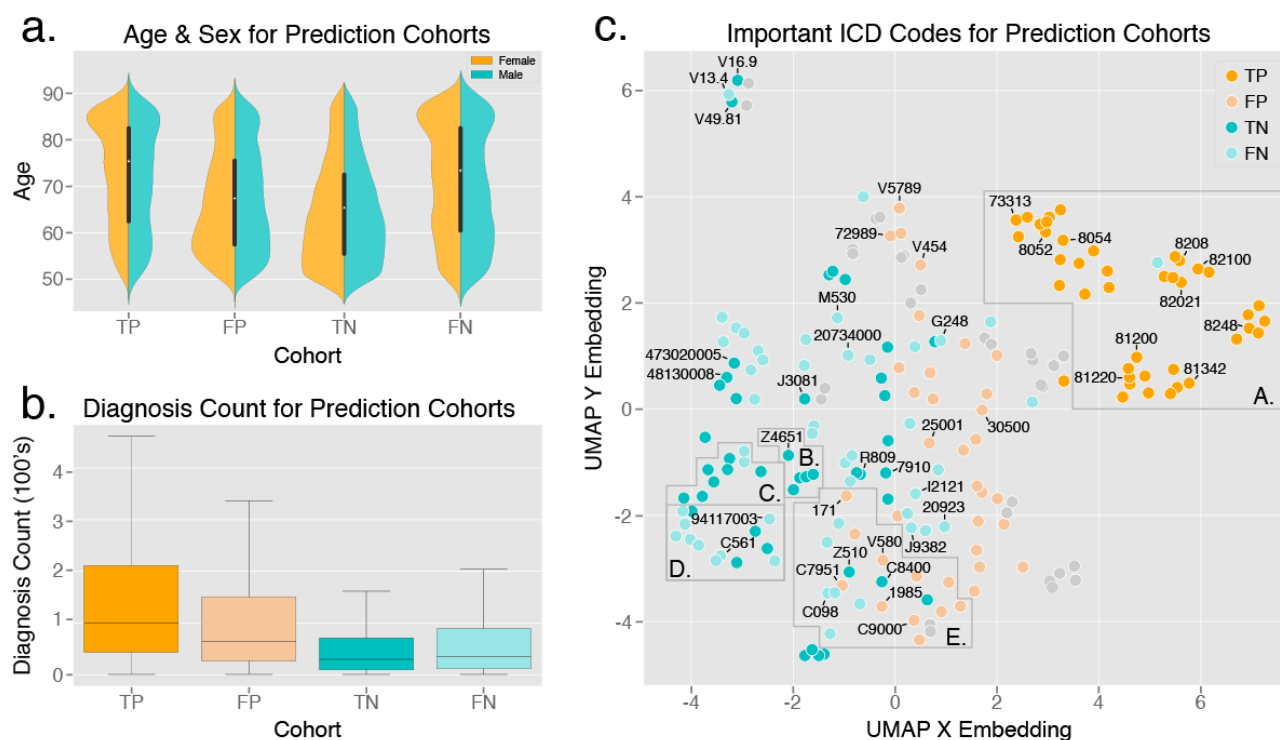
### Limitations of the Current Approach

Various limitations exist for the approaches described, particularly from the inherent complications of using real-world data. The techniques described rely upon ICD codes recorded in electronic health record systems, which will impact the performance and validity of the models if diagnoses are not detected, incorrectly recorded, or missed due to patient dropout. Indeed, most vertebral fragility fractures are clinically silent and hence not captured in electronic health records [70]. While an approach utilizing only ICD codes is potentially more comprehensive and straightforward for real-world implementation due to the quality of coverage and descriptive nature of diagnosis codes, we may miss salient clinical features captured elsewhere in the electronic health record. For example, there exist ICD codes associated with obesity, osteopenia, and osteoporosis, which represent measurements of BMI and bone mineral density on a categorical level. However, these do not reflect exact clinical measurements; the exclusion of these quantitative measurements may limit the performance and clinical impact of the algorithm. Nevertheless, it may be advantageous to utilize these ICD codes rather than the quantitative measures, as such measures in an electronic health record frequently contain human error and may not always be readily available.

In addition to data set challenges, there exist limitations inherent to assumptions of the modeling approach. The suppositions of constant time between diagnosis codes and uniform sequence length may affect performance. Exploration of more advanced methods that do not require such assumptions could improve the model and is an area of future work.

Perhaps the greatest limitation of the described approaches is that they are generally considered black box approaches and lack significant interpretability. Developing methods for improved interpretation of deep learning models is an active area of research. We have performed an initial exploration of this for the ICD code vectorization and long short-term memory model in [Figure 4](#), which compares various characteristics of the four prediction cohorts of the confusion matrix for the test set (true positive [TP], false positive [FP], true negative [TN], false negative [FN]). Within each of these groups, we performed exploratory analysis on the associated samples for each of the input features in the model: age, sex, total diagnosis count, and ICD codes. Results of this analysis are described in detail in [Multimedia Appendix 1](#). While this serves as an initial evaluation of model interpretability, a deeper exploration of interpretability techniques is an area for future work in these algorithms.

**Figure 4.** Exploration of model interpretability by comparison of various characteristics of the input data for the 4 prediction cohorts of the confusion matrix. FN: false negative; FP: false positive; ICD: International Classification of Diseases; TN: true negative; TP: true positive; UMAP: uniform manifold approximation and projection.



Another limitation of this study is the inability to perform direct comparisons with established risk calculators such as FRAX. Additionally, this approach has yet to be validated with external data, which is the subject of future work.

### Potential Applications

We foresee numerous applications of this work in the health care system, with benefits for patients, providers, and payers alike. For payers, Crystal Bone provides a unique opportunity to explore population health, enabling insurers to identify and address patients in need of evaluation or intervention, and preventing the large expenses associated with fracture events. For providers, direct electronic health record integration would facilitate patient care, and help identify at-risk patients who are not currently identified as such. That being said, effective implementation requires additional understanding on the impact of interventions on short-term fracture risk; while there is evidence to suggest that rapid acting treatments and bone-forming agents can significantly decrease fracture risk on a shortened time frame [8,46,47,63-69], a more detailed exploration of the optimal care pathways for various Crystal Bone risk scores would likely be required to facilitate real-world use of the algorithm.

Crystal Bone addresses the need for an automated and largely physician-independent tool that is effective at predicting

short-term fracture risk. It is the first such approach that takes longitudinal patient trajectories into account, rather than focusing primarily on cross-sectional information, enabling a more personalized assessment of fracture risk. Furthermore, with automated aggregation of patient histories in an electronic health record system, the prediction of fracture risk could be entirely hands-off, without requiring a doctor or patient to manually enter any information into the software. This unique approach may facilitate broader adoption of the algorithm. Still, the lack of clinical guidelines for 1- and 2-year risk may limit adoption in the near future.

Such a tool, if widely applied, could facilitate early patient identification, and help reduce the morbidity and mortality associated with fractures. The retrospective human-level performance comparison suggests that Crystal Bone would identify patients who are currently missed in the health care system, potentially minimizing the burden on patients and the health care system overall. Given the prevalence and anticipated increase of fractures due to osteoporosis and low bone mass as the population ages, as well as the enormous personal, clinical, and economic costs associated with such fractures, Crystal Bone could provide a meaningful positive impact through reduced burden and improved outcomes.

### Acknowledgments

This study was funded by Amgen Inc. The costs covered by Amgen Inc were licensing of the Optum data set, access to the computational resources required to develop the model, and compensation for listed Amgen Inc employees. No additional funding was provided for this study.



Thank you to Optum for providing access to and assistance with the data. We would like to additionally thank the following individuals for their guidance and support in conducting this study and creating this manuscript: Inbal Lapid, Tammy Lindberg, Howard Chen, Mandy Suggitt, Lisa Humphries, Marc Doble, Nkem Ogbechie, John Page, Michi He, Akhila Balasubramanian, and Erle Davis.

### Authors' Contributions

YA is the first author. Technical conception, design and direction: YA, AR, PZ, and KW. Medical direction and interpretation: CH, MO, EM, and SRC. Data analysis and interpretation: YA, PZ, AWM, RP and AM. Writing of the manuscript: YA, AR, PZ, KW, CH, MO, EM, and SRC. Authors EM and SR contributed equally. All authors contributed to critical revisions of the draft and approved the final manuscript.

### Conflicts of Interest

YA, PZ, RP, AM, KW, and MO are employees and stock owners at Amgen Inc, the funders of this study. AR, AWM, and CH are former employees and stock owners at Amgen Inc. EM is a consulting fee recipient, grant recipient, and speaker on behalf of Amgen Inc, as well as a member of the International Osteoporosis Foundation. SRC is a consulting fee recipient and grant recipient from Amgen Inc.

### Multimedia Appendix 1

#### Supplementary Information.

[DOC File, 1001 KB - [jmir\\_v22i10e22550\\_app1.doc](https://www.jmir.org/2020/10/e22550_app1.doc)]

### References

1. Haczynski J, Jakimiuk A. Vertebral fractures: a hidden problem of osteoporosis. *Med Sci Monit* 2001;7(5):1108-1117. [Medline: [11535963](#)]
2. Svedbom A, Hernlund E, Ivergård M, Compston J, Cooper C, Stenmark J, EU Review Panel of IOF. Osteoporosis in the European Union: a compendium of country-specific reports. *Arch Osteoporos* 2013 Oct 11;8(1-2):137 [FREE Full text] [doi: [10.1007/s11657-013-0137-0](#)] [Medline: [24113838](#)]
3. Davies KM, Stegman MR, Heaney RP, Recker RR. Prevalence and severity of vertebral fracture: The saunders county bone quality study. *Osteoporosis Int* 1996 Mar;6(2):160-165. [doi: [10.1007/bf01623941](#)]
4. Facts and Statistics 2015. International Osteoporosis Foundation. URL: [https://www.iofbonehealth.org/facts-statistics#:~:text=Osteoporosis%20is%20estimated%20to%20affect,USA%20and%20Japan%20\(1\)](https://www.iofbonehealth.org/facts-statistics#:~:text=Osteoporosis%20is%20estimated%20to%20affect,USA%20and%20Japan%20(1)) [accessed 2020-01-05]
5. Kanis JA, Johnell O, Oden A, Borgstrom F, Zethraeus N, De Laet C, et al. The risk and burden of vertebral fractures in Sweden. *Osteoporos Int* 2004 Jan;15(1):20-26. [doi: [10.1007/s00198-003-1463-7](#)] [Medline: [14593450](#)]
6. Cooper C, Atkinson EJ, O'Fallon WM, Melton JL. Incidence of clinically diagnosed vertebral fractures: a population-based study in Rochester, Minnesota, 1985-1989. *J Bone Miner Res* 1992 Feb;7(2):221-227. [doi: [10.1002/jbmr.5650070214](#)] [Medline: [1570766](#)]
7. Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A. Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. *J Bone Miner Res* 2007 Mar;22(3):465-475 [FREE Full text] [doi: [10.1359/jbmr.061113](#)] [Medline: [17144789](#)]
8. Lötters FJB, van den Bergh JP, de Vries F, Rutten-van Mölken MPMH. Current and future incidence and costs of osteoporosis-related fractures in The Netherlands: combining claims data with BMD measurements. *Calcif Tissue Int* 2016 Mar;98(3):235-243 [FREE Full text] [doi: [10.1007/s00223-015-0089-z](#)] [Medline: [26746477](#)]
9. Rosengren BE, Karlsson MK. The annual number of hip fractures in Sweden will double from year 2002 to 2050. *Acta Orthopaedica* 2014 Apr 30;85(3):234-237. [doi: [10.3109/17453674.2014.916491](#)]
10. Gullberg B, Johnell O, Kanis J. World-wide projections for hip fracture. *Osteoporos Int* 1997 Sep;7(5):407-413. [doi: [10.1007/pl00004148](#)]
11. Papadimitropoulos EA, Coyte PC, Josse RG, Greenwood CE. Current and projected rates of hip fracture in Canada. *CMAJ* 1997 Nov 15;157(10):1357-1363. [Medline: [9371065](#)]
12. Office of the Surgeon General (US). Bone Health and Osteoporosis: A Report of the Surgeon General. Rockville (MD): Office of the Surgeon General (US) 2004:67-105. [Medline: [20945569](#)]
13. What is osteoporosis and what causes it? National Osteoporosis Foundation. 2016. URL: <https://www.nof.org/patients/what-is-osteoporosis/> [accessed 2020-01-05]
14. Harvey N, Earl S, Cooper C. Epidemiology of osteoporotic fractures. In: Favus MJ, editor. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism* 6th ed. Washington, DC: American Society for Bone and Mineral Research; 2006:244-248.
15. Cooper C. The crippling consequences of fractures and their impact on quality of life. *Am J Med* 1997 Aug 18;103(2A):12S-17S. [doi: [10.1016/s0002-9343\(97\)90022-x](#)] [Medline: [9302893](#)]

16. Hip Fracture Outcomes in People Age 50 and Over-Background Paper: OTA-BP-H- 120. US Congress Office of Technology Assessment. Washington, DC: US Government Printing Office; 1994 Jul. URL: <http://www.princeton.edu/~ota/disk1/1994/9413/9413.PDF> [accessed 2020-10-01]
17. Tajeu GS, Delzell E, Smith W, Arora T, Curtis JR, Saag KG, et al. Death, debility, and destitution following hip fracture. *J Gerontol A Biol Sci Med Sci* 2014 Mar;69(3):346-353 [FREE Full text] [doi: [10.1093/gerona/glt105](https://doi.org/10.1093/gerona/glt105)] [Medline: [23873945](#)]
18. Sabesan VJ, Valikodath T, Childs A, Sharma VK. Economic and social impact of upper extremity fragility fractures in elderly patients. *Aging Clin Exp Res* 2015 Aug 24;27(4):539-546. [doi: [10.1007/s40520-014-0295-y](https://doi.org/10.1007/s40520-014-0295-y)] [Medline: [25708827](#)]
19. Ray NF, Chan JK, Thamer M, Melton LJ. Medical expenditures for the treatment of osteoporotic fractures in the United States in 1995: report from the National Osteoporosis Foundation. *J Bone Miner Res* 1997 Jan 01;12(1):24-35 [FREE Full text] [doi: [10.1359/jbmr.1997.12.1.24](https://doi.org/10.1359/jbmr.1997.12.1.24)] [Medline: [9240722](#)]
20. Hall SE, Williams JA, Senior JA, Goldswain PR, Criddle RA. Hip fracture outcomes: quality of life and functional status in older adults living in the community. *Aust N Z J Med* 2000 Jun;30(3):327-332. [doi: [10.1111/j.1445-5994.2000.tb00833.x](https://doi.org/10.1111/j.1445-5994.2000.tb00833.x)] [Medline: [10914749](#)]
21. Marottoli RA, Berkman LF, Cooney LM. Decline in physical function following hip fracture. *J Am Geriatr Soc* 1992 Sep 27;40(9):861-866. [doi: [10.1111/j.1532-5415.1992.tb01980.x](https://doi.org/10.1111/j.1532-5415.1992.tb01980.x)] [Medline: [1512379](#)]
22. Nevitt MC, Ettinger B, Black DM, Stone K, Jamal SA, Ensrud K, et al. The association of radiographically detected vertebral fractures with back pain and function: a prospective study. *Ann Intern Med* 1998 May 15;128(10):793-800. [doi: [10.7326/0003-4819-128-10-199805150-00001](https://doi.org/10.7326/0003-4819-128-10-199805150-00001)] [Medline: [9599190](#)]
23. Pasco JA, Henry MJ, Korn S, Nicholson GC, Kotowicz MA. Morphometric vertebral fractures of the lower thoracic and lumbar spine, physical function and quality of life in men. *Osteoporos Int* 2009 May 19;20(5):787-792. [doi: [10.1007/s00198-008-0744-6](https://doi.org/10.1007/s00198-008-0744-6)] [Medline: [18802658](#)]
24. Fischer S, Kapinos KA, Mulcahy A, Pinto L, Hayden O, Barron R. Estimating the long-term functional burden of osteoporosis-related fractures. *Osteoporos Int* 2017 Oct 24;28(10):2843-2851. [doi: [10.1007/s00198-017-4110-4](https://doi.org/10.1007/s00198-017-4110-4)] [Medline: [28647804](#)]
25. Dyer SM, Crotty M, Fairhall N, Magaziner J, Beaupre LA, Cameron ID, Fragility Fracture Network (FFN) Rehabilitation Research Special Interest Group. A critical review of the long-term disability outcomes following hip fracture. *BMC Geriatr* 2016 Sep 02;16:158 [FREE Full text] [doi: [10.1186/s12877-016-0332-0](https://doi.org/10.1186/s12877-016-0332-0)] [Medline: [27590604](#)]
26. Abimanyi-Ochom J, Watts JJ, Borgström F, Nicholson GC, Shore-Lorenti C, Stuart AL, et al. Changes in quality of life associated with fragility fractures: Australian arm of the International Cost and Utility Related to Osteoporotic Fractures Study (AusICUROS). *Osteoporos Int* 2015 Jun 20;26(6):1781-1790 [FREE Full text] [doi: [10.1007/s00198-015-3088-z](https://doi.org/10.1007/s00198-015-3088-z)] [Medline: [25792491](#)]
27. Brenneman SK, Barrett-Connor E, Sajjan S, Markson LE, Siris ES. Impact of recent fracture on health-related quality of life in postmenopausal women. *J Bone Miner Res* 2006 Jun 06;21(6):809-816 [FREE Full text] [doi: [10.1359/jbmr.060301](https://doi.org/10.1359/jbmr.060301)] [Medline: [16753011](#)]
28. Palacios S, Neyro JL, Fernández de Cabo S, Chaves J, Rejas J. Impact of osteoporosis and bone fracture on health-related quality of life in postmenopausal women. *Climacteric* 2014 Feb 30;17(1):60-70. [doi: [10.3109/13697137.2013.808182](https://doi.org/10.3109/13697137.2013.808182)] [Medline: [23710562](#)]
29. Roux C, Wyman A, Hooven FH, Gehlbach SH, Adachi JD, Chapurlat RD, GLOW investigators. Burden of non-hip, non-vertebral fractures on quality of life in postmenopausal women: the Global Longitudinal study of Osteoporosis in Women (GLOW). *Osteoporos Int* 2012 Dec 8;23(12):2863-2871 [FREE Full text] [doi: [10.1007/s00198-012-1935-8](https://doi.org/10.1007/s00198-012-1935-8)] [Medline: [22398855](#)]
30. Crans GG, Silverman SL, Genant HK, Glass EV, Krege JH. Association of severe vertebral fractures with reduced quality of life: reduction in the incidence of severe vertebral fractures by teriparatide. *Arthritis Rheum* 2004 Dec;50(12):4028-4034 [FREE Full text] [doi: [10.1002/art.20671](https://doi.org/10.1002/art.20671)] [Medline: [15593198](#)]
31. Kado DM, Browner WS, Palermo L, Nevitt MC, Genant HK, Cummings SR. Vertebral fractures and mortality in older women: a prospective study. Study of Osteoporotic Fractures Research Group. *Arch Intern Med* 1999 Jun 14;159(11):1215-1220. [doi: [10.1001/archinte.159.11.1215](https://doi.org/10.1001/archinte.159.11.1215)] [Medline: [10371229](#)]
32. Bentler SE, Liu L, Obrizan M, Cook EA, Wright KB, Geweke JF, et al. The aftermath of hip fracture: discharge placement, functional status change, and mortality. *Am J Epidemiol* 2009 Nov 15;170(10):1290-1299 [FREE Full text] [doi: [10.1093/aje/kwp266](https://doi.org/10.1093/aje/kwp266)] [Medline: [19808632](#)]
33. Bliuc D, Nguyen ND, Nguyen TV, Eisman JA, Center JR. Compound risk of high mortality following osteoporotic fracture and refracture in elderly women and men. *J Bone Miner Res* 2013 Nov 18;28(11):2317-2324 [FREE Full text] [doi: [10.1002/jbmr.1968](https://doi.org/10.1002/jbmr.1968)] [Medline: [23616397](#)]
34. Hu F, Jiang C, Shen J, Tang P, Wang Y. Preoperative predictors for mortality following hip fracture surgery: a systematic review and meta-analysis. *Injury* 2012 Jun;43(6):676-685. [doi: [10.1016/j.injury.2011.05.017](https://doi.org/10.1016/j.injury.2011.05.017)] [Medline: [21683355](#)]
35. Jette AM, Harris BA, Cleary PD, Campion EW. Functional recovery after hip fracture. *Arch Phys Med Rehabil* 1987 Oct;68(10):735-740. [Medline: [3662784](#)]

36. Leibson CL, Tosteson ANA, Gabriel SE, Ransom JE, Melton LJ. Mortality, disability, and nursing home use for persons with and without hip fracture: a population-based study. *J Am Geriatr Soc* 2002 Oct 17;50(10):1644-1650. [doi: [10.1046/j.1532-5415.2002.50455.x](https://doi.org/10.1046/j.1532-5415.2002.50455.x)] [Medline: [12366617](https://pubmed.ncbi.nlm.nih.gov/12366617/)]
37. Cooper C, Atkinson EJ, Jacobsen SJ, O'Fallon WM, Melton LJ. Population-based study of survival after osteoporotic fractures. *Am J Epidemiol* 1993 May 01;137(9):1001-1005. [doi: [10.1093/oxfordjournals.aje.a116756](https://doi.org/10.1093/oxfordjournals.aje.a116756)] [Medline: [8317445](https://pubmed.ncbi.nlm.nih.gov/8317445/)]
38. Morin S, Lix LM, Azimaee M, Metge C, Caetano P, Leslie WD. Mortality rates after incident non-traumatic fractures in older men and women. *Osteoporos Int* 2011 Sep 16;22(9):2439-2448. [doi: [10.1007/s00198-010-1480-2](https://doi.org/10.1007/s00198-010-1480-2)] [Medline: [21161507](https://pubmed.ncbi.nlm.nih.gov/21161507/)]
39. van Geel TACM, Eisman JA, Geusens PP, van den Bergh JPW, Center JR, Dinant GJ. The utility of absolute risk prediction using FRAX® and Garvan Fracture Risk Calculator in daily practice. *Maturitas* 2014 Feb;77(2):174-179. [doi: [10.1016/j.maturitas.2013.10.021](https://doi.org/10.1016/j.maturitas.2013.10.021)] [Medline: [24287178](https://pubmed.ncbi.nlm.nih.gov/24287178/)]
40. van Geel TA, van den Bergh JPW, Dinant GJ, Geusens PP. Individualizing fracture risk prediction. *Maturitas* 2010 Feb;65(2):143-148. [doi: [10.1016/j.maturitas.2009.12.007](https://doi.org/10.1016/j.maturitas.2009.12.007)] [Medline: [20053513](https://pubmed.ncbi.nlm.nih.gov/20053513/)]
41. Leslie WD, Lix LM. Comparison between various fracture risk assessment tools. *Osteoporos Int* 2014 Jan 25;25(1):1-21. [doi: [10.1007/s00198-013-2409-3](https://doi.org/10.1007/s00198-013-2409-3)] [Medline: [23797847](https://pubmed.ncbi.nlm.nih.gov/23797847/)]
42. Leslie WD, Berger C, Langsetmo L, Lix LM, Adachi JD, Hanley DA, Canadian Multicentre Osteoporosis Study Research Group. Construction and validation of a simplified fracture risk assessment tool for Canadian women and men: results from the CaMos and Manitoba cohorts. *Osteoporos Int* 2011 Jun 22;22(6):1873-1883 [FREE Full text] [doi: [10.1007/s00198-010-1445-5](https://doi.org/10.1007/s00198-010-1445-5)] [Medline: [20967422](https://pubmed.ncbi.nlm.nih.gov/20967422/)]
43. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. *BMJ* 2012 May 22;344(may22 1):e3427-e3427. [doi: [10.1136/bmj.e3427](https://doi.org/10.1136/bmj.e3427)] [Medline: [22619194](https://pubmed.ncbi.nlm.nih.gov/22619194/)]
44. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. *Osteoporos Int* 2008 Apr 22;19(4):385-397 [FREE Full text] [doi: [10.1007/s00198-007-0543-5](https://doi.org/10.1007/s00198-007-0543-5)] [Medline: [18292978](https://pubmed.ncbi.nlm.nih.gov/18292978/)]
45. NOGG 2017: Clinical guideline for the prevention and treatment of osteoporosis. National Osteoporosis Guideline Group. 2017. URL: <https://www.sheffield.ac.uk/NOGG/NOGG%20Guideline%202017.pdf> [accessed 2020-01-05]
46. Lewiecki EM, Laster AJ, Miller PD, Bilezikian JP. More bone density testing is needed, not less. *J Bone Miner Res* 2012 Apr 20;27(4):739-742 [FREE Full text] [doi: [10.1002/jbmr.1580](https://doi.org/10.1002/jbmr.1580)] [Medline: [22383457](https://pubmed.ncbi.nlm.nih.gov/22383457/)]
47. Siris ES, Boonen S, Mitchell PJ, Bilezikian J, Silverman S. What's in a name? What constitutes the clinical diagnosis of osteoporosis? *Osteoporos Int* 2012 Aug 28;23(8):2093-2097. [doi: [10.1007/s00198-012-1991-0](https://doi.org/10.1007/s00198-012-1991-0)] [Medline: [22543575](https://pubmed.ncbi.nlm.nih.gov/22543575/)]
48. Chen Y, Miller PD, Barrett-Connor E, Weiss TW, Sajjan SG, Siris ES. An approach for identifying postmenopausal women age 50-64 years at increased short-term risk for osteoporotic fracture. *Osteoporos Int* 2007 Sep 27;18(9):1287-1296. [doi: [10.1007/s00198-007-0380-6](https://doi.org/10.1007/s00198-007-0380-6)] [Medline: [17464525](https://pubmed.ncbi.nlm.nih.gov/17464525/)]
49. Miller PD, Barlas S, Breneman SK, Abbott TA, Chen Y, Barrett-Connor E, et al. An approach to identifying osteopenic women at increased short-term risk of fracture. *Arch Intern Med* 2004 May 24;164(10):1113-1120. [doi: [10.1001/archinte.164.10.1113](https://doi.org/10.1001/archinte.164.10.1113)] [Medline: [15159269](https://pubmed.ncbi.nlm.nih.gov/15159269/)]
50. Black DM, Steinbuch M, Palermo L, Dargent-Molina P, Lindsay R, Hoseyni MS, et al. An assessment tool for predicting fracture risk in postmenopausal women. *Osteoporos Int* 2001 Aug 1;12(7):519-528. [doi: [10.1007/s001980170072](https://doi.org/10.1007/s001980170072)] [Medline: [11527048](https://pubmed.ncbi.nlm.nih.gov/11527048/)]
51. Ferizi U, Honig S, Chang G. Artificial intelligence, osteoporosis and fragility fractures. *Current Opinion in Rheumatology* 2019;31(4):368-375. [doi: [10.1097/bor.0000000000000607](https://doi.org/10.1097/bor.0000000000000607)]
52. Kruse C, Eiken P, Vestergaard P. Machine Learning Principles Can Improve Hip Fracture Prediction. *Calcif Tissue Int* 2017 Apr 14;100(4):348-360. [doi: [10.1007/s00223-017-0238-7](https://doi.org/10.1007/s00223-017-0238-7)] [Medline: [28197643](https://pubmed.ncbi.nlm.nih.gov/28197643/)]
53. Kim S, Yoo T, Oh E, Kim D. Osteoporosis risk prediction using machine learning and conventional methods. *Conf Proc IEEE Eng Med Biol Soc. - 2013*;188:191. [doi: [10.1109/embc.2013.6609469](https://doi.org/10.1109/embc.2013.6609469)] [Medline: [24109656](https://pubmed.ncbi.nlm.nih.gov/24109656/)]
54. Wright NC, Daigle SG, Melton ME, Delzell ES, Balasubramanian A, Curtis JR. The design and validation of a new algorithm to identify incident fractures in administrative claims data. *J Bone Miner Res* 2019 Oct;34(10):1798-1807 [FREE Full text] [doi: [10.1002/jbmr.3807](https://doi.org/10.1002/jbmr.3807)] [Medline: [31170317](https://pubmed.ncbi.nlm.nih.gov/31170317/)]
55. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv.org* 2013 Sep 07:1-12 [FREE Full text]
56. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. *JOSS* 2018 Sep;3(29):861. [doi: [10.21105/joss.00861](https://doi.org/10.21105/joss.00861)]
57. Nowak J, Taspinar A, Scherer R. LSTM Recurrent Neural Networks for Short Text and Sentiment Classification. In: Rutkowski L, Korytkowski M, Scherer R, Tadeusiewicz R, Zadeh L, Zurada J, editors. *Artificial Intelligence and Soft Computing. 16th International Conference, ICAISC 2017, Zakopane, Poland, June 11-15, 2017, Proceedings, Part II*. Cham: Springer International Publishing; Jun 15, 2017:553-562.
58. Le QV, Mikolov T. Distributed representations of sentences and documents. *Stanford University Quoc Le Profile*. URL: [https://cs.stanford.edu/~quocle/paragraph\\_vector.pdf](https://cs.stanford.edu/~quocle/paragraph_vector.pdf) [accessed 2020-10-07]

59. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.: ACM Press; 2016 Aug Presented at: International Conference on Knowledge Discovery and Data Mining (KDD); August 13-17, 2020; San Francisco p. 785-794 URL: <https://dl.acm.org/doi/10.1145/2939672.2939785> [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
60. Cosman F, de Beur SJ, LeBoff MS, Lewiecki EM, Tanner B, Randall S, National Osteoporosis Foundation. Clinician's guide to prevention and treatment of osteoporosis. *Osteoporos Int* 2014 Oct 15;25(10):2359-2381 [FREE Full text] [doi: [10.1007/s00198-014-2794-2](https://doi.org/10.1007/s00198-014-2794-2)] [Medline: [25182228](https://pubmed.ncbi.nlm.nih.gov/25182228/)]
61. Eastell R, Rosen CJ, Black DM, Cheung AM, Murad MH, Shoback D. Pharmacological Management of Osteoporosis in Postmenopausal Women: An Endocrine Society\* Clinical Practice Guideline. *J Clin Endocrinol Metab* 2019 May 01;104(5):1595-1622. [doi: [10.1210/jc.2019-00221](https://doi.org/10.1210/jc.2019-00221)] [Medline: [30907953](https://pubmed.ncbi.nlm.nih.gov/30907953/)]
62. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020 Mar 25;m689. [doi: [10.1136/bmj.m689](https://doi.org/10.1136/bmj.m689)]
63. Hodsman AB, Bauer DC, Dempster DW, Dian L, Hanley DA, Harris ST, et al. Parathyroid hormone and teriparatide for the treatment of osteoporosis: a review of the evidence and suggested guidelines for its use. *Endocr Rev* 2005 Aug;26(5):688-703. [doi: [10.1210/er.2004-0006](https://doi.org/10.1210/er.2004-0006)] [Medline: [15769903](https://pubmed.ncbi.nlm.nih.gov/15769903/)]
64. Forteo Package Insert. Lilly USA, LLC. 2020 Apr 06. URL: <https://pi.lilly.com/us/forteo-pi.pdf> [accessed 2020-05-11]
65. Neer RM, Arnaud CD, Zanchetta JR, Prince R, Gaich GA, Reginster JY, et al. Effect of parathyroid hormone (1-34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med* 2001 May 10;344(19):1434-1441. [doi: [10.1056/NEJM200105103441904](https://doi.org/10.1056/NEJM200105103441904)] [Medline: [11346808](https://pubmed.ncbi.nlm.nih.gov/11346808/)]
66. Lindsay R, Kregg JH, Marin F, Jin L, Stepan JJ. Teriparatide for osteoporosis: importance of the full course. *Osteoporos Int* 2016 Aug;27(8):2395-2410 [FREE Full text] [doi: [10.1007/s00198-016-3534-6](https://doi.org/10.1007/s00198-016-3534-6)] [Medline: [26902094](https://pubmed.ncbi.nlm.nih.gov/26902094/)]
67. Tymlos Package Insert. Radius Health, Inc. 2018 Oct. URL: <https://radiuspharm.com/wp-content/uploads/tymlos/tymlos-prescribing-information.pdf> [accessed 2020-05-10]
68. Miller PD, Hattersley G, Riis BJ, Williams GC, Lau E, Russo LA, ACTIVE Study Investigators. Effect of abaloparatide vs placebo on new vertebral fractures in postmenopausal women with osteoporosis: a randomized clinical trial. *JAMA* 2016 Aug 16;316(7):722-733. [doi: [10.1001/jama.2016.11136](https://doi.org/10.1001/jama.2016.11136)] [Medline: [27533157](https://pubmed.ncbi.nlm.nih.gov/27533157/)]
69. Kendler DL, Marin F, Zerbini CAF, Russo LA, Greenspan SL, Zikan V, et al. Effects of teriparatide and risedronate on new fractures in post-menopausal women with severe osteoporosis (VERO): a multicentre, double-blind, double-dummy, randomised controlled trial. *Lancet* 2018 Jan 20;391(10117):230-240. [doi: [10.1016/S0140-6736\(17\)32137-2](https://doi.org/10.1016/S0140-6736(17)32137-2)] [Medline: [29129436](https://pubmed.ncbi.nlm.nih.gov/29129436/)]
70. Ballane G, Cauley JA, Luckey MM, El-Hajj Fuleihan G. Worldwide prevalence and incidence of osteoporotic vertebral fractures. *Osteoporos Int* 2017 Feb 6;28(5):1531-1542. [doi: [10.1007/s00198-017-3909-3](https://doi.org/10.1007/s00198-017-3909-3)]

## Abbreviations

**AUPRC:** area under the precision-recall curve

**AUROC:** area under the receiver operating characteristics curve

**FRAX:** University of Sheffield Fracture Risk Assessment Tool

**GIH-BFRC:** Garvan Institute of Health Bone Fracture Risk Calculator

**ICD:** International Classification of Diseases

*Edited by G Eysenbach; submitted 15.07.20; peer-reviewed by C Fincham, M Pradhan; comments to author 06.08.20; revised version received 05.09.20; accepted 12.09.20; published 16.10.20.*

### *Please cite as:*

Almog YA, Rai A, Zhang P, Moulaison A, Powell R, Mishra A, Weinberg K, Hamilton C, Oates M, McCloskey E, Cummings SR. Deep Learning With Electronic Health Records for Short-Term Fracture Risk Identification: Crystal Bone Algorithm Development and Validation

*J Med Internet Res* 2020;22(10):e22550

URL: <http://www.jmir.org/2020/10/e22550/>

doi: [10.2196/22550](https://doi.org/10.2196/22550)

PMID: [32956069](https://pubmed.ncbi.nlm.nih.gov/32956069/)

©Yasmeen Adar Almog, Angshu Rai, Patrick Zhang, Amanda Moulaison, Ross Powell, Anirban Mishra, Kerry Weinberg, Celeste Hamilton, Mary Oates, Eugene McCloskey, Steven R Cummings. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 16.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution

License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Integrating the Practical Robust Implementation and Sustainability Model With Best Practices in Clinical Decision Support Design: Implementation Science Approach

Katy E Trinkley<sup>1,2,3,4</sup>, PharmD, PhD; Michael G Kahn<sup>5</sup>, PhD, MD; Tellen D Bennett<sup>4,5</sup>, MS, MD; Russell E Glasgow<sup>4,6</sup>, PhD; Heather Haugen<sup>7</sup>, PhD; David P Kao<sup>2,3</sup>, MD; Miranda E Kroehl<sup>8</sup>, PhD; Chen-Tan Lin<sup>2,3</sup>, MD; Daniel C Malone<sup>9</sup>, RPh, PhD; Daniel D Matlock<sup>2,4,10</sup>, MD, MPH

<sup>1</sup>Department of Clinical Pharmacy, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

<sup>2</sup>Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

<sup>3</sup>Clinical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

<sup>4</sup>Adult and Child Consortium for Outcomes Research and Delivery Science, Aurora, CO, United States

<sup>5</sup>Section of Informatics and Data Science, Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

<sup>6</sup>Department of Family Medicine, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

<sup>7</sup>Colorado Clinical and Translational Sciences Institute, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

<sup>8</sup>Charter Communications Corporation, Greenwood Village, CO, United States

<sup>9</sup>Department of Pharmacotherapy, Skaggs College of Pharmacy, University of Utah, Salt Lake City, UT, United States

<sup>10</sup>VA Eastern Colorado Geriatric Research Education and Clinical Center, Aurora, CO, United States

## Corresponding Author:

Katy E Trinkley, PharmD, PhD

Department of Clinical Pharmacy

Skaggs School of Pharmacy and Pharmaceutical Sciences

University of Colorado Anschutz Medical Campus

12850 E Montview Blvd C238

Aurora, CO, 80045

United States

Phone: 1 3037246563

Email: [katy.trinkley@cuanschutz.edu](mailto:katy.trinkley@cuanschutz.edu)

## Abstract

**Background:** Clinical decision support (CDS) design best practices are intended to provide a narrative representation of factors that influence the success of CDS tools. However, they provide incomplete direction on evidence-based implementation principles.

**Objective:** This study aims to describe an integrated approach toward applying an existing implementation science (IS) framework with CDS design best practices to improve the effectiveness, sustainability, and reproducibility of CDS implementations.

**Methods:** We selected the Practical Robust Implementation and Sustainability Model (PRISM) IS framework. We identified areas where PRISM and CDS design best practices complemented each other and defined methods to address each. Lessons learned from applying these methods were then used to further refine the integrated approach.

**Results:** Our integrated approach to applying PRISM with CDS design best practices consists of 5 key phases that iteratively interact and inform each other: multilevel stakeholder engagement, designing the CDS, design and usability testing, thoughtful deployment, and performance evaluation and maintenance. The approach is led by a dedicated implementation team that includes clinical informatics and analyst builder expertise.

**Conclusions:** Integrating PRISM with CDS design best practices extends user-centered design and accounts for the multilevel, interacting, and dynamic factors that influence CDS implementation in health care. Integrating PRISM with CDS design best practices synthesizes the many known contextual factors that can influence the success of CDS tools, thereby enhancing the reproducibility and sustainability of CDS implementations. Others can adapt this approach to their situation to maximize and sustain CDS implementation success.

**KEYWORDS**

clinical decision support; PRISM; implementation science

## Introduction

### Background

Clinical decision support (CDS) tools within electronic health records (EHRs) have led to some improvements in patient care [1-4]. However, there are also numerous examples of CDS tools leading to low adoption or negative impact on outcomes. Up to 95% of CDS tools are dismissed [5], 52%-66% improve process outcomes such as appropriate drug selection, and only 21%-43%

lead to improvements in clinical outcomes [6-9]. CDS design best practices may be a way to improve the impact of CDS tools [10,11]. As a framework, CDS design best practices are intended to provide a narrative representation of the key determinants that influence the success of CDS tools [12-18]. [Textbox 1](#) provides a high-level summary of the best practices that include a user-centered design process. Initially established by experts in the field, retrospective studies support the potential benefit of applying CDS design best practices [13-16,19,20].

**Textbox 1.** Overview of clinical decision support design best practices and examples of how to address the best practice.

#### Minimize alert fatigue

- Ensure accuracy and completeness of data and information recommended and used to formulate recommendation. Continual performance evaluation and end user feedback throughout implementation. Evaluations and feedback should be used to iteratively update the clinical decision support (CDS). CDS customization to fit the end user and institutional needs, including type (interruptive or passive), presence of a dismiss option, and frequency and timing of alert. Consideration of whether the CDS can support agenda setting

#### Support team-based care

- Comprehensive inclusion of care team members in which the CDS is tailored to the workflows and roles of each member

#### Fit within the end user's workflow when considering other internal and external drivers

- Presentation of the CDS in such a way that it is available when needed, supports (versus impedes) end users, and human factors principles that makes it easy for them to synthesize and apply the information displayed (eg, visual cues such as size, position or color; prioritization; standardization). Flexibility to delay or defer CDS to another time or person can optimize workflow integration. Response options that reflect all possible patient situations (eg, other). Obtaining end user input on the design (user-centered design) with consideration of other internal and external factors such as national guidelines or value-based performance measures

#### Present pertinent and transparent information that supports and does not impair autonomy of decision-making

- Provide the rationale and supporting information (eg, references) on why the CDS is displayed so as to allow the end user to evaluate whether to apply the recommendation. Avoid giving a perception of shaming or use of insulting language

#### Make it easy, and incentivize users to follow the recommendation

- Provide actionable recommendations and functionality to save the end user time (eg, ability to place orders within display; links to review or update patient data). Provide a relative advantage to using the CDS (eg, peer or patient recognition for actions taken; save time).

CDS design best practices acknowledge the importance of implementation but do not provide a comprehensive framework or direction on evidence-based implementation principles. Some implementation science (IS) frameworks are available for health information technologies (health IT) [21-24]. However, there are no published reports that are comprehensive in providing direction for all stages of health IT or CDS implementation. A comprehensive IS framework could provide direction on how to account for field-specific nuances and contextual factors that influence the success of CDS tools [25]. Such an IS framework can allow health systems to proactively address the areas where things often break down in the process of adoption, implementation, and maintenance.

### Objectives

Here, we describe how to apply an existing IS framework to CDS to improve effectiveness, sustainability, and reproducibility. The specific objectives of this report are: (1) discuss how an IS framework can be integrated with CDS design best practices, (2) describe how to apply this integrated approach using illustrative case studies, and (3) discuss directions for future research and application of this integrated approach to CDS implementation.

## Methods

### Selection of an IS Framework

Although there are many IS frameworks [26], we selected the Practical Robust Implementation and Sustainability Model (PRISM) because it: (1) is a process, evaluation and

determinants framework, (2) comprehensively addresses the interactions between the intervention and stakeholders and both organizational and external factors, (3) is directly tied to real-world, pragmatic implementation outcomes, and (4) is easy to use, thereby maximizing scalability [26-30]. As a process model, PRISM provides direction on how to address factors that influence implementation. As a determinant framework, PRISM makes it possible to reproduce the process for considering key factors that may influence implementation success [26]. As a comprehensive framework, PRISM considers all stages of implementation (preimplementation planning and

design, implementation operations, postimplementation evaluation) and all groups or levels of influences within and external to the organization. As illustrated below, it is also compatible with CDS design best practices. Finally, PRISM builds on the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework [31] for key implementation outcome measures. [Textbox 2](#) describes the domains of PRISM and how they apply to CDS. [Figure 1](#) depicts the domains of PRISM, their interactions, and how they influence CDS.

**Textbox 2.** PRISM domains and application to clinical decision support.

**Intervention: organizational perspective includes leadership, management, clinicians, and frontline staff**

An intervention is more likely to be successful if:

- It is aligned with the organization's mission and readiness for change
- The strength of evidence supporting the intervention is strong
- It addresses a barrier to or gap in health care
- It has been observed to be beneficial before a long-term commitment (observability, trialability, reversibility)
- It is simple and inexpensive

**Intervention: patient perspective**

An intervention is more likely to be successful if it is:

- Patient centered
- Simple and inexpensive
- Accessible to and understood by a wide variety of patients (cultural backgrounds, literacy, or numeracy levels)
- Addresses key patient concerns, not limited to clinical issues

**Recipients: organizational characteristics includes leadership, management, clinicians, and frontline staff**

Characteristics of the organization can impact the success of an intervention, such as financial health, tendency to take risks or deviate from the norm, and morale.

An intervention is more likely to be successful when:

- Management is supportive
- Goals are cohesive and clearly communicated across the organization
- Input is provided across all levels of the organization, including senior leadership, midlevel management, and pertinent frontline clinicians and staff

**Recipients: patient characteristics**

- Characteristics of patients can impact the success of an intervention, including socioeconomic factors such as affordability and access or transportation barriers to the intervention

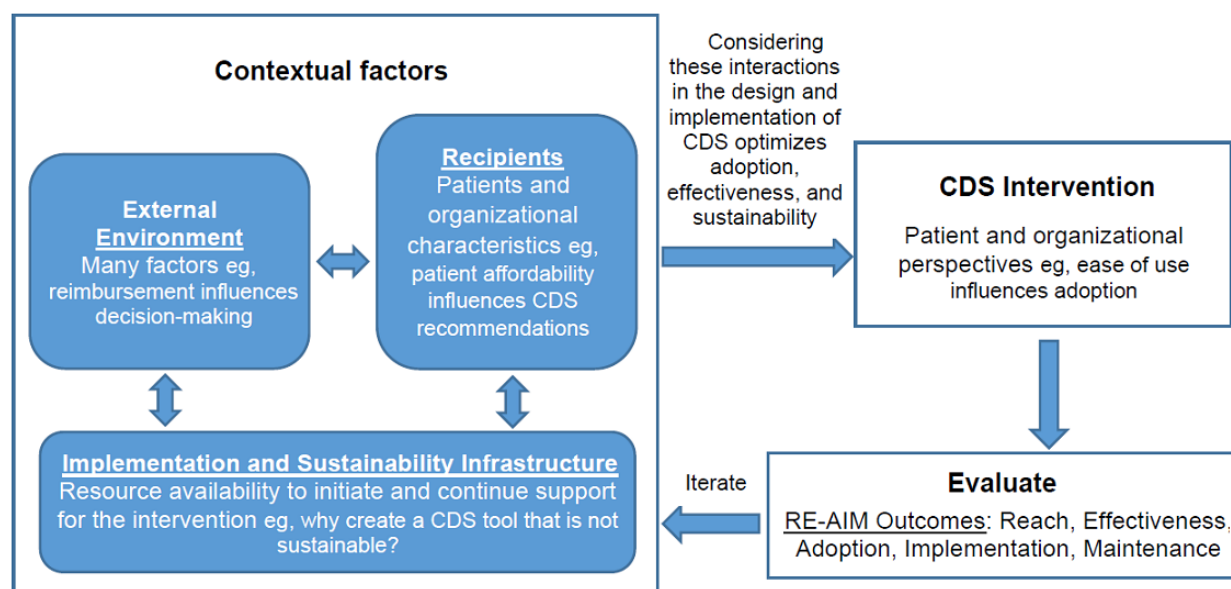
**External environment**

- Factors outside of the organization can influence the organization, such as reporting on performance metrics (public face), policy, guidelines, and reimbursement issues

**Implementation and sustainability infrastructure**

- The implementation plan should be carefully crafted with a dedicated team for implementation, input from management and other stakeholders, and consideration of sustainability and dissemination from the beginning.
- Adequate resources and ongoing assessment or audit and feedback system should be in place

**Figure 1.** Domains of Practical Robust Implementation and Sustainability Model, their interactions, and how they influence clinical decision support. CDS: clinical decision support.



### Integration of the PRISM With Clinical Decision Support Design Best Practices

To account for the specific contextual factors of CDS in health care, we integrated PRISM with CDS design best practices, which we hereafter refer to as the PRISM/CDS best practice approach. Table 1 highlights areas where PRISM and CDS design best practices complement each other and the corresponding methods to address each. For example, clinician focus groups address both the PRISM domain of the *intervention* from the organizational perspective and CDS design best practice principle of *support team-based care*. Therefore, the focus group should include questions to understand who is involved in the specific care process, organizational barriers to prescribing, and whether the subject matter experts (SMEs) feel

the strength of evidence is strong and clinically relevant. When performing the cross-walk between the 2 frameworks, there are situations where the individual principles and domains do not complement each other (empty cell in Table 1), which is appropriate. For example, when designing a CDS for clinicians to use, methods to address the best practice principle of *fit within end users workflow* would generally be unrelated to methods used to address the PRISM domain of *intervention* from the patient perspective. In general, instances in which cells are empty are because 1 framework focuses on the clinician's perspective, the other focuses on the patient perspective, and there is no common area or overlap that integrates the 2 perspectives. Lessons learned from applying the methods outlined in Table 1 were then used to further refine the PRISM/CDS best practice approach.

**Table 1.** Practical Robust Implementation and Sustainability Model and clinical decision support design best practices: complementary areas and corresponding methods to address.

PRISM <sup>a</sup> domains		Overarching CDS <sup>b</sup> design best practice principles			
	Minimize alert fatigue	Support team-based care	Fit within the end user's workflow when considering other internal and external drivers	Present pertinent and transparent information that supports and does not impair autonomy of decision making	Make it easy and incentivize users to follow the recommendation
Intervention: organizational perspective	EU <sup>c</sup> /clinician focus groups	EU/clinician focus groups	EU/clinician focus groups	EU/clinician design/usability testing	EU/clinician design/usability testing
	EU/clinician usability testing	EU/clinician design/usability testing	EU/clinician design/usability testing		
Intervention: patient perspective	N/A <sup>d</sup>	Patient focus groups and interviews	N/A	N/A	EU/patient focus groups and interviews
Recipients: organizational characteristics	N/A	EU/clinician focus groups	EU/clinician focus groups	EU/clinician design/usability testing	EU/clinician design/usability testing
			Clinician design/usability testing		
Recipients: patient characteristic	N/A	EU/clinician design/usability testing	Early engagement of leadership/ management		
		EU/clinician focus groups	EU/clinician focus groups	N/A	N/A
External environment	N/A	Patient focus groups			
		N/A	Alignment with national payor and guideline metrics	N/A	Alignment with national payor and guideline metrics
Implementation and sustainability infrastructure	Scheduled performance evaluation and update	EU/clinician design/usability testing	EU/clinician design/usability testing	EU/clinician design/usability testing	EU/clinician design/usability testing including testing of training materials

<sup>a</sup>PRISM: Practical Robust Implementation and Sustainability Model.<sup>b</sup>CDS: clinical decision support.<sup>c</sup>EU: end user.<sup>d</sup>N/A: Not applicable. Situations where the individual principles and domains do not complement each other.

## Case Study

Throughout this paper, we will refer to a CDS tool to improve the prescription of evidence-based beta blockers for patients with heart failure and reduced ejection fraction. CDS was deployed within the EHR of primary care practices across a large regional health system. The PRISM/CDS best practice approach was applied to the design and implementation of the CDS, as described in the 5 phases below. We note that the methods and results of a randomized controlled trial (RCT) evaluating the PRISM/CDS best practices approach are under review separately, but our focus here is on the application and integration of the approach, not outcome results.

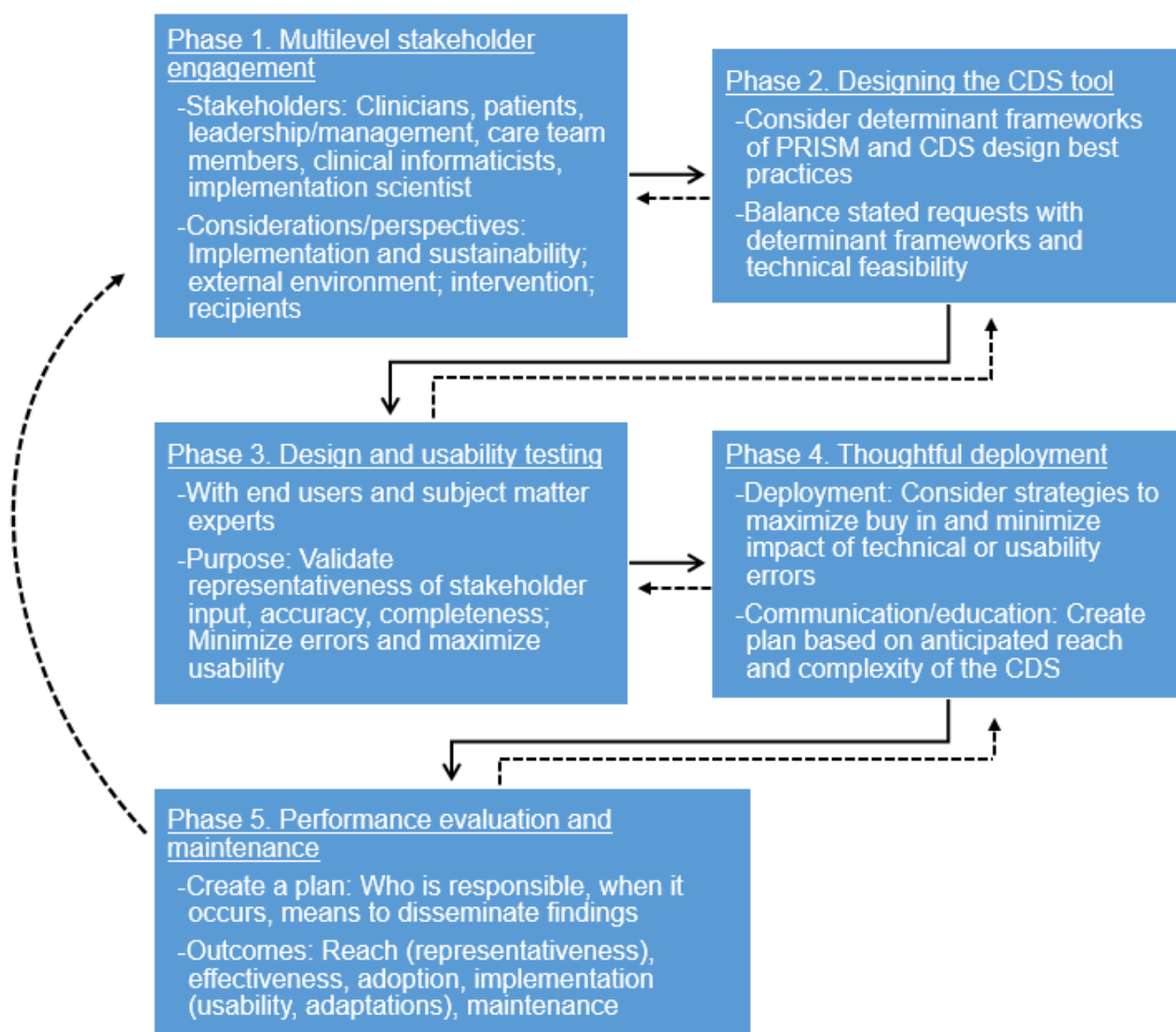
## Results

### Overview

Our integrated approach to applying PRISM to CDS consists of 5 phases: (1) multilevel stakeholder engagement, (2) designing the CDS tool, (3) design and usability testing, (4) thoughtful deployment, and (5) performance evaluation and maintenance. Although there is some logical sequence to these phases, the process is not linear. The phases interact and iteratively inform each other. [Figure 2](#) provides an overview of the phases, which emphasizes that the process is iterative and agile. Here, we describe each of the phases and the key determinants of implementation success from PRISM and CDS design best practices. These phases are intended to be adapted for each unique health system and CDS implementation.



**Figure 2.** Phases of applying the Practical Robust Implementation and Sustainability Model to clinical decision support implementation. CDS: clinical decision support.



### Phase 1. Multilevel Stakeholder Engagement

Multilevel stakeholder engagement is central to PRISM and extends the concept of user-centered design by incorporating stakeholders beyond the end user. PRISM emphasizes factors that can influence implementation success: perspectives of the patient and organization as well as characteristics of the recipients. Patient and organizational beliefs, technology support, and patient ability to afford medical care are some of the characteristics that can influence CDS implementation (Textbox 2 and Figure 1). Following PRISM, it is important to cultivate positive stakeholder perspectives of the CDS tool and identify recipient characteristics that can influence implementation success.

The types of stakeholders will vary for each CDS implementation and can range from a small homogenous group to large groups representing diverse disciplines and specialties. To build CDS tools that are trusted by end users, engaging the right stakeholders with the necessary skills is critical [32]. Stakeholders unique to health IT and CDS include health IT

leadership or governance, clinical informaticists, and analyst builders. Of these stakeholders, a clinical informaticist and analyst builder should be part of a dedicated implementation team, whereas the role of health IT leadership or governance is generally to provide support and approval for implementation. The dedicated implementation team is a key component of the PRISM's implementation and sustainability infrastructure. Often overlooked, clinical informaticists span across disciplines and are uniquely trained to empathize with clinicians, apply the determinants of effective CDS implementation, and balance what is asked for with what is technically possible [33-37]. Clinical informaticists should ideally have experience in IS or an IS expert should be engaged.

Gaining early support from health IT leadership is generally the first step, followed by input from SMEs and end users, and then formal governance approval. When patients are not the end users, their engagement is likely best reserved until after formal governance approval, just in case approval is not granted. The system's health IT leadership assists in determining whether a particular CDS tool is generally aligned with the system's

priorities and whether resources are available to build the CDS tool. Health IT leadership can also provide direction regarding steps needed to secure necessary approval from the health system's formal governance process. Although not all health systems have a formal health IT governance approval process, an increasing number do [38,39]. The organization and process for securing governance approval can vary greatly across health systems and can be complex; thus, early understanding is important. The formal governance approval process includes a review of CDS appropriateness from a workflow and safety perspective and considers potential overlapping or competing system-level initiatives.

SME and end user engagement is a dynamic process that often requires the clinical informaticist, leading the effort to iteratively engage each party until consensus on the design of the CDS tool is established. Obtaining input from SMEs is key to validating the clinical evidence, whereas input from end users is key to ensuring acceptance and practicality of the CDS tool. End user engagement is key to optimizing the workflow

integration. The workflow and preferences of clinician end users can vary greatly across disciplines (eg, respiratory therapists, physicians) and practice settings (eg, acute care, dialysis center). The type of information elicited from patients or caregivers will depend on the type of CDS, notably whether the patient will be an end user or whether the CDS tool makes recommendations that might impact patient treatment decisions. Although patient-facing CDS tools are becoming more common, the majority are still clinician facing. When clinician facing, the line of questioning directed at patients or caregivers is on treatment priorities and values to ensure alignment with the CDS tool's recommendations. When conducting focus groups or interviews with stakeholders, questions should be tailored to the situation [40] and informed by the determinants of effective implementation, which include CDS design best practices. The objectives of SMEs and end user engagement are to garner their support and define the general scope of the CDS tool. Such a written CDS scope can be used to facilitate the formal governance approval process. [Textbox 3](#) describes our approach to engaging stakeholders.

### Textbox 3. Phase 1 case study.

#### Case study:

- For the beta-blocker CDS tool, the Chief Medical Information Officer (CMIO) was contacted at the start of the project. The CMIO provided support contingent on approval from other stakeholders and the formal governance process. To engage patient and clinician stakeholders in preimplementation design activities, we conducted focus groups rather than individual interviews to maximize the efficiency of resources and to facilitate idea generation among participants. However, we also sought to elicit individual patient and clinician thoughts.
- Clinicians, after the open-ended group discussion regarding needs and preferences for a CDS tool, were asked to individually design on paper their ideal CDS [41], which provided valuable insights. For example, most clinicians expressed strong dislike for interruptive CDS during group discussion, but many individually described their ideal CDS as being interruptive. The discordance may be the result of peer influence or reflect differences between their preferences and their perceptions of what is most effective. These findings led to an interruptive CDS tool [41].
- In our case, the CDS tool was clinician facing; thus, the goal of the patient focus groups was to evaluate the factors that influence patients' decisions to take heart failure medications. The focus groups occurred early in the design process to ensure that the prototypes were driven by patient-centered factors. We found that patient values and preferences for heart failure medications aligned with clinical guideline recommendations that prioritize benefits over risks, cost, and the inconvenience of taking medications. These findings provided reassurance that we were designing a CDS tool that reflected patient priorities.

## Phase 2. Designing the CDS Tool

Following PRISM, stakeholder input is used to design the intervention to maximize the recipient's perceived value and ensure sufficient support infrastructure ([Textbox 2](#)). Such interventions are more likely to be successful and sustained over time. PRISM emphasizes the importance of designing interventions with sustainability in mind.

Once there is general agreement or saturation of ideas [42] from the stakeholders, this information is used to draft the build scope of the CDS tool. The CDS scope includes an idea of the user interface (UI) content, workflow integration, and format (eg, interruptive alert versus passive, mobile app) for interfacing with end users. The clinical informaticist drafts the build scope based on what is clinically relevant, the stated workflow needs and preferences of the stakeholders, organizational priorities, and patient-centered considerations. Aligned with the learning health care system and agile design principles, the first CDS build scope is not intended to be perfect, rather a first iteration [43]. However, every iteration must be vetted to *do no harm*.

Stakeholders suggest many ideas. If they are all followed, the CDS UI may not be intuitive because too much information is presented. It is the role of the clinical informaticist to identify the most salient stakeholder requests and determine the appropriate balance between user-centered design and other determinants of implementation success outlined by PRISM and CDS design best practices. Finding this balance is challenging and varies for each situation. [Textbox 4](#) provides an example of a situation in which stakeholder requests were balanced with CDS design best practices. CDS in health care is fraught with nuances, and design decisions require thoughtful consideration of multiple dynamic and interacting contextual factors. The clinical informaticist also liaises between the stakeholders and build analysts (or they may be the builder), advocating for the stakeholders as appropriate and creatively adapting to the constraints of the EHR technical infrastructure, standards, and local resources to ultimately fit within the clinical workflows. Standards can include norms around the appearance and presentation of CDS tools within the organization or external technical standards to promote interoperability, such as Health Level 7 and use of standardized vocabularies such as SNOMED (Systematized Nomenclature of Medicine) to classify diagnoses.

An important consideration is the most appropriate technical integration format for the CDS tool, which could include the use of native EHR software, web services, or Substitutable Medical Apps Reusable Technology on Fast Healthcare Interoperability Resources applications. Each institution's EHR

offers a unique set of options to integrate CDS tools natively or interface them with external software. A clinical informaticist will be abreast of what technical integration options each institution has available and guides the design accordingly.

**Textbox 4.** Phase 2 case study.

Case study:

When designing the beta-blocker CDS tool, some stakeholders requested that a complete list of patient medications be included in the UI. However, most patients with heart failure are on many medications. Compliance with that request would have significantly decreased intuitiveness of the CDS tool. Balancing this request with CDS design best practices, it was decided not to include a complete medication list within the UI. However, this decision was difficult, given that this was a consistent request from stakeholders and the information influenced their decision-making. Without inclusion of this information in the UI, clinicians who did not recall concurrent medications would have to leave the CDS tool and search elsewhere in the EHR for the information, which decreases the relative advantage of the CDS tool. Later iterations of the CDS tool will consider creative ways to integrate medications with the CDS tool (eg, info button functionality).

### Phase 3. Design and Usability Testing

Following PRISM, at each stage of stakeholder engagement, new insights are learned. As CDS development activities evolve to testing, the format and nature of stakeholder engagement changes, but patient safety issues are always considered. Aligned with PRISM, design and usability testing aim to ensure that the CDS tool is designed well and simple to use. Design testing includes both SMEs and end users, whereas SMEs may be conditionally included in usability testing if their inclusion optimizes general buy-in. Design testing does not necessarily require build completion. However, before usability testing begins, the build should be complete within EHR testing environments and thoroughly tested to ensure that there are no errors. The resolution of build errors maximizes the focus during usability testing on optimizing the end user experience. When possible, the use of >1 UI during design and usability testing can enrich preimplementation design activities [44].

Testing can be completed in a variety of ways. Exemplars of design and usability testing can be found outside of health care [45], but for a variety of reasons have proven difficult to apply in health care. Therefore, here we describe an approach that can be practically applied in health care. Testing in actual clinical scenarios is not always possible in health care; thus, we propose design testing in which static screenshots of the CDS UI are shared with stakeholders, and usability testing with simulated patient scenarios followed by open-ended discussion. Design testing serves to validate whether the stated needs and preferences of stakeholders are accurately represented in the

UI. During design testing, the UI is iteratively updated before commencement of the more resource-intensive usability testing.

During usability testing, proctored simulation coupled with the *think aloud* protocol [46] and open-ended discussion can serve to inform educational materials, identify usability issues before going live, and identify additional areas for improvement, such as unintended consequences. Using the *think aloud* protocol during usability testing simulations can help identify barriers to following the CDS tool. Such barriers can be elaborated on during the open-ended discussion and, when actionable, the CDS tool can be redesigned to address the barrier. Proctored, simulated patient scenarios may not always be feasible given time or other resource constraints. Less resource-intensive usability testing methods could consist of asking end users to test the CDS tool remotely in EHR testing environments at their convenience and reporting back electronically with any feedback.

During design and usability testing, it is important to get end users in the mindset of their clinical workflow and consider (1) all members of team-based care, (2) factors that would impede or aid their workflow (cause alert fatigue versus fit into workflow), (3) whether the CDS tool supports their ability to make an informed decision, (4) whether it makes it easy to take (or incentivize) action, and (5) potential barriers to following the CDS tool's recommendation, especially over time. The determinants of effective implementation and sustainability from PRISM (Textbox 2) and CDS design best practices (Textbox 1) should inform specific questions to be asked during testing. Textbox 5 describes our approach to design and usability testing.

**Textbox 5.** Phase 3 case study.

Case study:

For our beta-blocker CDS tool, design testing occurred via email. Stakeholders were asked for input with specific questions, which we found to be effective and efficient. During usability testing, we incorporated our educational handout as part of the simulations, which resulted in substantial revisions to the handout as a result of end user feedback. Furthermore, during usability testing, we discovered that many clinicians were unaware that respiratory disease was not a contraindication to beta blockers; thus, a statement to address this misconception was added to the UI. Multimedia Appendix 1 provides examples of feedback from stakeholders during design and usability testing and reasons for or not incorporating into the CDS design.

#### Phase 4. Thoughtful Deployment

The implementation and sustainability infrastructure domain of PRISM considers practical measures to facilitate the ease of workflow integration. Adoption is improved when the necessary support to use the intervention is provided, and there is a means to quickly resolve any unintended consequences. Thoughtful introduction into clinical workflows is imperative for the adoption of CDS tools. Deployment should consider how to move the CDS tool into actual clinical workflows and how to communicate the change with end users. Re-engagement with clinical leadership before deployment is also key, especially in situations in which the preimplementation process occurred over a long period of time or for health systems undergoing certain changes.

The decision to implement the CDS tool in a pilot cohort or begin with widespread deployment is an important consideration to resolve any remaining usability or technical errors and to facilitate buy-in. In some instances, it is prudent to begin with a pilot group of users, which may include one clinical department or group of users across departments. The decision to deploy the CDS tool on a large or small scale initially should be informed by the anticipated frequency of exposure, acuity of the clinical situation, and workflow disruption of the CDS tool. In the case of a CDS tool with infrequent end user exposure, widespread dissemination can serve as a natural pilot

when appropriately monitored. Implementing CDS in a pilot fashion can help to bolster buy-in and to discover unintended issues before widespread roll-out.

With every CDS go-live, some communication with end users is needed but the extent should vary based on the complexity of the CDS tool and the anticipated frequency of end user exposure. When exposure to a CDS tool is infrequent or highly intuitive, end user education may not be necessary. However, an interruptive CDS tool that recommends discontinuation of nonsteroidal anti-inflammatory drugs in patients with cardiovascular disease would likely require some education of end users because of its alert frequency and obtrusiveness. In the latter example, end user education may provide a more detailed explanation of why this new tool is being implemented and what the response options provided within the UI mean. For example, the response options may include *never appropriate* without space to explain within the CDS UI that selecting this option will suppress the CDS tool forever for all clinicians. It would be helpful to share such information in the educational material. Many clinicians wish to know the consequences of their actions in response to a CDS tool. Communication plans should be informed by the given health system's standard processes and adapted based on the type of CDS and anticipated impact on the end user. [Textbox 6](#) describes key experiences from our deployment.

#### Textbox 6. Phase 4 case study.

##### Case study:

- For the beta-blocker CDS tool, re-engagement with the leadership of the clinical practices before deployment was pivotal. During our preimplementation planning process, the health system acquired several outpatient practice groups, which resulted in changes in priorities and approaches to decision support. At the beginning of the preimplementation planning stage, we secured leadership approval to deploy the CDS tool across all practices; however, given the changes that occurred, this approval was no longer applicable. Therefore, we needed to solicit approval for deployment from individual practices. Our experience emphasizes the need to maintain frequent engagement with leadership and a nonlinear approach to CDS implementation.
- When soliciting approval from individual practices, we piloted the CDS tool in 2 of the largest practices. These 2 pilot practices expressed early support for the CDS tool. When soliciting approval from other practices, they found it reassuring to know the tool was already accepted by their peers and being tested.

#### Phase 5. Performance Evaluation and Maintenance

PRISM includes the RE-AIM evaluation framework and outcomes. RE-AIM captures a broad and balanced evaluation of the nuances and pragmatic nature of implementation in clinical workflows [31]. Such a multilevel IS framework assesses the representativeness of participants, the extent to which the intervention needs to be adapted, and its sustainability [27-30]. Adaptions to interventions should be anticipated and evaluated. RE-AIM also encourages continuous evaluation and

dissemination of findings to promote observability and thereby optimize implementation and sustainability. [Table 2](#) provides an example of how RE-AIM can be applied to CDS. For example, changes in clinical outcomes are difficult to associate with CDS tools; thus, effectiveness is often measured as a change in behavior. Other instruments and tools can also be used to evaluate CDS tools. The usability of CDS tools can be evaluated using the validated System Usability Scale (SUS) [47].



**Table 2.** Reach, Effectiveness, Adoption, Implementation, and Maintenance evaluation framework applied to clinical decision support.

RE-AIM <sup>a</sup> domain	Described	Potential CDS <sup>b</sup> outcome measures
Reach (individual level)	Proportion and representativeness of those impacted by the intervention (and reasons for these results)	<ul style="list-style-type: none"> <li>• Number of patients the CDS tool fired for divided by the number of patients the CDS tool should have fired for</li> <li>• Characteristics of each group in numerator and denominator</li> <li>• Investigation of reasons not fired</li> </ul>
Effectiveness (individual level)	Impact of the intervention, including heterogeneity across subgroups and any negative outcomes (and reasons for these results)	<ul style="list-style-type: none"> <li>• Number of patients the CDS tool changed care for divided by the number of patients the CDS tool fired for</li> <li>• Characteristics of each group in numerator and denominator</li> <li>• Reasons care did or did not change</li> <li>• Number and type of unintended or negative outcomes</li> </ul>
Adoption (setting and staff at multiple levels)	<ul style="list-style-type: none"> <li>• Proportion and representativeness of those accepting or using the intervention</li> <li>• At levels of health systems, departments, and individuals (and reasons for these results)</li> </ul>	<ul style="list-style-type: none"> <li>• Number of clinicians who responded<sup>c</sup> to the CDS tool (did not outright dismiss) divided by the number of clinicians the CDS tool fired for</li> <li>• Number of patients who the CDS fired for that were not outright dismissed divided by the number of patients the CDS tool fired for</li> <li>• Number of practices, setting or clinicians participating divided by the number invited</li> <li>• Characteristics of each group in the numerators and denominators above</li> <li>• Reasons for or not to participate or dismiss</li> </ul>
Implementation (setting and staff at multiple levels)	<ul style="list-style-type: none"> <li>• Fidelity of the intervention and implementation strategy</li> <li>• Adaptations</li> <li>• Burden of delivery, including costs</li> </ul>	<ul style="list-style-type: none"> <li>• Adaptation: number and type of changes to the CDS build or workflow integration after deployment</li> <li>• Usability of the CDS tool (eg, SUS<sup>d</sup>)</li> <li>• Interviews on experience and adaptations</li> <li>• Cost of implementing (eg, time, resources)</li> </ul>
Maintenance (individual level and setting and staff at multiple levels)	Long-term effects of the intervention and extent the intervention becomes a routine part of care	<ul style="list-style-type: none"> <li>• Long-term outcomes (eg, change in mortality)</li> <li>• Sustained workflow integration and effectiveness</li> <li>• Interviews on intended or actual sustainment or further modification</li> </ul>

<sup>a</sup>RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance framework.

<sup>b</sup>CDS: clinical decision support.

<sup>c</sup>Technically in PRISM or RE-AIM, adoption is defined as only initial agreement to participate in (or be trained in) a program. In this paper, it will be defined as above to be consistent with how this term is used in informatics and to reflect the fact that end users do not always have the choice to interface with a CDS tool.

<sup>d</sup>SUS: System Usability Scale.

CDS tools are critical components of the learning health care system and should be regularly evaluated for performance and safety. Performance and safety evaluations are necessary for CDS maintenance and assist in minimizing alert fatigue. Evaluations should be scheduled with defined procedures of who is responsible, what the evaluation entails, and when it occurs. What is evaluated should be informed by operational leadership and influenced by external drivers, such as regulatory requirements and pay for performance metrics. The evaluation can also be informed by input from stakeholders regarding unintended consequences. For example, stakeholders might express concern that a CDS tool may lead to an increased risk

of bradycardia; thus, this is an outcome that should be monitored.

Evaluations should not be limited to 1 instance and should lead to action when appropriate. CDS implementation should be an iterative, thoughtful, and continuous improvement process. Proactively seeking end user input during and post exposure also demonstrates commitment to improvement. Findings from evaluations should be disseminated to stakeholders, including appropriate levels of leadership, SMEs, and end users. Appropriate levels of leadership should vary based on the specific CDS tool and do not necessarily require notification at the chief executive level. Communication with individual SMEs



and end users may be at the discretion of and via their direct leaders but should be offered at a minimum. This level of transparent communication can promote trust, improve adoption,

and reduce the culture of negativity that currently exists around CDS and EHRs [48]. [Textbox 7](#) describes key findings from our evaluation approach.

#### Textbox 7. Phase 5 case study.

##### Case study:

- We monitored clinician responses and feedback to the beta-blocker CDS tool weekly to screen for unintended consequences and errors. After 1 month of deployment, a clinician alerted us to an error in the build of our CDS tool, which we were able to resolve quickly. Had our CDS tool fired more frequently, we would likely have identified this error during our 2-week pilot. Such errors are common when a CDS tool is deployed, reinforcing the importance of ongoing monitoring and clear lines of communication between the informatics team and the end user.
- We also conducted brief structured interviews with clinicians exposed to the CDS tool. Together with completion of the SUS survey, the interviews provided valuable open-ended feedback that was used to refine both the CDS tool and our PRISM/CDS best practices approach. Key discoveries made during the interviews were the importance of carefully considering whether to include a dismiss button and avoiding a sense of shaming or disrespect. These discoveries were explicitly added as determinants of implementation success within CDS design best practices in [Textbox 1](#).

## Discussion

### Principal Findings

The PRISM/CDS best practices approach accounts for the multilevel interactions and dynamic factors that influence CDS implementation in health care. IS frameworks make social science more replicable and, by adding the context of CDS to PRISM, the reproducibility and sustainability of CDS implementations should be enhanced. *Integrating PRISM with CDS design best practices synthesizes the many known contextual factors that can influence the success of CDS implementation*, thereby elevating the experience from implementation to IS. This approach can be adapted for other health systems and CDS tools and used to guide resource allocation in a manner that optimizes CDS implementation success.

However, an effective approach may not be optimally efficient. When applying PRISM to CDS, resources include the time and availability of skilled personnel. Especially during the preimplementation period, this approach can require resources that may not be available at every health system or be appropriate allocation to every CDS instance. It may not be appropriate to allocate extensive resources to implementation efforts for CDS tools that are minimally invasive (eg, infobuttons) or address infrequent care gaps or low severity clinical situations. When resources are limited or the situation does not justify the allocation of full resources, stakeholder engagement may be abbreviated. Abbreviated stakeholder engagement may be limited to a smaller sample, representing a few representatives from each stakeholder group. Irrespective of resource availability, stakeholders should include representation from SMEs and end users. If allowed by resources, greater representation reaching saturation and general agreement from SMEs and end users is ideal.

The results of an RCT demonstrating the positive effect of the PRISM/CDS best practice approach on prescribing for heart failure in primary care is currently under review for publication. However, there are several limitations of the PRISM/CDS best practice approach. Although our study team represents diverse expertise, including informatics and IS, the approach was created based on our knowledge and experiences. Inclusion of additional

expertise and experiences across other clinical contexts and situations would likely refine the PRISM/CDS best practices approach. Our approach also relies on one of many IS frameworks and existing CDS design best practices. Other frameworks may have different advantages or disadvantages when applied to CDS. As CDS design best practices and technical capabilities evolve, the PRISM/CDS best practices approach will need to adapt. Further, the PRISM/CDS best practices need to be adapted based on a given institution's available resources and skilled personnel. Despite these limitations, the PRISM/CDS best practices approach provides a basis for advancing the science of CDS implementation.

Future research is needed to apply PRISM to additional real-world CDS implementations to capture all contextual factors and understand its impact on CDS adoption and effectiveness. Although the PRISM/CDS best practices approach was created for adaptation across any health care setting, it was refined within primary care settings across a large health care system. Future research should explore the application of PRISM to diverse CDS formats (eg, mobile apps) in a variety of patient care situations and practice settings of different sizes and identify means to refine the model to maximize effectiveness and efficiency. Use of the PRISM/CDS best practices approach should be documented in terms of its costs and benefits relative to other approaches. A key issue for both PRISM and CDS design best practices is the degree of iteration and number of cycles required for a given implementation. Efforts to improve the efficiency of the PRISM/CDS best practices approach are needed and may include integration with principles of rapid prototyping or tailoring the approach based on the severity of the clinical situation or the anticipated reach or workflow interruption of a given CDS tool. This tradeoff between resource allocation and benefits will have to be evaluated for every setting and clinical situation. Further, given that many of the contextual issues related to CDS are applicable to other types of health IT solutions, the applicability of this integrated approach to health IT beyond CDS should also be explored. Although the PRISM/CDS best practices approach provides guidance on evaluation, future research should consider how the approach can be expanded to include a systematic and standardized knowledge management process to evaluate and update CDS tools.

## Conclusions

We described an approach for applying PRISM to design, implement, and evaluate CDS tools that are integrated with CDS design best practices. Others are encouraged to adapt this approach to their situation to maximize CDS implementation success. This approach considers the many dynamic and interacting contextual factors that influence CDS implementation

success and sustainability in health care and suggests specific methods for designing and implementing CDS. Informed by an evidence-based IS framework, such an approach is foundational to maximizing the success of CDS implementation and the necessary platform from which cutting-edge innovations in CDS can be created to significantly improve and sustain health care outcomes.

## Acknowledgments

This work was supported in part by the National Institutes of Health or The National Center for Advancing Translational Sciences Colorado CTSA (Clinical and Translational Science Award) Grant Number UL1 TR002535 and by the NHLBI K12 Training Grant Number K12HL137862. Contents are the authors' sole responsibility and do not necessarily represent the official views of the National Institutes of Health.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Summary of constructive feedback received during design and usability testing.

[DOCX File, 22 KB - [jmir\\_v22i10e19676\\_app1.docx](#)]

## References

1. Mungall DR, Anbe D, Forrester PL, Luoma T, Genovese R, Mahan J, et al. A prospective randomized comparison of the accuracy of computer-assisted versus GUSTO nomogram--directed heparin therapy. *Clin Pharmacol Ther* 1994 May;55(5):591-596. [doi: [10.1038/clpt.1994.73](#)] [Medline: [8181203](#)]
2. White RH, Hong R, Venook AP, Daschbach MM, Murray W, Mungall DR, et al. Initiation of warfarin therapy: comparison of physician dosing with computer-assisted dosing. *J Gen Intern Med* 1987;2(3):141-148. [doi: [10.1007/bf02596140](#)] [Medline: [3295148](#)]
3. Chertow GM, Lee J, Kuperman GJ, Burdick E, Horsky J, Seger DL, et al. Guided medication dosing for inpatients with renal insufficiency. *J Am Med Assoc* 2001 Dec 12;286(22):2839-2844. [doi: [10.1001/jama.286.22.2839](#)] [Medline: [11735759](#)]
4. Wolfstadt JI, Gurwitz JH, Field TS, Lee M, Kalkar S, Wu W, et al. The effect of computerized physician order entry with clinical decision support on the rates of adverse drug events: a systematic review. *J Gen Intern Med* 2008 Apr;23(4):451-458 [FREE Full text] [doi: [10.1007/s11606-008-0504-5](#)] [Medline: [18373144](#)]
5. Bryant AD, Fletcher GS, Payne TH. Drug interaction alert override rates in the meaningful use era: no evidence of progress. *Appl Clin Inform* 2014;5(3):802-813 [FREE Full text] [doi: [10.4338/ACI-2013-12-RA-0103](#)] [Medline: [25298818](#)]
6. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *J Am Med Assoc* 1998 Oct 21;280(15):1339-1346. [doi: [10.1001/jama.280.15.1339](#)] [Medline: [9794315](#)]
7. Hemens BJ, Holbrook A, Tonkin M, Mackay JA, Weise-Kelly L, Navarro T, CCDSS Systematic Review Team. Computerized clinical decision support systems for drug prescribing and management: a decision-maker-researcher partnership systematic review. *Implement Sci* 2011 Aug 3;6:89. [doi: [10.1186/1748-5908-6-89](#)] [Medline: [21824383](#)]
8. Roshanov PS, Misra S, Gerstein HC, Garg AX, Sebaldt RJ, Mackay JA, CCDSS Systematic Review Team. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review. *Implement Sci* 2011 Aug 3;6:92 [FREE Full text] [doi: [10.1186/1748-5908-6-92](#)] [Medline: [21824386](#)]
9. Jaspers MW, Smeulders M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc* 2011 May 1;18(3):327-334 [FREE Full text] [doi: [10.1136/amiajnl-2011-000094](#)] [Medline: [21422100](#)]
10. Kashfi H. Applying a user centered design methodology in a clinical context. *Stud Health Technol Inform* 2010;160(Pt 2):927-931. [Medline: [20841820](#)]
11. McCullagh LJ, Sofianou A, Kannry J, Mann DM, McGinn TG. User centered clinical decision support tools: adoption across clinician training level. *Appl Clin Inform* 2014;5(4):1015-1025 [FREE Full text] [doi: [10.4338/ACI-2014-05-RA-0048](#)] [Medline: [25589914](#)]
12. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc* 2003;10(6):523-530 [FREE Full text] [doi: [10.1197/jamia.M1370](#)] [Medline: [12925543](#)]

13. Osheroff J, Teich J, Levick D, Saldana L, Velasco F, Sittig D. Improving outcomes with clinical decision support: An implementers Guide. 2nd Edition. Chicago, IL: Healthcare Information Management Systems Society (HIMSS); 2012.
14. Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. *J Biomed Inform* 2012 Dec;45(6):1202-1216 [FREE Full text] [doi: [10.1016/j.jbi.2012.09.002](https://doi.org/10.1016/j.jbi.2012.09.002)] [Medline: [22995208](https://pubmed.ncbi.nlm.nih.gov/22995208/)]
15. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. *J Biomed Inform* 2008 Apr;41(2):387-392 [FREE Full text] [doi: [10.1016/j.jbi.2007.09.003](https://doi.org/10.1016/j.jbi.2007.09.003)] [Medline: [18029232](https://pubmed.ncbi.nlm.nih.gov/18029232/)]
16. Marcilly R, Ammenwerth E, Roehrer E, Niès J, Beuscart-Zéphir MC. Evidence-based usability design principles for medication alerting systems. *BMC Med Inform Decis Mak* 2018 Jul 24;18(1):69 [FREE Full text] [doi: [10.1186/s12911-018-0615-9](https://doi.org/10.1186/s12911-018-0615-9)] [Medline: [30041647](https://pubmed.ncbi.nlm.nih.gov/30041647/)]
17. Van de Velde S, Kunnamo I, Roshanov P, Kortteisto T, Aertgeerts B, Vandvik PO, GUIDES expert panel. The GUIDES checklist: development of a tool to improve the successful use of guideline-based computerised clinical decision support. *Implement Sci* 2018 Jun 25;13(1):86 [FREE Full text] [doi: [10.1186/s13012-018-0772-3](https://doi.org/10.1186/s13012-018-0772-3)] [Medline: [29941007](https://pubmed.ncbi.nlm.nih.gov/29941007/)]
18. Kannry J, McCullagh L, Kushniruk A, Mann D, Edonyabo D, McGinn T. A framework for usable and effective clinical decision support: experience from the iCPR randomized clinical trial. *EGEMS (Wash DC)* 2015;3(2):1150 [FREE Full text] [doi: [10.13063/2327-9214.1150](https://doi.org/10.13063/2327-9214.1150)] [Medline: [26290888](https://pubmed.ncbi.nlm.nih.gov/26290888/)]
19. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc* 2003;10(6):523-530 [FREE Full text] [doi: [10.1197/jamia.M1370](https://doi.org/10.1197/jamia.M1370)] [Medline: [12925543](https://pubmed.ncbi.nlm.nih.gov/12925543/)]
20. van de Velde S, Heselmans A, Delvaux N, Brandt L, Marco-Ruiz L, Spitaels D, et al. A systematic review of trials evaluating success factors of interventions with computerised clinical decision support. *Implement Sci* 2018 Aug 20;13(1):114 [FREE Full text] [doi: [10.1186/s13012-018-0790-1](https://doi.org/10.1186/s13012-018-0790-1)] [Medline: [30126421](https://pubmed.ncbi.nlm.nih.gov/30126421/)]
21. Schoville RR, Titler MG. Guiding healthcare technology implementation: a new integrated technology implementation model. *Comput Inform Nurs* 2015 Mar;33(3):99-107; quiz E1. [doi: [10.1097/CIN.0000000000000130](https://doi.org/10.1097/CIN.0000000000000130)] [Medline: [25799235](https://pubmed.ncbi.nlm.nih.gov/25799235/)]
22. Hu PJ, Chau PY, Sheng OR, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. *J Manag Inf Syst* 2015 Dec 2;16(2):91-112. [doi: [10.1080/07421222.1999.11518247](https://doi.org/10.1080/07421222.1999.11518247)]
23. Venkatesh, Morris, Davis, Davis. User acceptance of information technology: toward a unified view. *MIS Q* 2003;27(3):425. [doi: [10.2307/30036540](https://doi.org/10.2307/30036540)]
24. Holden RJ, Asan O, Wozniak EM, Flynn KE, Scanlon MC. Nurses' perceptions, acceptance, and use of a novel in-room pediatric ICU technology: testing an expanded technology acceptance model. *BMC Med Inform Decis Mak* 2016 Nov 15;16(1):145 [FREE Full text] [doi: [10.1186/s12911-016-0388-y](https://doi.org/10.1186/s12911-016-0388-y)] [Medline: [27846827](https://pubmed.ncbi.nlm.nih.gov/27846827/)]
25. Greenes RA, Bates DW, Kawamoto K, Middleton B, Osheroff J, Shahar Y. Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. *J Biomed Inform* 2018 Feb;78:134-143 [FREE Full text] [doi: [10.1016/j.jbi.2017.12.005](https://doi.org/10.1016/j.jbi.2017.12.005)] [Medline: [29246790](https://pubmed.ncbi.nlm.nih.gov/29246790/)]
26. Moullin JC, Sabater-Hernández D, Fernandez-Llamos F, Benrimoj SI. A systematic review of implementation frameworks of innovations in healthcare and resulting generic implementation framework. *Health Res Policy Syst* 2015 Mar 14;13:16 [FREE Full text] [doi: [10.1186/s12961-015-0005-z](https://doi.org/10.1186/s12961-015-0005-z)] [Medline: [25885055](https://pubmed.ncbi.nlm.nih.gov/25885055/)]
27. Feldstein AC, Glasgow RE. A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. *Jt Comm J Qual Patient Saf* 2008 Apr;34(4):228-243. [doi: [10.1016/s1553-7250\(08\)34030-6](https://doi.org/10.1016/s1553-7250(08)34030-6)] [Medline: [18468362](https://pubmed.ncbi.nlm.nih.gov/18468362/)]
28. Liles EG, Schneider JL, Feldstein AC, Mosen DM, Perrin N, Rosales AG, et al. Implementation challenges and successes of a population-based colorectal cancer screening program: a qualitative study of stakeholder perspectives. *Implement Sci* 2015 Mar 29;10:41 [FREE Full text] [doi: [10.1186/s13012-015-0227-z](https://doi.org/10.1186/s13012-015-0227-z)] [Medline: [25890079](https://pubmed.ncbi.nlm.nih.gov/25890079/)]
29. McCreight MS, Rabin BA, Glasgow RE, Ayele RA, Leonard CA, Gilmartin HM, et al. Using the Practical, Robust Implementation and Sustainability Model (PRISM) to qualitatively assess multilevel contextual factors to help plan, implement, evaluate, and disseminate health services programs. *Transl Behav Med* 2019 Nov 25;9(6):1002-1011. [doi: [10.1093/tbm/ibz085](https://doi.org/10.1093/tbm/ibz085)] [Medline: [31170296](https://pubmed.ncbi.nlm.nih.gov/31170296/)]
30. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. Re-aim planning and evaluation framework: adapting to new science and practice with a 20-year review. *Front Public Health* 2019;7:64 [FREE Full text] [doi: [10.3389/fpubh.2019.00064](https://doi.org/10.3389/fpubh.2019.00064)] [Medline: [30984733](https://pubmed.ncbi.nlm.nih.gov/30984733/)]
31. Gaglio B, Shoup JA, Glasgow RE. The RE-AIM framework: a systematic review of use over time. *Am J Public Health* 2013 Jun;103(6):e38-e46. [doi: [10.2105/AJPH.2013.301299](https://doi.org/10.2105/AJPH.2013.301299)] [Medline: [23597377](https://pubmed.ncbi.nlm.nih.gov/23597377/)]
32. Richardson JE, Middleton B, Platt JE, Blumenfeld BH. Building and maintaining trust in clinical decision support: recommendations from the Patient-Centered CDS Learning Network. *Learn Health Syst* 2020 Apr;4(2):e10208 [FREE Full text] [doi: [10.1002/lrh2.10208](https://doi.org/10.1002/lrh2.10208)] [Medline: [32313835](https://pubmed.ncbi.nlm.nih.gov/32313835/)]
33. Kannry J, Sengstack P, Thyvalikakath TP, Poikonen J, Middleton B, Payne T, et al. The Chief Clinical Informatics Officer (CCIO): AMIA Task Force Report on CCIO Knowledge, Education, and Skillset Requirements. *Appl Clin Inform* 2016;7(1):143-176 [FREE Full text] [doi: [10.4338/ACI-2015-12-R-0174](https://doi.org/10.4338/ACI-2015-12-R-0174)] [Medline: [27081413](https://pubmed.ncbi.nlm.nih.gov/27081413/)]

34. Warm D, Thomas B. A review of the effectiveness of the clinical informaticist role. *Nurs Stand* 2011 Jul 6;25(44):35-38. [doi: [10.7748/ns2011.07.25.44.35.c8612](https://doi.org/10.7748/ns2011.07.25.44.35.c8612)] [Medline: [27101059](https://pubmed.ncbi.nlm.nih.gov/27101059/)]
35. Kukhareva P, Weir CR, Staes C, Borbolla D, Slager S, Kawamoto K. Integration of clinical decision support and electronic clinical quality measurement: domain expert insights and implications for future direction. *AMIA Annu Symp Proc* 2018;2018:700-709 [FREE Full text] [Medline: [30815112](https://pubmed.ncbi.nlm.nih.gov/30815112/)]
36. Hagland M. Show us the evidence. *Healthc Inform* 2010 Jan;27(1):34-38. [Medline: [20120893](https://pubmed.ncbi.nlm.nih.gov/20120893/)]
37. Sable JH, Hales JW, Bopp KD. Medical informatics in healthcare organizations: a survey of healthcare information managers. *Proc AMIA Symp* 2000:744-748 [FREE Full text] [Medline: [11153356](https://pubmed.ncbi.nlm.nih.gov/11153356/)]
38. Hoehn BJ. Clinical information technology governance. *J Healthc Inf Manag* 2010;24(2):13-14. [Medline: [20397326](https://pubmed.ncbi.nlm.nih.gov/20397326/)]
39. Kawamoto K, Flynn MC, Kukhareva P, ElHalta D, Hess R, Gregory T, et al. A pragmatic guide to establishing clinical decision support governance and addressing decision support fatigue: a case study. *AMIA Annu Symp Proc* 2018;2018:624-633 [FREE Full text] [Medline: [30815104](https://pubmed.ncbi.nlm.nih.gov/30815104/)]
40. Kitzinger J. Qualitative research. Introducing focus groups. *Br Med J* 1995 Jul 29;311(7000):299-302 [FREE Full text] [doi: [10.1136/bmj.311.7000.299](https://doi.org/10.1136/bmj.311.7000.299)] [Medline: [7633241](https://pubmed.ncbi.nlm.nih.gov/7633241/)]
41. Trinkley KE, Blakeslee WW, Matlock DD, Kao DP, Van Matre AG, Harrison R, et al. Clinician preferences for computerised clinical decision support for medications in primary care: a focus group study. *BMJ Health Care Inform* 2019 Apr;26(1). [doi: [10.1136/bmjhci-2019-000015](https://doi.org/10.1136/bmjhci-2019-000015)] [Medline: [31039120](https://pubmed.ncbi.nlm.nih.gov/31039120/)]
42. Krueger R, Casey M. *Focus Groups: a Practical Guide for Applied Research*. 5th Edition. Los Angeles, UK: SAGE Publications Inc; 2015.
43. Ray JM, Ratwani RM, Sinsky CA, Frankel RM, Friedberg MW, Powsner SM, et al. Six habits of highly successful health information technology: powerful strategies for design and implementation. *J Am Med Inform Assoc* 2019 Oct 1;26(10):1109-1114. [doi: [10.1093/jamia/ocz098](https://doi.org/10.1093/jamia/ocz098)] [Medline: [31265064](https://pubmed.ncbi.nlm.nih.gov/31265064/)]
44. Bailey B. Consider as Many Design Alternatives as Possible: The Value of Parallel Design Internet. Usability. URL: <https://www.usability.gov/get-involved/blog/2006/02/parallel-design.html> [accessed 2019-11-24]
45. Olsen D. *The Lean Product Playbook: How to Innovate with Minimum Viable Products and Rapid Customer Feedback*. Hoboken, New Jersey: John Wiley & Sons; 2015.
46. Ericsson K, Simon H. *Protocol analysis Rev'd Edition: Verbal reports as data*. Cambridge, MA: Massachusetts Institute of Technology; 1993.
47. System Usability Scale (SUS). URL: <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> [accessed 2020-10-04]
48. Rogers E. *Diffusion of Innovations*, 5th Edition. New York, USA: Free Press; 2003.

## Abbreviations

**CDS:** clinical decision support

**CTSA:** Clinical and Translational Science Award

**EHR:** electronic health record

**Health IT:** health information technology

**IS:** implementation science

**PRISM:** Practical Robust Implementation and Sustainability Model

**RE-AIM:** Reach, Effectiveness, Adoption, Implementation, and Maintenance

**SME:** subject matter expert

**SUS:** System Usability Scale

**UI:** user interface

*Edited by G Eysenbach; submitted 27.04.20; peer-reviewed by M Dorsch, B Middleton; comments to author 12.06.20; revised version received 18.07.20; accepted 15.09.20; published 29.10.20.*

### *Please cite as:*

Trinkley KE, Kahn MG, Bennett TD, Glasgow RE, Haugen H, Kao DP, Kroehl ME, Lin CT, Malone DC, Matlock DD  
*Integrating the Practical Robust Implementation and Sustainability Model With Best Practices in Clinical Decision Support Design: Implementation Science Approach*  
*J Med Internet Res* 2020;22(10):e19676  
URL: <http://www.jmir.org/2020/10/e19676/>  
doi: [10.2196/19676](https://doi.org/10.2196/19676)  
PMID: [33118943](https://pubmed.ncbi.nlm.nih.gov/33118943/)

©Katy E Trinkley, Michael G Kahn, Tellen D Bennett, Russell E Glasgow, Heather Haugen, David P Kao, Miranda E Kroehl, Chen-Tan Lin, Daniel C Malone, Daniel D Matlock. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Research Participants' Perspectives on Using an Electronic Portal for Engagement and Data Collection: Focus Group Results From a Large Epidemiologic Cohort

Erika Rees-Punia<sup>1</sup>, PhD, MPH; Alpa V Patel<sup>1</sup>, PhD; Asher Beckwitt<sup>2</sup>, PhD; Corinne R Leach<sup>1</sup>, PhD, MPH, MS; Susan M Gapstur<sup>1</sup>, PhD; Tenbroeck G Smith<sup>1</sup>, MS

<sup>1</sup>American Cancer Society, Atlanta, GA, United States

<sup>2</sup>Asher Consulting, Atlanta, GA, United States

**Corresponding Author:**

Erika Rees-Punia, PhD, MPH

American Cancer Society

250 Williams St

Atlanta, GA, 30303

United States

Phone: 1 4049823684

Email: [erika.rees-punia@cancer.org](mailto:erika.rees-punia@cancer.org)

## Abstract

**Background:** Epidemiologic cohort studies have begun to leverage electronic research participant portals to facilitate data collection, integrate wearable technologies, lower costs, and engage participants. However, little is known about the acceptability of portal use by research participants.

**Objective:** The aim of this study is to conduct focus groups among a sample of Cancer Prevention Study-3 (CPS-3) participants to better understand their preferences and concerns about research portals.

**Methods:** CPS-3 participants were stratified based on sex, race and ethnicity, age, and cancer status, and randomly invited to participate. Focus groups used an exploratory case design with semistructured guides to prompt discussion. Using a constant comparison technique, transcripts were assigned codes to identify themes.

**Results:** Participants (31/59, 52% women; 52/59, 88% White/non-Latinx) were favorably disposed toward using a research participant portal to take surveys, communicate with the study staff, and upload data. Most participants indicated that a portal would be beneficial and convenient but expressed concerns over data safety. Participants stressed the importance of an easy-to-use and trustworthy portal that is compatible with mobile devices.

**Conclusions:** In addition to being beneficial to researchers, portals may also benefit participants as long as the portals are secure and simple. Participants believe that portals can provide convenient ways to report data and remain connected to the study.

(*J Med Internet Res* 2020;22(10):e18556) doi:[10.2196/18556](https://doi.org/10.2196/18556)

**KEYWORDS**

focus groups; health information technology; epidemiologic studies

## Introduction

**Background**

A common, contemporary application of health information technology is the electronic patient portal [1]. Electronic patient portals are secure web-based systems that provide patients 24-hour access to their personal health information, enable two-way communication with health care providers, and potentially decrease costs through patient self-service [2,3]. Currently, less than half of US adults use patient portals

annually, although use is more common among younger (younger than 65 years), more educated people with broadband internet access [4,5]. Patient portals also have the potential to facilitate the integration of new technologies for remote monitoring, such as home blood pressure monitors, ecological momentary assessment apps, and other wearable devices (eg, activity trackers and heart rate monitors) at a lower cost [6]. Collectively, patient portals can simplify the collection of patient-generated health data (PGHD), including vital signs (eg, temperature, blood pressure, and blood glucose), lifestyle data

(eg, caloric intake, diet, exercise, and medication adherence), and patient-reported outcomes data (eg, mood, sleep quality, and pain) [7].

The use of electronic health portals has spread beyond the clinical setting. The adoption of commercial products such as Apple HealthKit [8], 23andme [9], and PatientsLikeMe [10] illustrates that an increasingly large number of people are willing to exchange medical and health behavior information with professionals other than their health care providers. Electronic portals have also been leveraged to survey and engage participants in research studies. Similar to the various benefits of patient portals, research participant portals could facilitate the collection of PGHD and data from wearable technologies, provide a way to administer targeted surveys to different subsets of the study population, collect electronic health data, and save costs and time associated with printing and mailing logistics [11]. Portals may be useful for frequently engaging participants in long-term studies, potentially reducing attrition. In addition to the aforementioned researcher benefits, participants may find portals convenient for storage and anytime access to study materials such as copies of consent forms, surveys, and study newsletters or updates.

Despite the potential benefits, very little is known about the feasibility and acceptability of electronic portal use by research participants in longitudinal cohort studies. Participant portals have been adopted by a few prospective cohorts thus far, including the University of California-San Francisco's PRIDE study [12] and the National Institutes of Health All of Us study [13]; however, none of them have published details regarding their portal development and testing. A better understanding of research participants' preferences and concerns about research portals could inform portal design and maximize the benefits to and engagement of participants and researchers.

## Objectives

In this study, focus groups were implemented among a sample of Cancer Prevention Study-3 (CPS-3) participants to facilitate the open sharing of insights largely unavailable through pre-existing surveys, especially (1) research participants' perceptions of the facilitators and barriers to utilizing an electronic research portal; (2) participants' views about using an interactive research portal to submit health records, complete lifestyle and medical surveys, and communicate with study staff; and (3) participants' recommendations for electronic research portals.

## Methods

### Sample Recruitment

CPS-3 is a contemporary prospective epidemiologic cohort study initiated by the American Cancer Society (ACS) [14]. In total, 303,682 participants aged 30-65 years across the United States, including Puerto Rico, were enrolled in CPS-3 between 2006 and 2013. Participants are sent surveys approximately every 3 years to update exposure and medical information and will be followed to identify incident cancers and cause-specific mortality through linkages with state cancer registries and the

National Death Index, respectively. The Emory University Institutional Review Board approved all aspects of CPS-3.

In 2017, plans to develop a CPS-3 participant portal began as a way to increase participant engagement and facilitate survey administration. This effort launched with a series of virtual and on-site focus groups conducted to gain insight into participant perspectives and willingness to share data through a participant portal. Virtual focus groups, offered via conference calls, enabled the inclusion of a broader geographic distribution of CPS-3 participants. To be eligible for virtual focus groups, participants must have been involved with CPS-3 since enrollment (2006-2013) and must have (1) a valid email address and (2) completed the 2018 triennial CPS-3 survey online. On-site focus groups took place in Atlanta, Georgia, and Albany, New York, and allowed participants who had not previously engaged with CPS-3 online to be included. To be eligible for on-site focus groups, participants must have (1) completed the 2018 survey (online or paper) and (2) resided in one of the predefined zip codes near the on-site locations.

Eligible CPS-3 participants were stratified based on sex, race and ethnicity, age group, and cancer status and invited to participate in the portal feasibility focus groups. We used this purposive sampling strategy to ensure adequate representations in focus groups by race and ethnicity, sex, age, and cancer history. Overall, 608 participants were invited, 78 registered to participate, and 59 participated in the focus groups.

### Study Design

The focus groups used an exploratory case design, which is used when there are no clear or predictable outcomes [15,16]. To guide the focus group discussion, semistructured guides were developed before conducting focus groups and were used to prompt group discussions regarding perceived portal benefits, barriers, and preferences (Multimedia Appendix 1). All focus group participants were assigned a pseudonym for anonymity and were asked to introduce themselves using their pseudonym before sharing opinions to ease transcription.

### Qualitative Analysis

All focus groups were audio recorded and transcribed verbatim, excluding phrases like *um* and *uh*. A ground-up or grounded theory approach was used to analyze the data [17]. Open coding was initially performed to code each line of the text. During this process, in vivo codes were generated based on the participants' words. Following the open coding process, the data were coded using axial coding. The open-coded data were iteratively and systematically compared and then linked together using one or more codes. Finally, selective coding was used to identify the core categories that emerged from the data. Selective coding was completed by re-reading the data and coding data that related to these categories.

Coding consensus was used to validate the data. One researcher coded all data, and two additional researchers reviewed all codes and their assignment to the text; any differences in coding assignments were adjudicated until consensus was reached. NVivo 12 (QSR International) was used to store, code, query, and organize all data [18].

## Results

### Sample Characteristics

A total of 11 focus groups were conducted in November 2018, with 7 virtual focus groups (2-8 participants each) and 4 on-site (6-8 participants each). Of the 11 focus groups, 5 consisted entirely of participants sharing an underrepresented demographic

characteristic (1 focus group of all African Americans, 1 virtual and 1 on-site focus group of all older adults [60 years and older], and 1 virtual and 1 on-site focus group of all cancer survivors). The focus group participant (n=59) demographics are presented in [Table 1](#). About half of the participants were women (31/59, 52%) and most were non-Latinx White (52/59, 88%). The age range of participants was 41 to 72 years (mean 59, SD 6 years), with about half over the age of 60 years (32/59, 54%).

**Table 1.** Individual-level demographics (n=608 invited, n=59 participants).

Characteristic	Participants invited (n=608)	Participants registered (n=78)	Focus group participants (n=59)
Women, n (%)	314 (51.6)	41 (52)	31 (52)
<b>Race and ethnicity, n (%)</b>			
Non-Latinx White	458 (75.3)	65 (83)	52 (88)
Black	50 (8)	5 (6)	3 (5)
Latinx	50 (8)	6 (7)	2 (3)
Other	50 (8)	2 (2)	2 (3)
<b>Age group, n (%)</b>			
59 years and under	507 (83.4)	62 (80)	27 (46)
60 years and older	101 (16.6)	16 (21)	32 (54)
Cancer survivor, n (%)	100 (16.4)	11 (14)	8 (15)
<b>Focus group location, n (%)</b>			
Virtual	406 (67)	46 (59)	31 (52)
On-site	202 (33)	32 (41)	28 (48)

As an introduction to the topic of electronic portals, participants were asked about their current portal use. Most, although not all, of the participants were familiar with the concept of online portals, and many already use them with their health care providers and/or banks. On the other hand, there were 3 on-site and 2 online focus group participants who stated they “have no experience with online portals,” “don’t know” if they have access to one or have ever used one, or “have not set up” a portal. Most participants described the benefits and barriers to general portal participation, including improved communication with health care providers, appreciation for 24-hour access to health care records, and concerns about private health information remaining secure. These benefits and barriers were also consistent with the existing literature on patient portals [19-23].

### Benefits of Research Portals

The majority of participants indicated that a research participant portal would be beneficial. Participants cited convenience as the main benefit of portal use. For example, one participant described the convenience of the all-hours access:

*...being able to access 24/7 is very helpful. It's the convenience, and then also you get two-way feedback. It's more interactive.*

Other benefits ([Textbox 1](#), theme 1) include appreciation for having all their study material at one place:

*...I would say convenience, being able to access [the portal] on my phone...all my data is in one place. I don't have to keep a lot of paper that I'm worried about getting in the wrong hands.*

**Textbox 1.** Additional quotes by theme.**Theme 1: Perceived portal benefits**

## Subtheme: convenience

1. "Same thing with the [CPS-3 study] newsletters, you can go back and access, but a lot of us read them and then we possibly throw it out or it ends up in a pile somewhere. We don't remember where it is. I think that's going to be a great advantage of having the portal."
2. "I would like having it all in that one portal. It's there, it's handy. You can go back. If I travel, I have the app on my phone. I can pull it up, it's there"

**Theme 2: Portal use concerns**

## Subtheme: worried about data privacy

1. "I volunteered to do this study for discovery of treatment or diagnoses to better peoples' health. I think that should be limited to that only. ... it's not happening in health data yet, but the selling of stuff to other people. That kind of thing should not happen."
2. "...what nags me about the online portals is, how much information am I giving out? Who is actually seeing this? And how do I know it's secure, whether it's coming or going?"

## Subtheme: not concerned about data privacy

1. "I'm okay with my information being used for educational purposes and by colleges and by universities and students and professors for purposes of figuring out what's going on to heal cancer patients and stuff like that."

**Theme 3: Information accessed or shared through the portal**

## Subtheme: comfortable sharing information

1. "I don't mind sharing all that information with the American Cancer Society in relation to this study, and I wouldn't mind my doctor sharing it directly with them. I have no qualms with any of that."
2. "...the whole point like somebody said, we chose to be participants in this study to gain information, so I would think you would want, and need said information. I'm fine. If I wasn't fine, I wouldn't participate in the study in the first place, so I'm fine sharing that information with them. Again, presuming it's all confidential..."

## Subtheme: not comfortable sharing certain information

1. "I'm okay with sharing most information, especially regarding research. The one piece of information I probably would not be as comfortable with is anything that relates to sex."
2. "I think it's hard for people to share any kinds of mental health or certain diagnoses, and I think cancer, or any kind of chronic illness might be that. Because if you need to find health insurance, sometimes I think there is a feeling that it's going to be...people are going to know it, and you're not going to be able to get insurance, or life insurance. If somehow, they can access this information."

**Theme 4: Communication with the American Cancer Society staff**

## Subtheme: comfortable communicating

1. "The ease of both ways communication with using a portal what strikes me if it's on my screen, I get a message, or I'm asked a question, I'm likely to respond. If I get something in the mail or that's going to get lost in the shuffle."
2. "I'm going to look at what the letters are after the name. Does it say RN? Is the person who's answering my question properly trained to have answered my question? That would be my concern, not whether it's John Smith, RN, but whether it's RN or not. In terms of knowing who all the people are who are involved, when would we ever know that? That doesn't seem relevant to me, whereas the qualifications of the person to respond matter a lot."

**Theme 5: Cancer Prevention Study-3 (CPS-3) portal desires**

## Subtheme: CPS-3 study information

1. "...any successes where the researchers have used our data to help point them at some conclusions that are going to have an impact, certainly we've all gotten in to this so that we would have an impact and if we can be rewarded with milestones, it would be wonderful."
2. "I'd like to see how many people participate and how many of these poor people have gotten cancer in the meantime and how long it's been going on and all that."

## Subtheme: security

1. "That two-factor authentication gives me a much more better comfort level."

## Subtheme: portal-based surveys

.

1. "...if I know ahead of time that the survey's gonna take longer then I'm probably gonna be okay with it... but if I start a survey and I don't know how long it's gonna take and it's going on forever, I'll quit it."
2. "I would want to know... can I save it and stop and come back to it? And/or is there any information that I need other than just my memory to take this survey, or to think about before I actually start taking it?"

Subtheme: technical support

1. "I personally find online customer service chats to be way more effective and easier than phone calls. You have a transcript, too, afterwards when you're done."
2. "I don't really perceive needing any assistance, just the basic. Being able to call an 800 number or chat with somebody online if there was a difficulty."
3. "I must have 35 passwords. Between remembering capital, star, or whatever it is, space and all that stuff and if you're not exact then you don't get in. If you're only doing it once every three months, it's difficult. It just needs to be easily you can retrieve a password."

## Research Portal Use Concerns

Participants' concerns were focused largely on data privacy, including data storage and use (Textbox 1, theme 2). For example, many participants felt uninformed about how their data would be used and by whom:

*I would be concerned as to how long that data would be stored. I mean if the study's only going to be 20 years, what happens to the data at the end of that 20 years?*

Several participants also specifically mentioned concerns about their data being accessed by employers or insurance companies, impacting their careers and/or insurability:

*I think that in light of the pre-existing condition kind of controversy out there, I would have the same concerns that my insurance company would get that information and use that somehow against me.*

A few participants shared that they generally do not trust the internet, with one explaining:

*I don't know that much about the internet and the ethernet or whatever they call it, but who knows who could get a hold of... my medical records going across the airwaves, I don't know.*

Conversely, some participants did not feel worried about data privacy. One participant reported the following:

*I don't know that I have concerns. I was assuming the information would be used for research in order to help prevent cancer...I'm trusting.*

## Information Accessed or Shared Through the Research Portal

Many participants expressed willingness to share certain types of information through portals, such as CPS-3 survey responses, medical records, and/or data from electronic devices (eg, wearable physical activity monitors and blood pressure monitors; Textbox 1, theme 3). When prompted, a few participants said they were willing to "share whatever [researchers] want". Another participant stated:

*One of the reasons we signed up for this study was we wanted to give information so that it could be used for future generations... I have no problem sharing everything.*

On the other hand, several participants expressed reluctance to share information they perceived as more sensitive, such as details regarding their mental health or sexual habits (Textbox 1, theme 3, subtheme: not comfortable sharing certain information, quote 1). One participant shared:

*I think anything with mental health, there's still a stigma there. So, if I have [mental health] issues... I would not feel comfortable sharing.*

Participants were also hesitant about providing too many personal identifiers on a portal:

*The things I don't feel comfortable sharing, I guess on health-related sites, would be my social security number, my Medicare number, things that other people can use to really get access to information about me and use in other harmful ways.*

Several participants drew a clear distinction between patient and research participant portals. They expressed a willingness to share details on patient portals but set limits on the amount of information they would share on research participant portals. One participant described the conditions required for sharing medical information with researchers:

*I'm a little concerned about keeping a clear distinction between researchers and medical providers...I'm willing to provide some of my medical records to researchers as long as I'm clear as to what they're doing with that and how they're asking that I transfer any kind of medical information aside from just surveys.*

In contrast, other participants clearly understood the distinction between patient and research participant portals and felt equally willing to share their information through both:

*I think the portal for our personal physician is a different situation. There it helps the patient/physician interaction and helps save time for both parties. As far as a portal for this study or studies like this, I think it's very important to assist research. I'm a big fan of sharing information in an anonymous fashion to help the research parties involved.*

## Communications With Study Staff

When asked, most participants expressed a willingness to receive and respond to messages through the participant portal (Textbox



1, theme 4). Further, participants felt that two-way communication would improve study engagement, as exemplified in this participant's thoughts:

*I would like to feel more connected to the study... I don't even know how often we get the opportunity to participate in answering a survey. And sometimes I don't feel connected to it. I feel like, "Have I missed one because I missed that email?"...more communication with the ACS in regards to the study might help me feel more connected to it.*

These perspectives were echoed by other participants:

*I think the communication piece is really important and even as we take surveys, now I sometimes find certain questions not as clear as I would like them to be, it would give us that avenue to offer good feedback to the team working on it in the first place.*

However, as another participant noted, some understanding of the study staff credentials would ease communication:

*I want to know a little bit about who's responding, so if the responder has some certification or some specialty in whatever the information they're providing, that would be really helpful to know. I don't want to take information on face value unless I know that it's coming from someone who has a background, at least, and experience.*

### Features of a Research Portal

Participants expressed a desire for a "well-designed [portal] that is user-friendly" and "doesn't have a lot of bells and whistles." Participants made several specific recommendations for features and content they would like to see on a research participant portal. For example, several participants hoped to find information regarding participant demographics (Textbox 1, theme 5), study design, and research findings on the portal:

*I'd like to see some updates, some CPS-3, some just in general, regarding the work they're doing. Then, also, maybe a section for comments or questions.*

Participants also spent time discussing their wish for a safe and secure portal, with several participants advocating specifically for two-step verification:

*I like a two-step validation or verification process...while it is a little inconvenient because it requires extra steps is, if I'm going to log onto a portal, the portal will send a message to my cell phone with a code and then to complete the login, I have to use the code. I like that kind of two-step authorization.*

Most participants expressed openness to completing regular surveys through the portal. During these discussions, the topic of preferred survey length frequently arose, with several participants stating that they were willing to commit "as long as it takes" or however much "time [the surveys] require." However, other participants had specific suggestions for maximum survey length and administration frequency:

*I think 15 to 20 minutes for a survey is an appropriate amount of time for me and how often, two or three times a year I think would be okay.*

Other participants suggested providing clear indicators of survey length (Textbox 1) by adding a completion bar or page countdown to help them manage their time while taking more lengthy surveys:

*I'm more likely to finish the survey if during it, it tells you how far you are so that you know if you're getting close to the end or not.*

Finally, many participants discussed the importance of technical support (Textbox 1, theme 5). Participants expressed concerns over having too many passwords and reiterated the importance of some sort of technical support to quickly retrieve a new password:

*I would like just a quick password reset email if necessary. Something quick as opposed to 20 minutes later and perhaps an online video if things get very technical.*

## Discussion

### Principal Findings

In this qualitative study, we investigated epidemiologic cohort research participants' perspectives and expectations of an electronic portal for study engagement and data collection. Through 11 focus groups, we found that most participants were favorably disposed toward the concept of using a portal to participate in the cohort study as long as the portal is secure and easy to use. Many themes that arose in this study were consistent across groups. Importantly, we were able to recruit enough participants to gain an understanding of how research participants in longitudinal cohort studies may view the implementation of an electronic portal. Given that the final focus group did not produce any significant new information, we can assume that we achieved data saturation. This suggests that participants in CPS-3 and other longitudinal cohort studies might be willing to use a research portal that has the potential to benefit both researchers and participants.

Some of the themes that arose in this study regarding research participant portals were similar to those documented in the literature regarding health care patient portals. For example, the most commonly discussed perceived benefit in this study was convenience, especially regarding access to research records anywhere, at any time, with a variety of electronic devices, which also applies to health care portals [19,20]. Similarly, the most commonly cited concern was data privacy and protection, including hacking, data sharing, and password management [20,24]. Although participants were informed of data privacy and protection at the time of CPS-3 enrollment, continuing to educate and inform participants is clearly important, and a participant portal would make it easier to do so. Given the similarities in the themes presented here and in patient portal research, it is possible that some findings regarding the design and implementation of patient portals may be applied to research portals. This is potentially important as there is considerably

more research on the use of portals for health care purposes than on the use of portals for research purposes.

Several novel concepts relating specifically to research participant portals have emerged. Generally, participants were willing to upload medical records, connect wearable devices, and share other PGHD through an electronic portal outside of a health care setting. Although most participants were open to completing surveys on the portal, many wished to be informed of the survey duration before and/or during administration. Similarly, participants expressed their preference to save responses and complete surveys at a time that is convenient for them. Participants' willingness to provide survey data appears to be based primarily on the altruistic belief that the data could benefit others through research findings. Although altruistic motivation for research participation may not be a novel concept [25], it is useful to understand these motivations when considering how to communicate the importance of logging in and participating on the portal. Finally, most participants were willing to communicate with the study staff through the portal, although several participants mentioned the importance of understanding the staff member's credentials.

On the basis of the participant feedback provided through the focus groups, we recommend that research participant portals be easy to use and function on all platforms (eg, mobile phones, tablets, and all commonly used web browsers). Given participant concerns regarding data privacy and protection, it is essential to design a portal with robust security (eg, dual-factor authentication) and easy access to data privacy policies from the signed informed consent form. Research participant portals should also provide timely technical support. Surveys administered through a portal should include progress bars and options to save and return to the survey later. The findings also suggest that a portal would be more appealing to participants if it connects and engages them through information on new study findings and discoveries.

## Strengths and Limitations

This study has some limitations. First, the level of transferability (ie, the degree to which these findings can be applied to other groups or initiatives) of these results is unclear, particularly as the participants in this study have already demonstrated a long-term commitment to a longitudinal cohort study. Additionally, although several underrepresented groups were present in our focus groups, we did not have large enough subgroup sample sizes to consider differences by group. Similarly, only 2 participants were present for one of the online focus groups, although more had initially registered; thus, group discussion may have been affected for that particular group. Although it is beyond the scope of this study, more consideration around the real-world integration of electronic research portals is needed, especially as participants presumably begin to gain access to more electronic patient portals through their health care providers. Future research should identify ways to simplify the implementation and management of research portals for both participants and study staff. Despite these potential limitations, this study is, to our knowledge, the first to systematically investigate the views of research participants regarding online research participant portal use.

## Conclusions

Participant portals may be beneficial to researchers for facilitating new data collection, integrating wearable technologies, administering targeted surveys on special topics to subset populations, saving costs and time compared with mailed surveys, and engaging participants between survey cycles. For participants, portals have the potential to reduce burden by providing convenient ways to report data, review study materials, and remain connected to the study. The rich data from these focus groups provide a rationale for building a research participant portal and information to inform the design of the portal. To achieve these potential benefits, portals should be easy to use, secure, trustworthy, and compatible with a variety of devices.

---

## Acknowledgments

The authors express sincere appreciation to all CPS-3 participants and to each member of the study and biospecimen management group. The views expressed here are those of the authors and do not necessarily represent the ACS or the ACS Cancer Action Network. The ACS funds the creation, maintenance, and updating of the CPS-3.

---

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Semistructured focus group guide.

[DOCX File, 18 KB - [jmir\\_v22i10e18556\\_app1.docx](#)]

---

## References

1. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. *Health Aff (Millwood)* 2011 Mar;30(3):464-471. [doi: [10.1377/hlthaff.2011.0178](#)] [Medline: [21383365](#)]

2. Otte-Trojel T, de Bont A, van de Klundert J, Rundall TG. Characteristics of patient portals developed in the context of health information exchanges: early policy effects of incentives in the meaningful use program in the United States. *J Med Internet Res* 2014 Nov 21;16(11):e258 [FREE Full text] [doi: [10.2196/jmir.3698](https://doi.org/10.2196/jmir.3698)] [Medline: [25447837](https://pubmed.ncbi.nlm.nih.gov/25447837/)]
3. What is a Patient Portal? Office of the National Coordinator for Health Information. URL: <https://www.healthit.gov/faq/What-Patient-Portal> [accessed 2019-10-10]
4. Sarkar U, Bates DW. Care partners and online patient portals. *J Am Med Assoc* 2014;311(4):357-358. [doi: [10.1001/jama.2013.285825](https://doi.org/10.1001/jama.2013.285825)] [Medline: [24394945](https://pubmed.ncbi.nlm.nih.gov/24394945/)]
5. Perzynski A, Roach M, Shick S, Callahan B, Gunzler D, Cebul R, et al. Patient portals and broadband internet inequality. *J Am Med Inform Assoc* 2017 Sep 1;24(5):927-932. [doi: [10.1093/jamia/ocx020](https://doi.org/10.1093/jamia/ocx020)] [Medline: [28371853](https://pubmed.ncbi.nlm.nih.gov/28371853/)]
6. Cannon C. Telehealth, mobile applications, and wearable devices are expanding cancer care beyond walls. *Semin Oncol Nurs* 2018 May;34(2):118-125. [doi: [10.1016/j.soncn.2018.03.002](https://doi.org/10.1016/j.soncn.2018.03.002)] [Medline: [29627143](https://pubmed.ncbi.nlm.nih.gov/29627143/)]
7. Shapiro M, Johnson D. Patient-Generated Health Data: White Paper. Office of the National Coordinator for Health Information. 2012. URL: [https://www.healthit.gov/sites/default/files/rti\\_pghd\\_whitepaper\\_april\\_2012.pdf](https://www.healthit.gov/sites/default/files/rti_pghd_whitepaper_april_2012.pdf) [accessed 2020-09-25]
8. Apple Developer. Apple HealthKit. URL: <https://developer.apple.com/healthkit/> [accessed 2019-10-10]
9. 23andMe. URL: <https://www.23andme.com/?myg06=true> [accessed 2019-10-10]
10. PatientsLikeMe. URL: <https://www.patientslikeme.com/> [accessed 2019-10-10]
11. Kelly-Pumarol IJ, Henderson PQ, Rushing JT, Andrews JE, Kost RG, Wagenknecht LE. Delivery of the research participant perception survey through the patient portal. *J Clin Transl Sci* 2018 Jun;2(3):163-168 [FREE Full text] [doi: [10.1017/cts.2018.32](https://doi.org/10.1017/cts.2018.32)] [Medline: [30370068](https://pubmed.ncbi.nlm.nih.gov/30370068/)]
12. Your Story. Your Health. The PRIDE Study. URL: <https://www.pridestudy.org/> [accessed 2019-10-10]
13. All of Us. URL: <https://www.joinallofus.org/En> [accessed 2019-10-10]
14. Patel AV, Jacobs EJ, Dudas DM, Briggs PJ, Lichtman CJ, Bain EB, et al. The American cancer society's cancer prevention study 3 (CPS-3): recruitment, study design, and baseline characteristics. *Cancer* 2017 Jun 1;123(11):2014-2024 [FREE Full text] [doi: [10.1002/cncr.30561](https://doi.org/10.1002/cncr.30561)] [Medline: [28171707](https://pubmed.ncbi.nlm.nih.gov/28171707/)]
15. Yin R. Qualitative Research from Start to Finish. Second Edition. New York, USA: Guilford Press; 2011.
16. Yin RK, Rk. Case Study Research: Design and Methods. Fourth Edition. Thousand Oaks, CA: Sage Publications; 2009.
17. Glaser, Barney, Anselm L. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago, USA: Aldine Publishing Company; 1967.
18. Nvivo. 2018. URL: <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home> [accessed 2020-09-24]
19. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. *Comput Inform Nurs* 2017 Nov;35(11):565-573. [doi: [10.1097/CIN.0000000000000377](https://doi.org/10.1097/CIN.0000000000000377)] [Medline: [28723832](https://pubmed.ncbi.nlm.nih.gov/28723832/)]
20. Wildenbos GA, Peute L, Jaspers M. Facilitators and barriers of electronic health record patient portal adoption by older adults: a literature study. *Stud Health Technol Inform* 2017;235:308-312. [Medline: [28423804](https://pubmed.ncbi.nlm.nih.gov/28423804/)]
21. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? Evidence and implications from a national sample of US adults. *Health Aff (Millwood)* 2018 Dec;37(12):1948-1954. [doi: [10.1377/hlthaff.2018.05117](https://doi.org/10.1377/hlthaff.2018.05117)] [Medline: [30633673](https://pubmed.ncbi.nlm.nih.gov/30633673/)]
22. Irizarry T, Shoemaker J, Nilsen ML, Czaja S, Beach S, DeVito Dabbs A. Patient portals as a tool for health care engagement: a mixed-method study of older adults with varying levels of health literacy and prior patient portal use. *J Med Internet Res* 2017 Mar 30;19(3):e99 [FREE Full text] [doi: [10.2196/jmir.7099](https://doi.org/10.2196/jmir.7099)] [Medline: [28360022](https://pubmed.ncbi.nlm.nih.gov/28360022/)]
23. Lyles CR, Sarkar U, Ralston JD, Adler N, Schillinger D, Moffet HH, et al. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: the diabetes and aging study. *J Am Med Inform Assoc* 2013;20(6):1128-1131 [FREE Full text] [doi: [10.1136/amiainl-2012-001567](https://doi.org/10.1136/amiainl-2012-001567)] [Medline: [23676243](https://pubmed.ncbi.nlm.nih.gov/23676243/)]
24. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient Portal Use and Experience Among Older Adults: Systematic Review. *JMIR Med Inform* 2017 Oct 16;5(4):e38 [FREE Full text] [doi: [10.2196/medinform.8092](https://doi.org/10.2196/medinform.8092)] [Medline: [29038093](https://pubmed.ncbi.nlm.nih.gov/29038093/)]
25. Howe N, Giles E, Newbury-Birch D, McColl E. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. *J Health Serv Res Policy* 2018 Apr 13;23(2):123-133. [doi: [10.1177/1355819617751555](https://doi.org/10.1177/1355819617751555)] [Medline: [29653503](https://pubmed.ncbi.nlm.nih.gov/29653503/)]

## Abbreviations

**ACS:** American Cancer Society  
**CPS-3:** Cancer Prevention Study-3  
**PGHD:** patient-generated health data

*Edited by G Eysenbach; submitted 04.03.20; peer-reviewed by G Kernohan, J Jones, K Fuji; comments to author 29.06.20; revised version received 03.08.20; accepted 02.09.20; published 01.10.20.*

*Please cite as:*

Rees-Punia E, Patel AV, Beckwitt A, Leach CR, Gapstur SM, Smith TG

*Research Participants' Perspectives on Using an Electronic Portal for Engagement and Data Collection: Focus Group Results From a Large Epidemiologic Cohort*

*J Med Internet Res* 2020;22(10):e18556

URL: <https://www.jmir.org/2020/10/e18556>

doi: [10.2196/18556](https://doi.org/10.2196/18556)

PMID: [33001033](https://pubmed.ncbi.nlm.nih.gov/33001033/)

©Erika Rees-Punia, Alpa V Patel, Asher Beckwitt, Corinne R Leach, Susan M Gapstur, Tenbroeck G Smith. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 01.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Threats of Bots and Other Bad Actors to Data Quality Following Research Participant Recruitment Through Social Media: Cross-Sectional Questionnaire

Rachel Pozzar<sup>1</sup>, RN, PhD; Marilyn J Hammer<sup>1</sup>, RN, DC, PhD; Meghan Underhill-Blazey<sup>1,2</sup>, APRN, PhD; Alexi A Wright<sup>3</sup>, MPH, MD; James A Tulskey<sup>4</sup>, MD; Fangxin Hong<sup>5</sup>, PhD; Daniel A Gundersen<sup>6</sup>, PhD; Donna L Berry<sup>1,7</sup>, RN, PhD

<sup>1</sup>Phyllis F Cantor Center for Research in Nursing and Patient Care Services, Dana-Farber Cancer Institute, Boston, MA, United States

<sup>2</sup>School of Nursing, University of Rochester, Rochester, NY, United States

<sup>3</sup>McGraw/Patterson Center for Population Sciences, Dana-Farber Cancer Institute, Boston, MA, United States

<sup>4</sup>Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Boston, MA, United States

<sup>5</sup>Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, United States

<sup>6</sup>Survey and Data Management Core, Dana-Farber Cancer Institute, Boston, MA, United States

<sup>7</sup>Department of Biobehavioral Nursing and Health Informatics, University of Washington, Seattle, WA, United States

**Corresponding Author:**

Rachel Pozzar, RN, PhD

Phyllis F Cantor Center for Research in Nursing and Patient Care Services

Dana-Farber Cancer Institute

450 Brookline Ave

LW517

Boston, MA,

United States

Phone: 1 857 215 0743

Email: [rachel\\_pozzar@dfci.harvard.edu](mailto:rachel_pozzar@dfci.harvard.edu)

## Abstract

**Background:** Recruitment of health research participants through social media is becoming more common. In the United States, 80% of adults use at least one social media platform. Social media platforms may allow researchers to reach potential participants efficiently. However, online research methods may be associated with unique threats to sample validity and data integrity. Limited research has described issues of data quality and authenticity associated with the recruitment of health research participants through social media, and sources of low-quality and fraudulent data in this context are poorly understood.

**Objective:** The goal of the research was to describe and explain threats to sample validity and data integrity following recruitment of health research participants through social media and summarize recommended strategies to mitigate these threats. Our experience designing and implementing a research study using social media recruitment and online data collection serves as a case study.

**Methods:** Using published strategies to preserve data integrity, we recruited participants to complete an online survey through the social media platforms Twitter and Facebook. Participants were to receive \$15 upon survey completion. Prior to manually issuing remuneration, we reviewed completed surveys for indicators of fraudulent or low-quality data. Indicators attributable to respondent error were labeled suspicious, while those suggesting misrepresentation were labeled fraudulent. We planned to remove cases with 1 fraudulent indicator or at least 3 suspicious indicators.

**Results:** Within 7 hours of survey activation, we received 271 completed surveys. We classified 94.5% (256/271) of cases as fraudulent and 5.5% (15/271) as suspicious. In total, 86.7% (235/271) provided inconsistent responses to verifiable items and 16.2% (44/271) exhibited evidence of bot automation. Of the fraudulent cases, 53.9% (138/256) provided a duplicate or unusual response to one or more open-ended items and 52.0% (133/256) exhibited evidence of inattention.

**Conclusions:** Research findings from several disciplines suggest studies in which research participants are recruited through social media are susceptible to data quality issues. Opportunistic individuals who use virtual private servers to fraudulently complete research surveys for profit may contribute to low-quality data. Strategies to preserve data integrity following research



participant recruitment through social media are limited. Development and testing of novel strategies to prevent and detect fraud is a research priority.

(*J Med Internet Res* 2020;22(10):e23021) doi:[10.2196/23021](https://doi.org/10.2196/23021)

## KEYWORDS

social media; internet; methods; data accuracy; fraud

## Introduction

Health research participants are increasingly recruited online [1]. Researchers may access potential research participants through a variety of online sources, including classified advertisements, search engine advertisements, survey panels, email listservs, crowdsourced online labor markets, and social media platforms [2-4]. Recruitment of health research participants through social media is particularly popular and has been reported in at least 69 unique papers published between 2011 and 2019 [5-7].

Recruitment of health research participants through social media may appeal to researchers for several reasons. First, 80% of US adults use social media, and rates of social media use exceed 60% in almost every sociodemographic category for which data are available [8]. Although only 40% of US adults aged 65 years and older use social media, this proportion has grown substantially from 12% in 2010 [8]. Second, social media platforms permit researchers to target advertisements to users according to their age, gender, education, location, interests, and behaviors [9]. Targeted social media advertisements enable researchers to direct their recruitment efforts toward individuals who are likely to meet study eligibility criteria. Third, the practical and ethical considerations of recruiting health research participants from social media have been well characterized. Guides to using social media to recruit participants to health research studies are available in the peer-reviewed literature and are increasingly produced by academic institutions [9-14]. Likewise, several authors have proposed approaches to ensure the protection of human research participants who are recruited through social media [11,13,15].

Researchers have sought to describe the extent to which participant recruitment through social media is cost-effective and efficient [2-4,7,16-19]. Although study results vary, some researchers suggest the use of social media may be more efficient and affordable than traditional recruitment methods in clinical settings [5]. Likewise, there is evidence that social media platforms effectively provide researchers with a way to access members of small or difficult-to-reach populations [7,11,12,16,20]. Despite these findings, studies in which research participants are recruited through social media are vulnerable to the same challenges associated with other methods of recruiting research participants online [21,22]. Respondent misrepresentation of eligibility criteria, duplicate enrollment, and automated enrollment by software applications known as

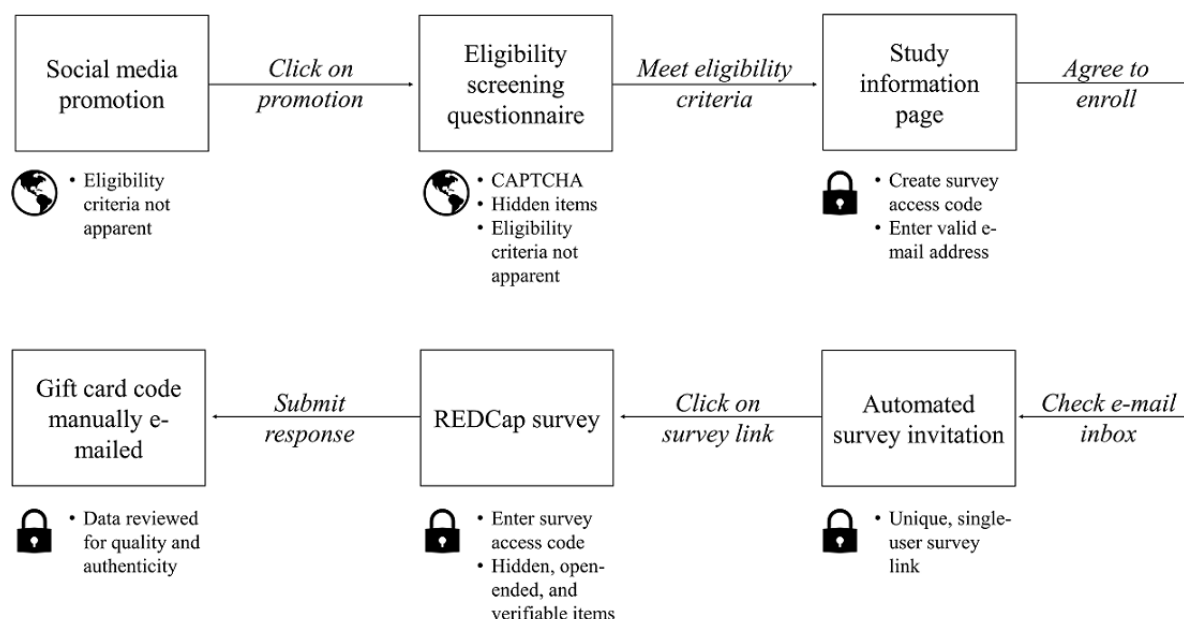
bots pose serious threats to sample validity and data integrity [23]. Nevertheless, these challenges are poorly described in the health sciences literature, particularly as they relate to the recruitment of health research participants through social media.

Ongoing development of best practices for all aspects of online research is necessary to encourage rigor and ensure judicious use of limited resources. The purpose of this paper is to describe and explain potential threats to sample validity and data integrity associated with the recruitment of health research participants through social media. We use our recent experience recruiting health research participants through social media as a case study. Drawing upon this example and from published research within and outside of the health sciences literature, we aim to provide a comprehensive overview of strategies that may be used to mitigate these threats.

## Methods

We designed a cross-sectional descriptive study that elicited patient perceptions of patient-provider communication in the ovarian cancer care setting. At the time of initial recruitment, eligible participants were English-speaking US adults diagnosed with ovarian cancer within the last 12 months. We planned to recruit participants through the Facebook and Twitter social media platforms and collect data online through a Research Electronic Data Capture (REDCap) survey [24]. Upon survey completion, valid participants would be issued a \$15 electronic gift card. The Dana-Farber/Harvard Cancer Center institutional review board (IRB) approved the study protocol.

We developed our study protocol (Figure 1) after reviewing published guides to recruiting health research participants online [9,11,23,25] and seeking advice from our institution's REDCap administrators and survey research core. First, we created a study page on Facebook and study account on Twitter. Next, we developed and planned to disseminate a set of Facebook posts, targeted Facebook advertisements, tweets (Twitter posts), and targeted Twitter advertisements. The Facebook page, Twitter account profile, Facebook posts, tweets, and targeted advertisements each included a brief overview of the study purpose and link to an eligibility screening questionnaire. Promotions described a gynecologic cancer communication study rather than an ovarian cancer communication study to prevent respondent misrepresentation of eligibility criteria [9,11]. No other details related to eligibility criteria were apparent from study promotions.

**Figure 1.** Procedure for participant recruitment and enrollment.

To access the eligibility screening questionnaire, respondents were required to pass a completely automated public Turing test to tell computers and humans apart (CAPTCHA) [23,25]. The eligibility screening questionnaire asked respondents to report how they heard about the study and used branching logic to deny access to ineligible respondents [25]. Respondents who reportedly met eligibility criteria were directed to a study information page that included all elements of informed consent. The study information page informed respondents that remuneration was limited to one gift card per participant and evidence of fraudulent activity may result in study removal [23]. Respondents who agreed to enroll in the study were prompted to provide their email address and create a survey passcode. Enrolled participants received automated emails containing a unique survey link and were required to enter their passcode to access the survey.

The survey included 124 closed- and 14 open-ended items. We pretested the survey and estimated that it would require 15 minutes to complete. We designed the survey to include several elements aimed at identifying low-quality or fraudulent responses. These included (1) a timestamp at the beginning and end of the survey, (2) hidden items, which are visible to bots but invisible to human respondents, and (3) pairs of items that could be used to identify inconsistent or illogical responses (eg, timestamp time zone and self-reported location). Prior to manually distributing participant remuneration, we planned to review completed surveys for evidence of inattention, duplicate or unusual responses to open-ended items, inconsistent responses to verifiable items, and evidence of automation. Specific examples from each of these categories are provided in the Results section.

We initiated recruitment with a single tweet that read “Help researchers learn about communication in gynecologic cancer care. Fill out a research survey from Dana-Farber Cancer

Institute and receive a \$15 Amazon gift card. Visit [link to the eligibility screening questionnaire] to learn more.” We also added the link to the eligibility screening questionnaire to the study Facebook page and Twitter account profile. We scheduled targeted advertisements to be launched at a later date.

## Results

Less than 7 hours after initiating recruitment, 576 respondents had completed the eligibility screening questionnaire. We suspected fraudulent activity after noting that although eligibility was limited to US residents, 82.5% (475/576) of responses to the eligibility screening questionnaire were submitted between the hours of midnight and 4:00 am Eastern Standard Time. In turn, we removed the tweet containing the link to the eligibility screening questionnaire, deleted the link from the Facebook page and Twitter account profile, and temporarily deactivated the survey.

Of the respondents who completed the eligibility screening questionnaire, 47.0% (271/576) reportedly met eligibility criteria, enrolled in the study, and completed the survey. Of the completed surveys, 47.2% (128/271) were submitted between the hours of 1:00 and 5:00 am in the participant’s reported time zone. The mean time to survey completion was 12.8 (SD 14.8) minutes. Three members of the study team (RP, MJH, and DLB) assessed completed surveys for quality and authenticity. We began by highlighting evidence of inattention, duplicate or unusual responses to open-ended items, inconsistent responses to verifiable items, and evidence of automation in each case. Next, we documented the specific indicators of low-quality or fraudulent data that were present in the data set. Indicators that could reasonably be attributed to respondent error or coincidence were labeled as suspicious, while those that strongly suggested automation or respondent misrepresentation were labeled as fraudulent (Table 1). Given the possibility that some legitimate

respondents could have completed the survey between the hours of 1:00 and 5:00 am, we opted not to include hour of survey submission on our list of indicators.

We classified cases with 1 fraudulent indicator or at least 3 suspicious indicators as fraudulent, cases with no fraudulent indicators and 1 to 2 suspicious indicators as suspicious, and cases with no fraudulent or suspicious indicators as legitimate.

In total, we classified 94.5% (256/271) as fraudulent, 5.5% (15/271) as suspicious, and none as legitimate. Most cases (241/271, 88.9%) exhibited more than 1 type of indicator of low-quality or fraudulent data. Of the fraudulent cases, 52.0% (133/256) exhibited evidence of inattention, with survey

completion times under 5 minutes in 24.6% (63/256) of cases and under 10 minutes in 27.3% (70/256) of cases. More than half of the fraudulent cases (138/256, 53.9%) included a duplicate or unusual response to an open-ended item. For example, in response to an item asking if participants wished to share anything else about communicating with doctors and other health professionals, 2 respondents entered “professional and technical personnel carry out film packaging management.” In response to an item asking participants what recommendations their clinicians had made about surgery, 6 respondents entered “the first choice surgery excision treatment, surgery pathology.”

**Table 1.** Indicators of low-quality or fraudulent data.

Indicator	Designation
<b>Evidence of inattention</b>	
Survey completion time <5 minutes	Fraudulent
Survey completion time <10 minutes	Suspicious
Same response provided to every closed-ended item on a survey page (straight lining)	Suspicious
<b>Duplicate or unusual responses to open-ended items</b>	
Exact response (consisting of more than 2-3 words) provided by more than one respondent	Fraudulent
Response is nonsensical or irrelevant to item	Suspicious
Several responses follow the same pattern in terms of phrasing or formatting	Suspicious
Response is an exact duplicate of text found on an existing website	Suspicious
<b>Inconsistent responses to verifiable items</b>	
Reported location and zip code prefix do not match	Suspicious
Reported location and timestamp time zone do not match	Suspicious
Reported treatment facility is not a cancer care facility	Suspicious
Timestamp time zone indicates survey was completed outside of the United States	Fraudulent
Response to “Where did you hear about this survey?” identified an organization that was not involved with recruitment	Suspicious
<b>Evidence of bot automation</b>	
Response provided to one or more hidden items	Fraudulent

In total, 86.7% (235/271) of cases included an inconsistent response to 1 or more verifiable items, and 16.2% (44/271) included a response to a hidden item. Every case that included a response to a hidden item and had valid timestamp data (25/271, 9.2%) exhibited a survey completion time under 2 minutes.

After consulting with our institution’s IRB, we removed fraudulent cases from the study without remuneration. We issued remuneration to the 15 respondents whose cases were classified as suspicious; however, we will exclude these cases from planned data analyses. We reinitiated recruitment by creating a duplicate REDCap project with a new URL. The new URL was not posted publicly; rather, promotions were limited to targeted Facebook advertisements and Facebook posts in private groups. Several months after successfully reinitiating recruitment in this fashion, we received 3 completed surveys in rapid succession. Upon review, we classified these cases as fraudulent. On review of Facebook user engagement with our targeted advertisements, we determined that a Facebook user who met

our targeting criteria had shared one of our advertisements in a public Facebook post. We promptly removed the advertisements from Facebook and reinitiated recruitment using a third REDCap project URL without further issues.

## Discussion

### Principal Findings

Our initial attempt to recruit health research participants through social media resulted in a large volume of low-quality and fraudulent data. Although we implemented strategies to prevent respondent misrepresentation of eligibility criteria and automated enrollment, hundreds of respondents navigated past checkpoints meant to restrict access to eligible human respondents.

Although our study protocol was informed by published guidance on the recruitment of health research participants through social media [9-14], discussions of data quality and

authenticity are largely absent from these works. Likewise, existing discussions of data quality and authenticity may be embedded in articles that discuss the challenges of online research more generally [23,25]. Researchers who consult the literature prior to recruiting health research participants through social media may overlook articles that do not refer to social media explicitly.

Our experience suggests studies in which research participants are recruited through social media are susceptible to many of the same pitfalls as studies in which participants are recruited through other online means [25-28]. In a related example, Dewitt and colleagues [22] conducted a cross-sectional descriptive study in which data were collected via web-based survey. The study team recruited research participants through an electronic mailing list and Facebook. Following data collection, they found that 60.5% (289/478) of completed survey responses were fraudulent. Similarly, Ballard and colleagues recruited research participants through an unspecified social media platform [21]. Following data collection, they determined that of the survey responses, 28.3% (117/414) were fraudulent and 10.1% (42/414) were potentially fraudulent. It is possible that the proportion of fraudulent responses was higher in our study because we shared the link to our eligibility screening questionnaire on both Facebook and Twitter. Nevertheless, these findings highlight the need to address issues of sample validity and data integrity as they pertain to the role of social media in health research.

Although issues of data quality and authenticity are not unique to studies in which research participants are recruited online, individuals who intend to defraud researchers may find that technology permits them to do so on a larger scale than would otherwise be possible. For example, bots can be programmed to rapidly complete online surveys. However, our experience and those of others suggest that the majority of fraudulent data cannot be attributed to bots alone [21,28]. All respondents in our study were able to pass a CAPTCHA, and only 16.2% (44/271) responded to one or more hidden survey items. Although some bots may be capable of passing a CAPTCHA and generating a fraudulent email address [22], access to our survey was restricted to respondents who provided a valid email address and had access to its inbox. Moreover, most respondents successfully identified a cancer treatment facility in the United States and entered a zip code prefix in the same geographic region. These activities require a degree of sophistication characteristic of human respondents [28].

Several authors have observed that *satisficing*, in which eligible respondents expend the minimal amount of cognitive effort needed to complete a survey, contributes to low-quality data

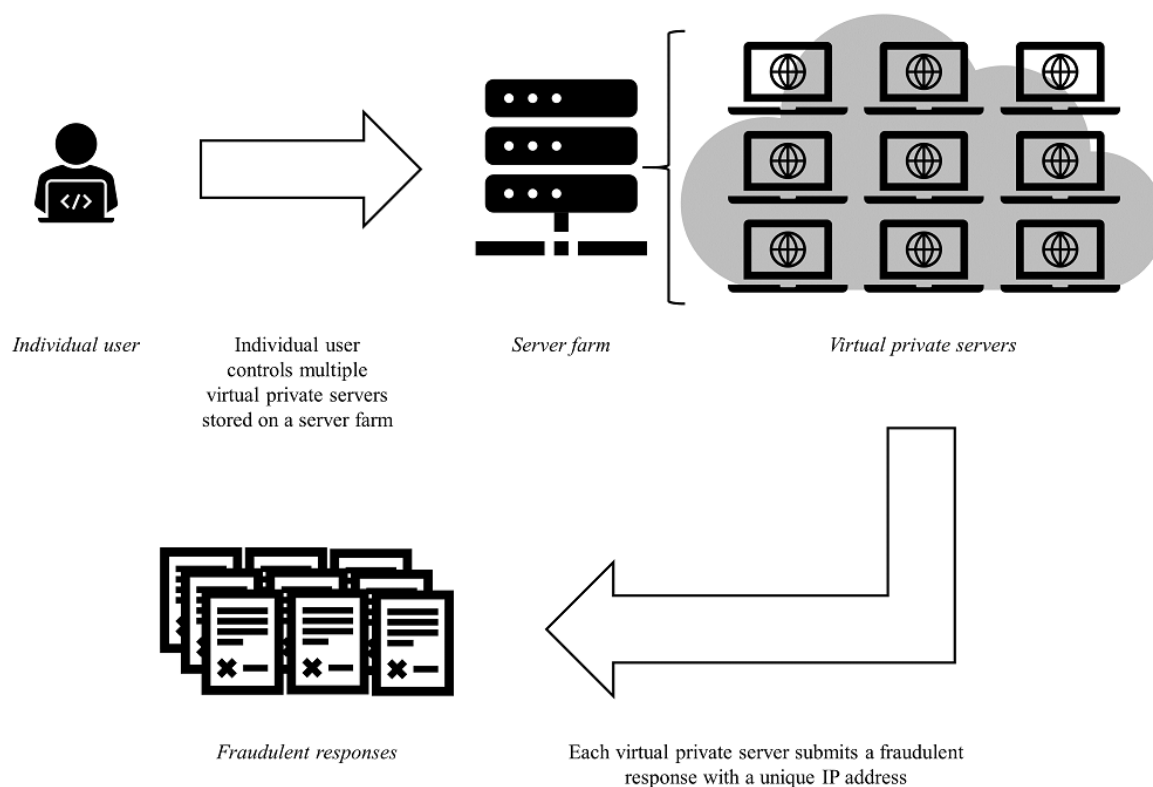
[29,30]. One limitation of our fraud detection protocol is the overlap between indicators of fraud and indicators of satisficing. However, given the speed with which we accumulated low-quality data, it is likely that our results largely reflect a coordinated effort by ineligible respondents to obtain remuneration rather than sample-wide satisficing. Groups of individuals who intend to defraud researchers may exchange information about online research studies that provide financial incentives [23]. Moreover, in a 2019 blog post, the founder of a company specializing in market research identified at least one website dedicated to training individuals to fraudulently complete large volumes of online surveys [31]. Although the phenomenon of respondent misrepresentation has been identified in the health sciences literature [25], the mechanisms by which low-quality and fraudulent survey responses are submitted by human respondents are poorly described. Improved understanding of this phenomenon is necessary to prevent the exploitation of research studies in which participants are recruited through social media and other online means.

### Role of Virtual Private Servers

Research from other disciplines offers insight into strategies used to defraud researchers who recruit research participants online. In the field of behavioral accounting, Dennis and colleagues [28] described 2 studies in which they recruited research participants through Amazon's Mechanical Turk, a crowdsourced online labor market. They received a large volume of responses that exhibited many of the same indicators of low-quality or fraudulent data that were present in our data. The authors used internet protocol (IP) address geolocation to identify the source of these responses and discovered multiple IP addresses with identical global positioning system coordinates. On further investigation, the authors determined that the IP addresses in question were associated with both a server farm and an internet service provider known to provide virtual private servers (VPSs).

Server farms are large collections of computer hardware housed in a single location. Server farms provide users with remote access to hardware with a processing capacity that exceeds that of a single computer. Each server farm can host a nearly unlimited number of VPSs, each of which functions like an individual computer but lacks its own physical hardware [28]. Like a physical computer, a VPS comprises data files, software programs, and an operating system [28]. An individual using more than one VPS would be able to use one physical computer to remotely program multiple VPSs to complete research surveys at the same time (Figure 2).



**Figure 2.** The role of virtual private servers in research participant misrepresentation.

Responses that originate from one individual using more than one VPS may be difficult to identify. Each VPS has a unique IP address associated with the physical location of the server farm rather than that of its user [28]. An individual using more than one VPS may masquerade as multiple respondents, each with a unique IP address. Furthermore, an individual using a VPS hosted on a server farm within the United States may mask his or her true location and circumvent strategies to limit study enrollment to US residents [28,32].

A VPS is not the only way in which an individual can mask his or her location. Virtual private networks, anonymous proxies, and spoofed IP addresses may be used for the same purpose. It is important to note that some individuals conceal their location or IP address out of privacy concerns and may not have malicious intent [32]. However, research suggests VPS use is associated with the collection of low-quality and fraudulent data following online recruitment of research participants.

Dennis and colleagues [28] used respondent IP addresses to compare the data they received from respondents who used a VPS to the data they received from those who did not. In open-ended item responses, respondents who used a VPS exhibited significantly higher proportions of English language misuse, incoherent or nonsensical phrases, duplicate responses, and responses that were copied verbatim from an existing website. The similarities between the responses received by Dennis and colleagues and our study team suggest individuals who use a VPS to defraud researchers are active outside of crowdsourced labor markets and may enroll in research studies that recruit participants through social media. The implications

of this finding for data quality are especially concerning given a recent analysis by Kennedy and colleagues [32], who analyzed 38 studies that recruited research participants through Amazon's Mechanical Turk and found evidence of respondents using a VPS as early as 2015.

### Strategies to Preserve Sample Validity and Data Integrity

Published papers within and outside of the health sciences literature offer suggestions to avoid collecting low-quality and fraudulent data from research participants recruited online. Although most strategies are applicable to studies that recruit research participants through social media, we provide additional suggestions that are specific to this approach. Limited research describes strategies to identify respondents using a VPS. Herein, we summarize the progress that has been made in this area to date and identify topics in need of further development. Strategies to prevent collection of low-quality or fraudulent data are proposed according to project phase below.

Preparation of study protocol and IRB application:

- Develop a written protocol for identifying and responding to low-quality data [22]
- Include language that permits the study team to verify respondent identities if needed (eg, via telephone call) [22,25]
- In consent document, state that participants will be removed from the study without remuneration in cases of fraud and participants will not receive additional remuneration for completing the study more than once [21,23]



- Mail remuneration to a physical address to avoid respondent misrepresentation of location-based eligibility criteria [21,25]
- Lower the value of or eliminate remuneration [23,25]
- Prepare study advertisements that do not explicitly state eligibility criteria [9,11]
- Seek guidance from institutional resources (eg, information systems, research computing, and the IRB)

#### Preparation of data collection instruments:

- Use a data collection platform with fraud prevention and detection features (eg, Qualtrics) [21-23,32]
- Use automated invitations to send each respondent a unique link to the data collection instrument [25]
- Ask respondents to identify where they heard about the study [25]
- Require respondents to pass a CAPTCHA [22,23,25]
- Collect respondent IP addresses (according to the Health Insurance Portability and Accountability Act Privacy Rule, IP address is considered an identifier) [21,23,25,28,32,33]
- Collect verifiable information, such as telephone number or physical address [21,23]
- Include at least one hidden item in each instrument. This can be accomplished by adding the @HIDDEN action tag to an item in REDCap or by adding custom JavaScript code to an item in Qualtrics
- Include a time stamp at the beginning and end of each instrument [21,23,25]
- Include (and consider requiring a response to) open-ended items [28]
- Include items with embedded directives (eg, “select the third option below”) [27]
- Include pairs of items that can be compared for consistency [23,25]
- Include items that require respondents to demonstrate insider knowledge [25]

#### Active recruitment and data collection:

- Avoid posting links to data collection instruments in the public space
- Use targeted advertisements to avoid promoting the study to ineligible respondents [9]
- Limit visibility of study-related social media profiles to audiences in the target geographic regions
- Monitor social media user engagement with study posts and advertisements (eg, for public shares or comments related to eligibility criteria)
- Monitor frequency and content of responses for suspicious patterns
- Identify respondents using a VPS with a tool such as the rIP R package or Shiny [32-34]

Researchers will need to weigh the potential benefits of each strategy against the financial and practical burden it may impose. For example, eliminating participant remuneration may remove the incentive for individuals who aim to defraud researchers [23,25]. However, survey completion and response rates are likely to be higher when remuneration is offered [35]. Entering participants into a raffle drawing for a larger incentive may serve as an acceptable compromise [25]. Alternatively, to verify

that respondents meet location-based eligibility criteria, researchers may elect to mail gift cards to a physical address rather than send them electronically [25]. Some researchers have reported successfully verifying respondent eligibility over the telephone [22,25], but as Teitcher and colleagues [23] observed, respondent eligibility verification is labor-intensive and may increase burden for legitimate participants.

Not every strategy mentioned will be appropriate for every research study. Similarly, no strategy will effectively preserve sample validity and data integrity when used alone. For example, although CAPTCHAs are intended to differentiate human respondents from bots, they are not always effective [22,23]. Likewise, although IP addresses can be used to verify that a respondent meets geographic eligibility criteria, IP-based geolocation is not always accurate [21,25]. Given that each strategy may be associated with one or more shortcomings, we recommend a comprehensive and multifaceted approach.

There is a need for research that develops and tests strategies to limit enrollment of individuals who may be using a VPS to defraud researchers. One approach has been proposed by Waggoner and colleagues [33], who developed a package called rIP for the statistical computing environment R (R Foundation for Statistical Computing). The rIP package provides researchers with the location of respondent IP addresses, information about likely VPS or server farm use, and a recommendation about whether to include the respondent's data in the data set. The team created an online version of the tool called Shiny that allows users to upload comma-separated value files for analysis in lieu of using R [33]. Although the rIP package and Shiny application have the potential to substantially reduce the workload associated with data quality review, prevention of low-quality responses is preferable to retrospective data classification. In a separate paper, Kennedy and colleagues [32] described embedding code in their Qualtrics survey to identify respondents whose IP address is associated with a server farm or VPS. The code used the IP verification website IP Hub [36] to identify these respondents and redirected them to a message informing them that they were ineligible to participate in the study. Additional solutions that capitalize on emerging knowledge of low-quality and fraudulent data sources are needed.

#### Limitations

Our study team did not collect the IP addresses of respondents. As such, we could not use the rIP R package or Shiny app [33] to determine whether a respondent used a VPS to access our survey. Future research that compares information provided by the rIP R package or Shiny app to the indicators of fraudulent or low-quality data that are described in this paper is warranted.

#### Conclusions

The recruitment of health research participants through social media is associated with several potential advantages. Nevertheless, studies in which research participants are recruited through social media are vulnerable to significant threats to sample validity and data integrity. There is a pressing need for best practices to prevent respondent misrepresentation of eligibility criteria and to identify low-quality and fraudulent

data. As health researchers increasingly turn to social media to access potential research participants, development of strategies to ensure rigor remains a priority.

## Acknowledgments

RP is supported by an American Cancer Society Postdoctoral Fellowship (133063-PF-19-102-01-CPPB) and a Gordon and Betty Moore Foundation Society for Medical Decision Making Fellowship in Medical Decision Making (GBMF7853).

## Conflicts of Interest

None declared.

## References

1. Rosa C, Campbell ANC, Miele GM, Brunner M, Winstanley EL. Using e-technologies in clinical trials. *Contemp Clin Trials* 2015 Nov;45(Pt A):41-54. [doi: [10.1016/j.cct.2015.07.007](https://doi.org/10.1016/j.cct.2015.07.007)] [Medline: [26176884](#)]
2. Dworkin J, Hessel H, Gliske K, Rudi JH. A comparison of three online recruitment strategies for engaging parents. *Fam Relat* 2016 Oct;65(4):550-561 [FREE Full text] [doi: [10.1111/fare.12206](https://doi.org/10.1111/fare.12206)] [Medline: [28804184](#)]
3. Guillory J, Kim A, Murphy J, Bradfield B, Nonnemaker J, Hsieh Y. Comparing Twitter and online panels for survey recruitment of e-cigarette users and smokers. *J Med Internet Res* 2016 Nov 15;18(11):e288 [FREE Full text] [doi: [10.2196/jmir.6326](https://doi.org/10.2196/jmir.6326)] [Medline: [27847353](#)]
4. Antoun C, Zhang C, Conrad FG, Schober MF. Comparisons of online recruitment strategies for convenience samples: Craigslist, Google AdWords, Facebook, and Amazon Mechanical Turk. *Field Methods* 2015 Sep 16;28(3):231-246. [doi: [10.1177/1525822X15603149](https://doi.org/10.1177/1525822X15603149)]
5. Reagan L, Nowlin SY, Birdsall SB, Gabbay J, Vorderstrasse A, Johnson C, et al. Integrative review of recruitment of research participants through Facebook. *Nurs Res* 2019;68(6):423-432. [doi: [10.1097/NNR.0000000000000385](https://doi.org/10.1097/NNR.0000000000000385)] [Medline: [31693547](#)]
6. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. *J Med Internet Res* 2016 Nov 7;18(11):e286 [FREE Full text] [doi: [10.2196/jmir.5698](https://doi.org/10.2196/jmir.5698)] [Medline: [27821383](#)]
7. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. *J Med Internet Res* 2017 Aug 28;19(8):e290 [FREE Full text] [doi: [10.2196/jmir.7071](https://doi.org/10.2196/jmir.7071)] [Medline: [28851679](#)]
8. Pew Research Center. Social Media Fact Sheet. Washington: Pew Internet and American Life Project; 2019. URL: <https://www.pewresearch.org/internet/fact-sheet/social-media/> [accessed 2020-07-27]
9. Carter-Harris L. Facebook targeted advertisement for research recruitment: a primer for nurse researchers. *Appl Nurs Res* 2016 Nov;32:144-147. [doi: [10.1016/j.apnr.2016.07.006](https://doi.org/10.1016/j.apnr.2016.07.006)] [Medline: [27969019](#)]
10. Akers L, Gordon JS. Using Facebook for large-scale online randomized clinical trial recruitment: effective advertising strategies. *J Med Internet Res* 2018 Nov 08;20(11):e290 [FREE Full text] [doi: [10.2196/jmir.9372](https://doi.org/10.2196/jmir.9372)] [Medline: [30409765](#)]
11. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. *Digit Health* 2018;4:2055207618771757 [FREE Full text] [doi: [10.1177/2055207618771757](https://doi.org/10.1177/2055207618771757)] [Medline: [29942634](#)]
12. Harvard Catalyst Regulatory Foundations, Ethics, & Law Program. The use of social media in recruitment to research: a guide for investigators and IRBs. 2017. URL: [https://catalyst.harvard.edu/pdf/regulatory/Social\\_Media\\_Guidance.pdf](https://catalyst.harvard.edu/pdf/regulatory/Social_Media_Guidance.pdf) [accessed 2020-07-27]
13. The use of social media in recruitment to research: a guide for investigators and IRBs. 2017. URL: [https://catalyst.harvard.edu/pdf/regulatory/Social\\_Media\\_Guidance.pdf](https://catalyst.harvard.edu/pdf/regulatory/Social_Media_Guidance.pdf) [accessed 2020-07-27]
14. Ramo D, Kaur M. Introduction to social media recruitment: opportunities and challenges. 2020. URL: <https://recruit.ucsf.edu/introduction-social-media-recruitment-opportunities-and-challenges> [accessed 2020-07-27]
15. Kamp K, Herbell K, Magginis WH, Berry D, Given B. Facebook recruitment and the protection of human subjects. *West J Nurs Res* 2019 Feb 07;41(9):1270-1281. [doi: [10.1177/0193945919828108](https://doi.org/10.1177/0193945919828108)] [Medline: [30729866](#)]
16. Wozney L, Turner K, Rose-Davis B, McGrath PJ. Facebook ads to the rescue? Recruiting a hard to reach population into an Internet-based behavioral health intervention trial. *Internet Interv* 2019 Sep;17:100246 [FREE Full text] [doi: [10.1016/j.invent.2019.100246](https://doi.org/10.1016/j.invent.2019.100246)] [Medline: [31080751](#)]
17. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. *JMIR Res Protoc* 2016;5(3):e161 [FREE Full text] [doi: [10.2196/resprot.5747](https://doi.org/10.2196/resprot.5747)] [Medline: [27511829](#)]
18. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. *Internet Interv* 2014 Apr;1(2):58-64. [doi: [10.1016/j.invent.2014.05.001](https://doi.org/10.1016/j.invent.2014.05.001)] [Medline: [25045624](#)]

19. Carter-Harris L, Bartlett ER, Warrick A, Rawl S. Beyond traditional newspaper advertisement: leveraging Facebook-targeted advertisement to recruit long-term smokers for research. *J Med Internet Res* 2016;18(6):e117 [FREE Full text] [doi: [10.2196/jmir.5502](https://doi.org/10.2196/jmir.5502)] [Medline: [27306780](https://pubmed.ncbi.nlm.nih.gov/27306780/)]
20. Nelson EJ, Loux T, Arnold LD, Siddiqui ST, Schootman M. Obtaining contextually relevant geographic data using Facebook recruitment in public health studies. *Health Place* 2019 Jan;55:37-42. [doi: [10.1016/j.healthplace.2018.11.002](https://doi.org/10.1016/j.healthplace.2018.11.002)] [Medline: [30466814](https://pubmed.ncbi.nlm.nih.gov/30466814/)]
21. Ballard AM, Cardwell T, Young AM. Fraud detection protocol for web-based research among men who have sex with men: development and descriptive evaluation. *JMIR Public Health Surveill* 2019 Feb 04;5(1):e12344 [FREE Full text] [doi: [10.2196/12344](https://doi.org/10.2196/12344)] [Medline: [30714944](https://pubmed.ncbi.nlm.nih.gov/30714944/)]
22. Dewitt J, Capistrant B, Kohli N, Rosser BRS, Mitteldorf D, Merengwa E, et al. Addressing participant validity in a small internet health survey (the Restore Study): protocol and recommendations for survey response validation. *JMIR Res Protoc* 2018 Apr 24;7(4):e96 [FREE Full text] [doi: [10.2196/resprot.7655](https://doi.org/10.2196/resprot.7655)] [Medline: [29691203](https://pubmed.ncbi.nlm.nih.gov/29691203/)]
23. Teitcher JEF, Bocking WO, Bauermeister JA, Hofer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. *J Law Med Ethics* 2015;43(1):116-133 [FREE Full text] [doi: [10.1111/jlme.12200](https://doi.org/10.1111/jlme.12200)] [Medline: [25846043](https://pubmed.ncbi.nlm.nih.gov/25846043/)]
24. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009 Apr;42(2):377-381 [FREE Full text] [doi: [10.1016/j.jbi.2008.08.010](https://doi.org/10.1016/j.jbi.2008.08.010)] [Medline: [18929686](https://pubmed.ncbi.nlm.nih.gov/18929686/)]
25. Kramer J, Rubin A, Coster W, Helmuth E, Hermos J, Rosenbloom D, et al. Strategies to address participant misrepresentation for eligibility in Web-based research. *Int J Methods Psychiatr Res* 2014 Mar;23(1):120-129 [FREE Full text] [doi: [10.1002/mpr.1415](https://doi.org/10.1002/mpr.1415)] [Medline: [24431134](https://pubmed.ncbi.nlm.nih.gov/24431134/)]
26. Conrique BG, McDade-Montez E, Anderson PM. Detection and prevention of data fraud in a study of community college career technical education students. *Community Coll J Res Pract* 2019 Jun 13;44(9):690-693. [doi: [10.1080/10668926.2019.1629126](https://doi.org/10.1080/10668926.2019.1629126)]
27. Jones MS, House LA, Gao Z. Respondent screening and revealed preference axioms: testing quarantining methods for enhanced data quality in web panel surveys. *Public Opin Q* 2015 Jun 02;79(3):687-709. [doi: [10.1093/poq/nfv015](https://doi.org/10.1093/poq/nfv015)]
28. Dennis S, Goodson B, Pearson C. Online worker fraud and evolving threats to the integrity of MTurk data: a discussion of virtual private servers and the limitations of IP-based screening procedures. *Behav Res Account* 2020;32(1):119-134. [doi: [10.2308/bria-18-044](https://doi.org/10.2308/bria-18-044)]
29. Hamby T, Taylor W. Survey satisficing inflates reliability and validity measures: an experimental comparison of college and Amazon Mechanical Turk samples. *Educ Psychol Meas* 2016 Dec;76(6):912-932 [FREE Full text] [doi: [10.1177/0013164415627349](https://doi.org/10.1177/0013164415627349)] [Medline: [29795893](https://pubmed.ncbi.nlm.nih.gov/29795893/)]
30. Lelkes Y, Krosnick JA, Marx DM, Judd CM, Park B. Complete anonymity compromises the accuracy of self-reports. *J Exp Soc Psychol* 2012 Nov;48(6):1291-1299. [doi: [10.1016/j.jesp.2012.07.002](https://doi.org/10.1016/j.jesp.2012.07.002)]
31. Pasternak O. Market research fraud: distributed survey farms exposed. 2019. URL: <https://persona.ly/blog/2019/01/market-research-fraud-distributed-survey-farms-exposed/> [accessed 2020-07-27]
32. Kennedy R, Clifford S, Burleigh T, Waggoner PD, Jewell R, Winter NJG. The shape of and solutions to the MTurk quality crisis. *Pol Sci Res Methods* 2020 Apr 24:1-16. [doi: [10.1017/psrm.2020.6](https://doi.org/10.1017/psrm.2020.6)]
33. Waggoner P, Kennedy R, Clifford S. Detecting fraud in online surveys by tracing, scoring, and visualizing IP addresses. *JOSS* 2019 May;4(37):1285. [doi: [10.21105/joss.01285](https://doi.org/10.21105/joss.01285)]
34. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2013.
35. Goritz A. Incentives in web studies: methodological issues and a review. *Int J Internet Sci* 2006;1(1):58-70 [FREE Full text]
36. IP Hub: Proxy and VPN detection API. 2020. URL: <https://iphub.info/> [accessed 2020-07-27]

## Abbreviations

**CAPTCHA:** completely automated, public Turing test to tell computers and humans apart

**IP:** internet protocol

**IRB:** institutional review board

**REDCap:** Research Electronic Data Capture

**VPS:** virtual private server

*Edited by G Eysenbach; submitted 30.07.20; peer-reviewed by M Katapodi, L Akers, W West; comments to author 14.09.20; revised version received 16.09.20; accepted 16.09.20; published 07.10.20.*

*Please cite as:*

*Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsy JA, Hong F, Gundersen DA, Berry DL*

*Threats of Bots and Other Bad Actors to Data Quality Following Research Participant Recruitment Through Social Media: Cross-Sectional Questionnaire*

*J Med Internet Res 2020;22(10):e23021*

URL: <http://www.jmir.org/2020/10/e23021/>

doi: [10.2196/23021](https://doi.org/10.2196/23021)

PMID: [33026360](https://pubmed.ncbi.nlm.nih.gov/33026360/)

©Rachel Pozzar, Marilyn J Hammer, Meghan Underhill-Blazey, Alexi A Wright, James A Tulsy, Fangxin Hong, Daniel A Gundersen, Donna L Berry. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 07.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Review

# The Role of Social Media in Enhancing Clinical Trial Recruitment: Scoping Review

Ida Darmawan<sup>1</sup>, MA; Caitlin Bakker<sup>2</sup>, MLIS, AHIP; Tabet A Brockman<sup>3,4</sup>, MA; Christi A Patten<sup>3,4</sup>, PhD; Milton Eder<sup>5,6</sup>, PhD

<sup>1</sup>Hubbard School of Journalism and Mass Communication, University of Minnesota, Minneapolis, MN, United States

<sup>2</sup>Health Sciences Libraries, University of Minnesota, Minneapolis, MN, United States

<sup>3</sup>Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, United States

<sup>4</sup>Center for Clinical and Translational Science, Mayo Clinic, Rochester, MN, United States

<sup>5</sup>Department of Family Medicine & Community Health, University of Minnesota, Minneapolis, MN, United States

<sup>6</sup>Clinical and Translational Science Institute, University of Minnesota, Minneapolis, MN, United States

**Corresponding Author:**

Ida Darmawan, MA  
Hubbard School of Journalism and Mass Communication  
University of Minnesota  
206 Church St SE  
Minneapolis, MN, 55455  
United States  
Phone: 1 612 735 8974  
Email: [darma002@umn.edu](mailto:darma002@umn.edu)

## Abstract

**Background:** Recruiting participants into clinical trials continues to be a challenge, which can result in study delay or termination. Recent studies have used social media to enhance recruitment outcomes. An assessment of the literature on the use of social media for this purpose is required.

**Objective:** This study aims to answer the following questions: (1) How is the use of social media, in combination with traditional approaches to enhance clinical trial recruitment and enrollment, represented in the literature? and (2) Do the data on recruitment and enrollment outcomes presented in the literature allow for comparison across studies?

**Methods:** We conducted a comprehensive literature search across 7 platforms to identify clinical trials that combined social media and traditional methods to recruit patients. Study and participant characteristics, recruitment methods, and recruitment outcomes were evaluated and compared.

**Results:** We identified 2371 titles and abstracts through our systematic search. Of these, we assessed 95 full papers and determined that 33 studies met the inclusion criteria. A total of 17 studies reported enrollment outcomes, of which 9 achieved or exceeded their enrollment target. The proportion of participants enrolled from social media in these studies ranged from 0% to 49%. Across all 33 studies, the proportion of participants recruited and enrolled from social media varied greatly. A total of 9 studies reported higher enrollment rates from social media than any other methods, and 4 studies reported the lowest cost per enrolled participant from social media.

**Conclusions:** While the assessment of the use of social media to improve clinical trial participation is hindered by reporting inconsistencies, preliminary data suggest that social media can increase participation and reduce per-participant cost. The adoption of consistent standards for reporting recruitment and enrollment outcomes is required to advance our understanding and use of social media to support clinical trial success.

(*J Med Internet Res* 2020;22(10):e22810) doi:[10.2196/22810](https://doi.org/10.2196/22810)

**KEYWORDS**

social media; clinical trial; recruitment methods; enrollment methods; review



## Introduction

Patient recruitment continues to be a major challenge in clinical trials, with 19% of trials being terminated due to poor recruitment and another one-third needing to extend recruitment time [1,2]. Delay and termination of clinical trials has significant scientific, ethical, and financial impact on patients, researchers, and society [3]. Termination of clinical trials due to recruitment issues leads to failure to obtain the necessary evidence to assess efficacy, safety, and effectiveness of an intervention, which could subsequently delay the implementation of a more beneficial therapy or allow an existing suboptimal standard therapy to remain in practice [3,4]. From an ethical perspective, study termination exposes enrolled patients to unnecessary risks and inconvenience due to the impossibility of generating sufficient data [3]. Furthermore, delay and termination also involve human, time, and financial opportunity costs, lower researchers' morale, and may reduce society's trust and subsequent willingness to participate in medical research [3].

Multiple factors contribute to recruitment failure in clinical trials, including lack of funds, complex and unclear trial design, and failure to find and interest eligible participants [5]. The inability to reach potential participants may result from overestimation of prevalence, competing trials, and ineffective advertising strategies [5]. Although researchers employ multiple strategies to improve clinical trials recruitment, evidence on best practices is lacking.

In recent years, researchers have increasingly augmented traditional methods, such as newspaper, radio or television advertisements, flyers, and signs on buses, with social media strategies to help accelerate enrollment and achieve their recruitment targets. Social media may offer distinct benefits compared with traditional methods due to its ability to target specific patient segments using customized messages that may resonate better with the target segments [6]. Additionally, several social media platforms have a higher proportion of users from minority groups (eg, African American use of Twitter), which can facilitate reach and diversity of recruitment for trials [6].

Research evaluating the effectiveness of social media in enhancing clinical trial recruitment in various settings has produced conflicting results [7-10]. Some studies find social media more effective and less costly than traditional methods [7,8], while others do not [9,11]. A scoping review on the use of social media in health research recruitment published in 2016 also yielded inconclusive findings [12]. Differences in the results reported appear to be due to the ways researchers defined social media, the types of studies included in the review, and variation in how studies compared and assessed recruitment and enrollment through social media and traditional methods.

In this study, we define social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" [13]. According to the Organization for Economic Co-operation and Development definition, user-generated content needs to be published on a website that is publicly accessible, show some

creative effort, and not be created in the context of professional work [14]. This definition excludes online media supported by internet-based platforms, such as email, instant messaging, regular websites, replications of existing content, and work created by companies for commercial use. We define traditional methods as offline (non-internet-based) platforms, such as television, print, and in-person recruitment. By using generally accepted definitions of social and traditional media, we address the heterogeneity observed in a prior scoping review and more clearly articulate the basis for including studies in this scoping review.

The prior scoping review included both observational and interventional studies [12], which potentially confounded the results. To avoid potential incommensurability across studies, this review includes only research studies that meet the National Institutes of Health's definition of a clinical trial [15]. Beyond refinements to the inclusion criteria, new studies comparing the use of social media and traditional methods for clinical trial recruitment have been published. As a result, two-thirds of the studies included in this scoping review were not evaluated in the prior scoping review.

Our scoping review examined the literature on the use of social media in conjunction with traditional methods in clinical trials to improve recruitment outcomes (success rate and cost). Specifically, this review addressed the following questions: (1) How is the use of social media, in combination with traditional approaches to enhance clinical trial recruitment and enrollment, represented in the literature? (2) Do the data on recruitment and enrollment outcomes presented in the literature allow for comparison across studies?

## Methods

### Review Method and Search Strategy

A scoping review was performed and reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR) guidelines, and the checklist is provided in [Multimedia Appendix 1](#) [16]. The protocol can be accessed through the Open Science Framework website [17]. Since this study focused on previously published literature and thus did not involve direct contact with human participants, institutional review board approval was not required [18].

We performed a comprehensive search in July 2019 using the following databases: PubMed; MEDLINE, EMBASE, and PsycInfo via Ovid; Cochrane Library via Wiley; Scopus; and Web of Science Core Collection. A combination of natural language and controlled vocabulary was employed in accordance with Methodological Expectations of Cochrane Intervention Reviews guidelines [19]. A complete search strategy is available in [Multimedia Appendix 2](#). To ensure that we did not overlook potentially relevant items, we also reviewed reference lists of related systematic and scoping reviews and included studies. Results were compiled and deduplicated in EndNote (version X.9; Clarivate Analytics).

## Inclusion and Exclusion Criteria

Papers employing strategies according to the stated definitions of social media and traditional methods were included. Papers also needed to be a clinical trial and describe at least one of the following outcomes: number of participants recruited or enrolled, cost of recruitment, or length of recruitment. Papers that contained no original data (editorials, letters to editors, opinions, conference proceedings, comments, systematic reviews), that recruited health care professionals as participants, or that did not report outcomes of interest were excluded.

## Screening

Papers were screened in 2 stages: (1) title and abstract screening and (2) full-text screening using the aforementioned inclusion and exclusion criteria. Two independent screeners reviewed papers at both the title and abstract stage and the full-text stage (ID, CB, and TAB). We resolved discrepancies through discussion and consensus or through the intervention of a third party when necessary. Study screening was facilitated using Rayyan (Qatar Computing Research Institute).

## Data Charting Process

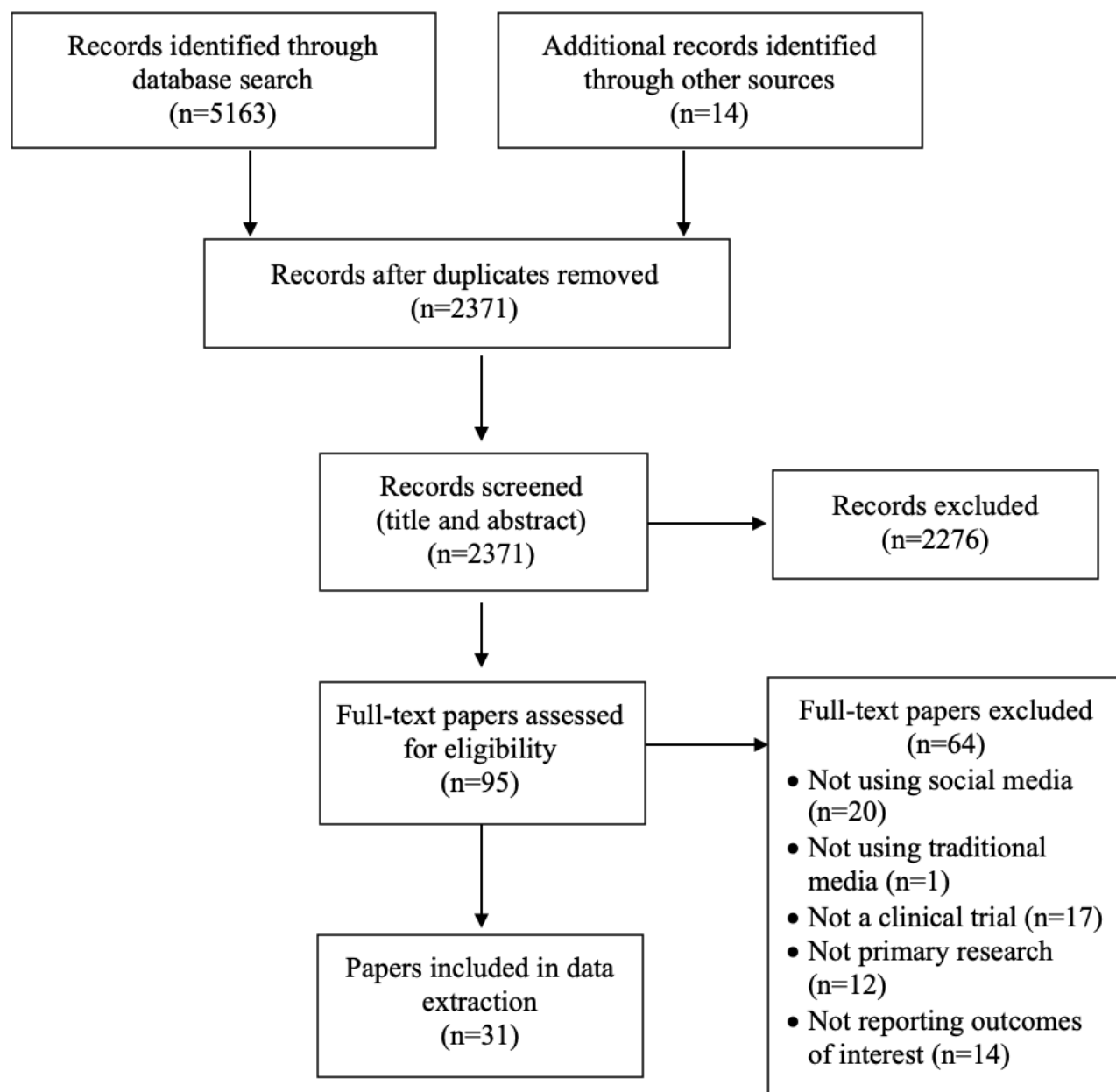
Two independent reviewers charted data using standardized collection forms created in REDCap (Vanderbilt University).

Again, discrepancies were resolved by discussion to achieve consensus. Data charted included study characteristics (goal of intervention, disease, mode of intervention), recruitment methods (social media platforms, traditional methods, other online methods), participant characteristics (age, sex, ethnicity, geographic location, type of residential area), and recruitment outcomes (number of participants recruited via each method, cost of recruitment, length of recruitment). We encountered differences in the way researchers described recruitment versus enrollment outcomes across studies. As a result, we defined recruitment as the first contact between prospective participants and study staff and enrollment as when participants were enrolled in the study after they signed informed consent. We extracted both recruitment and enrollment data when available.

## Results

### Paper Selection

Our search strategy identified a total of 5177 papers. After removing duplicates ( $n=2806$ ), the titles and abstracts of 2371 papers were screened against the exclusion criteria, resulting in 95 papers for full-text screening. From this, 64 papers did not meet inclusion criteria (Figure 1); the remaining 31 papers were included in this review and described in detail below.

**Figure 1.** PRISMA flow diagram.

### Study Characteristics

Of the 31 papers, 29 reported on a single study and 2 included reports on 2 studies each. This review identified a total of 33 studies that reported using social media and traditional methods in recruitment to clinical trials. The 33 studies described interventions for treatment (n=19) and prevention (n=14). A total of 19 studies consisted of in-person visits only, 10 were

online only, and 4 consisted of in-person and online participation. Study diseases or conditions focused on lifestyle and behavior (eg, weight loss, smoking), neurology, obstetrics/gynecology, and HIV. Most studies (32/33) recruited adults (18 years or older), with 5 studies focusing on older adults and seniors; 1 study recruited adolescents. Detailed characteristics of included studies are presented on [Table 1](#).

**Table 1.** Characteristics of included studies.

Study authors (year)	Purpose of intervention	Mode of intervention	Diseases or conditions	Participant's age (years)
Abbate et al (2017) [20]	Treatment	Online	Smoking	18+
Adam et al (2016) [21]	Prevention	In-person	Weight gain and diet	23-40
Alley et al (2016) [22]	Prevention	Online	Physical activity	18+
Bracken et al (2019) [23]	Prevention	In-person	Type 2 diabetes	50-74
Buckingham et al (2017) [24]	Prevention	In-person	HIV	18+
Burrell et al (2012) [25]	Prevention	In-person	HIV	18+
Cowie et al (2018) [26]	Treatment	In-person	Alzheimer disease	60+
Frandsen et al (2014) [27]	Treatment	In-person	Smoking	18+
Frandsen et al (2016) [9]	Treatment	In-person	Smoking	18+
Guthrie et al (2019) [28]	Treatment	In-person	Postmenopausal vulvovaginal symptoms	45-70
Heffner et al (2013) [29]	Treatment	Online	Smoking	18+
Huesch et al (2018) [30]	Treatment	In-person	Bipolar disorder	18+
Ince et al (2014) [31]	Treatment	Online	Depressive symptoms	18+
Jones et al (2012) [32]	Prevention	In-person	Physical activity	16-18 <sup>a</sup>
Jones et al (2017) [33]	Prevention	Online	HIV	18-29
Juraschek et al (2018) [34]	Prevention	Both	Cancer survivors	18+
Kayrouz et al (2016) [10]	Treatment	Online	Anxiety, depression	18-70
Kendall et al (2018) [35]	Treatment	In-person	Chronic dizziness, chronic neck pain	65-85
Kira et al (2016) [36]	Prevention	In-person	Smoking	16+
Kuhn et al (2017) [37]	Treatment	Online	PTSD <sup>b</sup>	18+
Nash et al (2017) [38]	Treatment	In-person	Hypertension	18-69
Partridge et al (2015) [39]	Prevention	Both	Weight gain	18-35
Rabin et al (2013) [40]	Prevention	Online	Physical activity	18-39
Raviotta et al (2016) [41]	Prevention	Both	HPV <sup>c</sup>	18-25
Rounds et al (2019) [42]	Treatment	Both	Obesity	18-65
Sanchez et al (2018a) [43]	Treatment	In-person	ARHL <sup>d</sup>	18+
Sanchez et al (2018b) [43]	Treatment	In-person	ARHL	50-89
Shere et al (2014) [44]	Treatment	In-person	Pregnancy	18-45
Usadi et al (2015a) [45]	Treatment	In-person	Infertility	18-40
Usadi et al (2015b) [45]	Treatment	In-person	Infertility	18-40
Volkova et al (2017) [46]	Prevention	Online	Healthy food purchase	18+
Waltman et al (2019) [47]	Prevention	In-person	Bone loss	19+
Watson et al (2018) [48]	Treatment	Online	Smoking	18+

<sup>a</sup>Participants were 11th-grade students. Age range is estimated.

<sup>b</sup>PTSD: posttraumatic stress disorder.

<sup>c</sup>HPV: human papillomavirus.

<sup>d</sup>ARHL: age-related hearing loss.

## Recruitment Methods

Facebook was the most commonly used social media platform (31/33), followed by Twitter (6/33). The majority of studies (21/33) used a single social media platform and the remainder used two or more social media platforms, with one study reporting the use of 4 platforms. Facebook use occurred across all studies reporting the use of more than one social media platform.

Studies that reported using Facebook for their recruitment used a variety of Facebook features: Facebook ads, Facebook pages (including boosted posts on a page) [10,36], or sending a friend invite to prospective participants using a Facebook account [32]. One study used untargeted Facebook ads, then switched to targeted Facebook ads that showed ads only to participants who met the study criteria [22]. Several studies started using social media after other recruitment methods were deployed [28,31,38,44], while others used social media at the beginning of their recruitment period.

The most commonly used traditional method was print (32/33), followed by in-person venues (17/33) and referrals from health care professionals (16/33). A total of 23 studies used other online

media, such as website ads (15/23), emails (11/23), and Craigslist (8/23).

## Recruitment and Enrollment Rates

A total of 17 out of 33 studies reported overall enrollment rates. Of these, 9 studies achieved or exceeded their enrollment target. Of these 9 studies, 8 reported the proportion of participants enrolled through social media, which ranged from 0% to 49%. One study that achieved its enrollment target reported that social media outperformed other recruitment methods [48].

About half of the studies (17/33) reported both recruitment and enrollment rates from social media, 3 studies reported recruitment rates only, 11 studies reported enrollment rates, and 2 did not report either (Table 2). The proportion of participants recruited and enrolled from social media varied greatly from study to study (Table 2) and across study types (Table 3). Studies with a high proportion (50% or greater) of participants enrolled from social media recruited adult participants. Studies recruiting seniors [26,35,43] and the 1 study involving adolescents [32] enrolled the majority of participants from traditional methods (Table 3).



**Table 2.** Reported recruitment and enrollment from social media.

Study authors (year)	Participants recruited from social media, n/N (%) <sup>a</sup>	Participants enrolled from social media, n/N (%) <sup>a</sup>
Abbate et al (2017) [20]	— <sup>b</sup>	24/151 (16)
Adam et al (2016) [21]	45/126 (36)	25/70 (36)
Alley et al (2016) [22]	205/278 (74)	74/140 (53) <sup>c</sup>
Bracken et al (2019) [23]	369/19,022 (2)	16/1007 (2)
Buckingham et al (2017) [24]	598/1945 (31)	48/96 (50)
Burrell et al (2012) [25]	—	24/105 (23)
Cowie et al (2018) [26]	621/857 (72)	—
Frandsen et al (2014) [27]	—	138/266 (52)
Frandsen et al (2016) [9]	228/414 (55)	92/175 (53)
Guthrie et al (2019) [28]	461/2627 (18)	25/302 (8)
Heffner et al (2013) [29]	—	11/222 (5)
Huesch et al (2018) [30]	117/147 (80)	11/17 (65)
Ince et al (2014) [31]	227/287 (79)	75/96 (78)
Jones et al (2012) [32]	—	43/589 (7)
Jones et al (2017) [33]	940/1435 (66)	153/247 (62)
Juraschek et al (2018) [34]	24/121 (6)	4/121 (3)
Kayrouz et al (2016) [10]	—	70/81 (86)
Kendall et al (2018) [35]	38/162 (23)	8/24 (33)
Kira et al (2016) [36]	1/74 (1)	1/24 (4)
Kuhn et al (2017) [37]	—	22/120 (18)
Nash et al (2017) [38]	—	—
Partridge et al (2015) [39]	20/1181 (2)	5/250 (2)
Rabin et al (2013) [40]	11/73 (15) <sup>d</sup>	0/12 (0)
Raviotta et al (2016) [41]	—	44/220 (20)
Rounds et al (2019) [42]	—	3/102 (3)
Sanchez et al (2018a) [43]	4/425 (1)	0/91 (0)
Sanchez et al (2018b) [43]	N/A <sup>e</sup>	0/79 (0)
Shere et al (2014) [44]	—	—
Usadi et al (2015a) [45]	7/3358 (0.2)	N/A
Usadi et al (2015b) [45]	3/3727 (0.1)	N/A
Volkova et al (2017) [46]	966/2448 (40)	584/1357 (43)
Waltman et al (2019) [47]	838/3033 (28)	44/276 (16)
Watson et al (2018) [48]	—	1299/2637 (49)

<sup>a</sup>Some studies allowed participants to be counted in multiple recruitment/enrollment methods.<sup>b</sup>Not available (not reported by original study).<sup>c</sup>Includes participants enrolled from targeted Facebook ads only.<sup>d</sup>Includes participants recruited from emails and Craigslist.<sup>e</sup>N/A: not applicable. Reported as N/A due to inconsistencies in recruitment and enrollment data.

**Table 3.** Social media recruitment and enrollment rates by study type.

Study type	Studies reporting recruitment rate, n	Range of participants recruited from social media, %	Studies reporting enrollment rate, n	Range of participants enrolled from social media, %
<b>Intervention purpose</b>				
Treatment	10	0-80	14	0-86
Prevention	10	1-74	14	0-62
<b>Disease or condition type</b>				
Lifestyle-related	6	1-74	13	0-53
Neurological	4	23-80	5	18-86
OB/GYN <sup>a</sup>	4	0-28	2	8-16
HIV	2	31-66	3	23-62
Others	4	1-6	5	0-20
<b>Mode of intervention</b>				
In-person	14	0-80	14	0-65
Online	4	15-79	10	0-86
Both	2	2-6	4	2-20
<b>Participant age group</b>				
Adults (18+ years)	15	0-80	23	0-86
Older adults/seniors	5	1-72	4	0-33
Adolescents	0	— <sup>b</sup>	1	7

<sup>a</sup>OB/GYN: obstetrics/gynecology.<sup>b</sup>Not available (not reported by original study).

Out of 20 studies that reported recruitment rates by method, 7 reported higher recruitment rates from social media than from any other methods. Similarly, out of 28 studies that reported enrollment rates by method, only 9 reported higher enrollment rates using social media than any other methods (Table 4). Among the 9 studies that reported the highest enrollment rates from social media, 6 studies were for treatment (3

lifestyle-related conditions, 3 neurological) and 3 were for prevention (2 HIV, 1 lifestyle). Of the 9 studies reporting enrollment, 5 involved an online intervention and the remainder used an in-person intervention. A total of 5 of these 9 studies used Facebook only, and the rest used more than one social media platform.

**Table 4.** Enrollment rates by recruitment method.<sup>a</sup>

Study authors (year)	Enrolled from social media, %	Enrolled from traditional methods, %	Enrolled from other online media, %
Abbate et al (2017) [20]	16	40	57
Adam et al (2016) [21]	36	64	N/A <sup>b</sup>
<i>Alley et al<sup>c</sup></i> (2016) [22]	53 <sup>d</sup>	47 <sup>e</sup>	N/A
Bracken et al (2019) [23]	2	94	1
<i>Buckingham et al</i> (2017) [24]	50	26	24
Burrell et al (2012) [25]	23	77 <sup>e</sup>	N/A
<i>Frandsen et al</i> (2014) [27]	52	47	N/A
<i>Frandsen et al</i> (2016) [9]	53	47	N/A
Guthrie et al (2019) [28]	8	92	N/A
Heffner et al (2013) [29]	5	25	70
<i>Huesch et al</i> (2018) [30]	65	35	N/A
<i>Ince et al</i> (2014) [31]	78	2	0
Jones et al (2012) [32]	7	94	N/A
<i>Jones et al</i> (2017) [33]	62	38	N/A
Juraschek et al (2018) [34]	3	77	N/A
<i>Kayrouz et al</i> (2016) [10]	86	14	N/A
Kendall et al (2018) [35]	33	67	N/A
Kira et al (2016) [36]	4	88	N/A
Kuhn et al (2017) [37]	18	22	60
Partridge et al (2015) [39]	2	71	25
Rabin et al (2013) [40]	0	100	N/A
Raviotta et al (2016) [41]	20	80	N/A
Rounds et al (2019) [42]	3	81	16
Sanchez et al (2018a) [43]	0	19	5
Sanchez et al (2018b) [43]	0	53	4
Volkova et al (2017) [46]	43	46	11
Waltman et al (2019) [47]	16	44	1
<i>Watson et al</i> (2018) [48]	49	11	39

<sup>a</sup>Some studies allowed participants to be counted in multiple recruitment and enrollment methods.

<sup>b</sup>N/A: not applicable.

<sup>c</sup>Italics signify studies that reported the highest enrollment rates from social media.

<sup>d</sup>Includes participants enrolled from targeted Facebook ads only.

<sup>e</sup>Includes participants enrolled from both traditional and other online media.

## Recruitment and Enrollment Costs

A total of 20 studies reported itemized recruitment and enrollment costs. Of these, almost all studies reported the costs to place the recruitment ads in various media (19/20), and 6 reported the costs to develop these ads. Only 5 studies reported the costs of staff involved in participant recruitment and enrollment.

Of the 19 studies that reported cost per enrolled participant, only 4 [10,21,30,33] reported lower cost per enrolled participant using social media than any other methods. Recruitment methods used in these 4 studies included Facebook, Instagram, print, television, radio, in-person recruitment, and website ads. Detailed costs per enrolled participant by media type can be found in Table 5.

**Table 5.** Costs per enrolled participant by media type.<sup>a</sup>

Study authors (year)	Cost per enrolled participant		
	Social media	Traditional methods	Other online media
Abbate et al (2017) [20]	N/A <sup>b</sup>	US \$8.28 <sup>c</sup>	N/A
Adam et al (2016) [21]	Can \$20.28	Can \$24.15	N/A
Bracken et al (2019) [23]	N/A	Aus \$594	N/A
Frandsen et al (2014) [27]	Aus \$56.34	Aus \$52.33	N/A
Frandsen et al (2016) [9]	Aus \$42.34	Aus \$21.52	N/A
Guthrie et al (2019) [28]	US \$593	US \$356	N/A
Heffner et al (2013) [29]	US \$172.76	US \$5.27-\$46.98	US \$26.19-\$50.26
Huesch et al (2018) [30]	US \$18	US \$635	N/A
Ince et al (2014) [31]	€33	N/A	N/A
Jones et al (2017) [33]	US \$66.46	US \$149.62	N/A
Juraschek et al (2018) [34]	US \$1426	US \$436-\$917	N/A
Kayrouz et al (2016) [10]	US \$37	US \$40	N/A
Kendall et al (2018) [35]	N/A	Aus \$2141 <sup>c</sup>	N/A
Nash et al (2017) [38]	Aus \$45.15-\$176	N/A	N/A
Partridge et al (2015) [39]	Aus \$945.33	Aus \$144.52-\$212.51	Aus \$11.98-\$571.45
Raviotta et al (2016) [41]	US \$110	US \$61	N/A
Volkova et al (2017) [46]	NZ \$5 <sup>d</sup>	NZ \$4-\$179	NZ \$4
Waltman et al (2019) [47]	US \$119.38	US \$29.36-\$926.90	US \$1000
Watson et al (2018) [48]	US \$40.51	US \$20.30	US \$13.95-34.71

<sup>a</sup>Currency exchange rates of Can \$1=US \$0.75, Aus \$1=US \$0.72, €1=US \$1.17, and NZ \$1=US \$0.66 were applicable at the time of publication.

<sup>b</sup>N/A: not applicable.

<sup>c</sup>Cost per enrolled participant across all media.

<sup>d</sup>Includes cost using Craigslist.

## Discussion

### Principal Findings

This scoping review examines literature describing the use of social media and traditional methods as well as other online media in clinical trial recruitment. We found that using social media resulted in the highest recruitment rates in 7 of 20 studies and the highest enrollment rate in 9 of 28 studies. We also found that social media resulted in the lowest cost per enrolled participant in 4 of 19 studies. However, the data reported about social media outcomes varied greatly across studies, obscuring our ability to evaluate if particular studies might benefit more from using social media.

Our review discovered a lack of consistency in defining and reporting recruitment and enrollment data across studies. Several studies seemed to use these terms interchangeably, and some failed to specify which data they reported (eg, when consent occurred). Furthermore, we noted inconsistencies resulting from the varied use of social media, even within a platform, and the opportunistic use of social media by several studies that started using social media only after traditional methods failed to achieve sufficient interest and enrollment. We also observed

that some studies did not indicate how many participants were counted as being reached by more than one recruitment or enrollment method. Lastly, we observed inconsistencies in cost reporting, with some studies reporting itemized costs and others reporting aggregated costs. Collectively, these inconsistencies contributed to the variations in describing the outcomes of using social media and traditional methods to support clinical trial recruitment and enrollment. The inconsistencies in reporting also rendered comparison of data across the studies included in this scoping review impossible.

To address these inconsistencies, we recommend that future studies use the terms “recruitment” and “enrollment” in alignment with the definitions used by the Food and Drug Administration’s advisory committee [49]. The term “recruitment” should report on all interactions with potential study participants prior to obtaining informed consent, while the term “enrollment” should report only on the total number of individuals who provide informed consent. Studies should report recruitment using as the denominator the total number of participants reached and as numerators the number of study participants reached through each method. Enrollment rates should be calculated using as the denominator the total number of participants providing informed consent and as numerators

the number of participants enrolled through each recruitment method. Additionally, screening that results in the exclusion of participants after they provide informed consent should be reported separately, enabling a more accurate count and cost of the individuals actively participating in the intervention. Using consistent definitions of key terms across research studies will allow researchers to improve the clarity and comparability of clinical trial management data. Establishing reporting standards will also allow researchers to adopt evidence-based recruitment strategies by choosing optimal recruitment methods to reduce cost and timelines [50].

To facilitate shared learning based on the experience of prior research, we further recommend that future studies report in detail their approaches to using each social media platform (eg, Facebook ad, Facebook page) and specify the number of participants recruited or enrolled from multiple methods, which will allow for more accurate response analyses and cost estimates. As the ability to capitalize on existing social media relationships may influence both the success of recruitment and enrollment as well as their subsequent costs, researchers should consider explicitly stating whether their social media recruitment strategies are related to a larger social media initiative.

When adding social media to existing recruitment methods, researchers should report both the incremental costs of social media and the personnel cost to develop and monitor posted content. Furthermore, researchers are encouraged to report trade-offs between cost and time when analyzing the effectiveness of each recruitment method. For example, social media may result in higher cost per enrolled participant but reduce the time to enroll participants. Although we recognize the complexity in analyzing and comparing cost in different time increments and the potential for historical trends to confound results, such reporting may provide deeper insights on different recruitment methods. Such analyses may help researchers, organizations, and corporations determine whether to build social media expertise directly into their research teams to optimize the full extent of social media capabilities or to engage social media experts as consultants. The costs associated with such arrangements should be identified and, ideally, reported.

### Comparison With Prior Work

The mixed results on the use of social media found in this study are similar to those found in a prior scoping review [12]. While

the prior scoping review attributed the different success rates to the effort researchers put into the use of social media as a recruitment method, we found that the variations in the use of social media and the inconsistencies in the way outcomes were reported prevented comparisons across studies. Having explored the ways researchers used and reported the use of social media in clinical trial recruitment, this study advances recommendations for achieving data consistency in reporting, which would facilitate comparison across results.

Similar to the findings from a systematic mapping study conducted by Frampton et al [51], we found that the use of social media to recruit participants involved a variety of diseases and health conditions. We also failed to find many studies involving minority populations (eg, African American or lesbian, gay, bisexual, transgender, and queer populations). As a result, we are unable to provide any insights on the use of social media to recruit diverse populations when compared with traditional recruitment methods.

### Limitations

Our scoping review is limited in its examination of data from clinical trials that combined social media and traditional methods in their recruitment strategies because we cannot guarantee that the review identified all relevant studies. We also cannot claim this review is comprehensive, as research not organized or reported as a clinical trial may inform reasoning about incorporating social media into the research recruitment armamentarium. The relatively recent use of social media also renders highly speculative any conclusions drawn about its appropriateness in recruitment and enrollment methods for particular diseases and conditions. We also recognize a limitation in our ability to assess the quality of the recruitment strategy, the recruitment materials, or the enrollment process. However, we advance recommendations to improve consistency in reporting on recruitment and enrollment that we believe to be necessary precursors for comparative analysis.

### Conclusions

The use of social media for clinical trial recruitment holds great promise. However, this scoping review identified continued inconsistency in reports on the use of social media for clinical trial recruitment and enrollment. We recommend that future studies incorporate the recommendations for collecting and reporting recruitment and enrollment data advanced here to facilitate comparison of study data and shared learning.

### Acknowledgments

This project was supported by grant number UL1 TR002494 and UL1 TR002377 from the National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

### Authors' Contributions

ID developed the project concept, performed database searches, screened papers, extracted data, analyzed social media use and recruitment and enrollment outcomes across studies, and was the lead in writing the manuscript. CB and TAB refined the project concept, screened papers, extracted data, and were major contributors in writing the manuscript. CAP was a major contributor in writing the manuscript. ME developed the project concept, analyzed recruitment and enrollment outcomes across studies, and was a major contributor in writing the manuscript.



**Conflicts of Interest**

None declared.

Multimedia Appendix 1

PRISMA-ScR checklist.

[PDF File (Adobe PDF File), 103 KB - [jmir\\_v22i10e22810\\_app1.pdf](#)]

Multimedia Appendix 2

Complete search strategy.

[XLSX File (Microsoft Excel File), 9 KB - [jmir\\_v22i10e22810\\_app2.xlsx](#)]

**References**

1. Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, STEPS group. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. *Health Technol Assess* 2007 Nov;11(48):iii, ix-iii,105 [FREE Full text] [doi: [10.3310/hta11480](#)] [Medline: [17999843](#)]
2. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. *Clin Trials* 2015 Feb;12(1):77-83 [FREE Full text] [doi: [10.1177/1740774514558307](#)] [Medline: [25475878](#)]
3. Malmqvist E, Juth N, Lynöe N, Helgesson G. Early stopping of clinical trials: charting the ethical terrain. *Kennedy Inst Ethics J* 2011 Mar;21(1):51-78. [doi: [10.1353/ken.2011.0002](#)] [Medline: [21598846](#)]
4. Task Force of the Working Group on Arrhythmias of the European Society of Cardiology. The early termination of clinical trials: causes, consequences, and control. With special reference to trials in the field of arrhythmias and sudden death. *Circulation* 1994 Jun;89(6):2892-2907. [doi: [10.1161/01.cir.89.6.2892](#)] [Medline: [8205706](#)]
5. Briel M, Olu KK, von Elm E, Kasenda B, Alturki R, Agarwal A, et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *J Clin Epidemiol* 2016 Dec;80:8-15. [doi: [10.1016/j.jclinepi.2016.07.016](#)] [Medline: [27498376](#)]
6. Caplan A, Friesen P. Health disparities and clinical trial recruitment: Is there a duty to tweet? *PLoS Biol* 2017 Mar;15(3):e2002040 [FREE Full text] [doi: [10.1371/journal.pbio.2002040](#)] [Medline: [28249024](#)]
7. Guillory J, Wiant KF, Farrelly M, Fiocco L, Alam I, Hoffman L, et al. Recruiting Hard-to-Reach Populations for Survey Research: Using Facebook and Instagram Advertisements and In-Person Intercept in LGBT Bars and Nightclubs to Recruit LGBT Young Adults. *J Med Internet Res* 2018 Jun 18;20(6):e197 [FREE Full text] [doi: [10.2196/jmir.9461](#)] [Medline: [29914861](#)]
8. Admon L, Haefner JK, Kolenic GE, Chang T, Davis MM, Moniz MH. Recruiting Pregnant Patients for Survey Research: A Head to Head Comparison of Social Media-Based Versus Clinic-Based Approaches. *J Med Internet Res* 2016 Dec 21;18(12):e326 [FREE Full text] [doi: [10.2196/jmir.6593](#)] [Medline: [28003174](#)]
9. Frandsen M, Thow M, Ferguson SG. The Effectiveness Of Social Media (Facebook) Compared With More Traditional Advertising Methods for Recruiting Eligible Participants To Health Research Studies: A Randomized, Controlled Clinical Trial. *JMIR Res Protoc* 2016 Aug 10;5(3):e161 [FREE Full text] [doi: [10.2196/resprot.5747](#)] [Medline: [27511829](#)]
10. Kayrouz R, Dear BF, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. *Internet Interv* 2016 May;4:1-10 [FREE Full text] [doi: [10.1016/j.invent.2016.01.001](#)] [Medline: [30135786](#)]
11. Moreno MA, Waite A, Pumper M, Colburn T, Holm M, Mendoza J. Recruiting Adolescent Research Participants: In-Person Compared to Social Media Approaches. *Cyberpsychol Behav Soc Netw* 2017 Jan;20(1):64-67. [doi: [10.1089/cyber.2016.0319](#)] [Medline: [27976951](#)]
12. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. *J Med Internet Res* 2016 Nov 07;18(11):e286 [FREE Full text] [doi: [10.2196/jmir.5698](#)] [Medline: [27821383](#)]
13. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* 2010 Jan;53(1):59-68. [doi: [10.1016/j.bushor.2009.09.003](#)]
14. Organisation for Economic Co-operation and Development. Participative web and user-created content: Web 2.0, wikis and social networking. Organisation for Economic Co-operation and Development. 2007. URL: [https://www.oecd-ilibrary.org/science-and-technology/participative-web-and-user-created-content\\_9789264037472-en](https://www.oecd-ilibrary.org/science-and-technology/participative-web-and-user-created-content_9789264037472-en) [accessed 2020-10-02]
15. NIH's Definition of a Clinical Trial. National Institutes of Health. URL: <https://grants.nih.gov/policy/clinical-trials/definition.htm> [accessed 2019-12-11]
16. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: [10.7326/M18-0850](#)] [Medline: [30178033](#)]
17. Darmawan I, Bakker C, Brockman T, Eder M. The role of social media in enhancing clinical trial recruitment - A scoping review. *Open Science Framework*. 2019. URL: <https://osf.io/hr6sa/> [accessed 2020-10-03]

18. Part 46 - Protection of Human Subjects. Electronic Code of Federal Regulations. 2018 Jul 19. URL: [https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&ptd=20180719&n=pt45.1.46&r=PART&ty=HTML#\\_top](https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&ptd=20180719&n=pt45.1.46&r=PART&ty=HTML#_top) [accessed 2020-10-02]
19. Methodological Expectations of Cochrane Intervention Reviews. Cochrane Methods. URL: <https://methods.cochrane.org/methodological-expectations-cochrane-intervention-reviews> [accessed 2020-10-02]
20. Abbate KJ, Hingle MD, Armin J, Giacobbi P, Gordon JS. Recruiting Women to a Mobile Health Smoking Cessation Trial: Low- and No-Cost Strategies. *JMIR Res Protoc* 2017 Nov 03;6(11):e219 [FREE Full text] [doi: [10.2196/resprot.7356](https://doi.org/10.2196/resprot.7356)] [Medline: [29101091](https://pubmed.ncbi.nlm.nih.gov/29101091/)]
21. Adam LM, Manca DP, Bell RC. Can Facebook Be Used for Research? Experiences Using Facebook to Recruit Pregnant Women for a Randomized Controlled Trial. *J Med Internet Res* 2016 Sep 21;18(9):e250 [FREE Full text] [doi: [10.2196/jmir.6404](https://doi.org/10.2196/jmir.6404)] [Medline: [27655184](https://pubmed.ncbi.nlm.nih.gov/27655184/)]
22. Alley S, Jennings C, Plotnikoff R, Vandelanotte C. An evaluation of web- and print-based methods to attract people to a physical activity intervention. *Journal of Science and Medicine in Sport* 2015 Dec;19:e81 [FREE Full text] [doi: [10.1016/j.jsams.2015.12.196](https://doi.org/10.1016/j.jsams.2015.12.196)]
23. Bracken K, Hague W, Keech A, Conway A, Handelsman DJ, Grossmann M, et al. Recruitment of men to a multi-centre diabetes prevention trial: an evaluation of traditional and online promotional strategies. *Trials* 2019 Jun 19;20(1):366 [FREE Full text] [doi: [10.1186/s13063-019-3485-2](https://doi.org/10.1186/s13063-019-3485-2)] [Medline: [31217024](https://pubmed.ncbi.nlm.nih.gov/31217024/)]
24. Buckingham L, Becher J, Voytek C, Fiore D, Dunbar D, Davis-Vogel A, et al. Going social: Success in online recruitment of men who have sex with men for prevention HIV vaccine research. *Vaccine* 2017 Jun 14;35(27):3498-3505 [FREE Full text] [doi: [10.1016/j.vaccine.2017.05.002](https://doi.org/10.1016/j.vaccine.2017.05.002)] [Medline: [28526330](https://pubmed.ncbi.nlm.nih.gov/28526330/)]
25. Burrell ER, Pines HA, Robbie E, Coleman L, Murphy RD, Hess KL, et al. Use of the location-based social networking application GRINDR as a recruitment tool in rectal microbicide development research. *AIDS Behav* 2012 Oct;16(7):1816-1820 [FREE Full text] [doi: [10.1007/s10461-012-0277-z](https://doi.org/10.1007/s10461-012-0277-z)] [Medline: [22851153](https://pubmed.ncbi.nlm.nih.gov/22851153/)]
26. Cowie JM, Gurney ME. The Use of Facebook Advertising to Recruit Healthy Elderly People for a Clinical Trial: Baseline Metrics. *JMIR Res Protoc* 2018 Jan 24;7(1):e20 [FREE Full text] [doi: [10.2196/resprot.7918](https://doi.org/10.2196/resprot.7918)] [Medline: [29367186](https://pubmed.ncbi.nlm.nih.gov/29367186/)]
27. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. *Nicotine Tob Res* 2014 Feb;16(2):247-251. [doi: [10.1093/ntr/ntt157](https://doi.org/10.1093/ntr/ntt157)] [Medline: [24127266](https://pubmed.ncbi.nlm.nih.gov/24127266/)]
28. Guthrie KA, Caan B, Diem S, Ensrud KE, Greaves SR, Larson JC, et al. Facebook advertising for recruitment of midlife women with bothersome vaginal symptoms: A pilot study. *Clin Trials* 2019 Oct;16(5):476-480. [doi: [10.1177/1740774519846862](https://doi.org/10.1177/1740774519846862)] [Medline: [31055949](https://pubmed.ncbi.nlm.nih.gov/31055949/)]
29. Heffner JL, Wyszynski CM, Comstock B, Mercer LD, Bricker J. Overcoming recruitment challenges of web-based interventions for tobacco use: the case of web-based acceptance and commitment therapy for smoking cessation. *Addict Behav* 2013 Oct;38(10):2473-2476 [FREE Full text] [doi: [10.1016/j.addbeh.2013.05.004](https://doi.org/10.1016/j.addbeh.2013.05.004)] [Medline: [23770645](https://pubmed.ncbi.nlm.nih.gov/23770645/)]
30. Huesch MD, Mukherjee D, Saunders EF. E-recruitment into a bipolar disorder trial using Facebook tailored advertising. *Clin Trials* 2018 Oct;15(5):522-523. [doi: [10.1177/1740774518784018](https://doi.org/10.1177/1740774518784018)] [Medline: [29952243](https://pubmed.ncbi.nlm.nih.gov/29952243/)]
31. Ünlü Ince B, Cuijpers P, van 't Hof E, Riper H. Reaching and recruiting Turkish migrants for a clinical trial through Facebook: A process evaluation. *Internet Interventions* 2014 Apr;1(2):74-83. [doi: [10.1016/j.invent.2014.05.003](https://doi.org/10.1016/j.invent.2014.05.003)]
32. Jones L, Saksvig BI, Grieser M, Young DR. Recruiting adolescent girls into a follow-up study: benefits of using a social networking website. *Contemp Clin Trials* 2012 Mar;33(2):268-272 [FREE Full text] [doi: [10.1016/j.cct.2011.10.011](https://doi.org/10.1016/j.cct.2011.10.011)] [Medline: [22101207](https://pubmed.ncbi.nlm.nih.gov/22101207/)]
33. Jones R, Lacroix LJ, Porcher E. Facebook Advertising to Recruit Young, Urban Women into an HIV Prevention Clinical Trial. *AIDS Behav* 2017 Nov;21(11):3141-3153 [FREE Full text] [doi: [10.1007/s10461-017-1797-3](https://doi.org/10.1007/s10461-017-1797-3)] [Medline: [28528463](https://pubmed.ncbi.nlm.nih.gov/28528463/)]
34. Juraschek SP, Plante TB, Charleston J, Miller ER, Yeh H, Appel LJ, et al. Use of online recruitment strategies in a randomized trial of cancer survivors. *Clin Trials* 2018 Apr;15(2):130-138 [FREE Full text] [doi: [10.1177/1740774517745829](https://doi.org/10.1177/1740774517745829)] [Medline: [29361843](https://pubmed.ncbi.nlm.nih.gov/29361843/)]
35. Kendall JC, French SD, Hartvigsen J, Azari MF. Chiropractic treatment including instrument-assisted manipulation for non-specific dizziness and neck pain in community-dwelling older people: a feasibility randomised sham-controlled trial. *Chiropr Man Therap* 2018;26:14 [FREE Full text] [doi: [10.1186/s12998-018-0183-1](https://doi.org/10.1186/s12998-018-0183-1)] [Medline: [29760878](https://pubmed.ncbi.nlm.nih.gov/29760878/)]
36. Kira A, Glover M, Walker N, Bauld L. Recruiting Pregnant Indigenous Women Who Smoke into a High Contact Incentivized Cessation Trial: A Feasibility Study. *Nicotine Tob Res* 2016 Oct;18(10):2036-2040. [doi: [10.1093/ntr/ntw106](https://doi.org/10.1093/ntr/ntw106)] [Medline: [27154970](https://pubmed.ncbi.nlm.nih.gov/27154970/)]
37. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. *J Consult Clin Psychol* 2017 Mar;85(3):267-273. [doi: [10.1037/ccp0000163](https://doi.org/10.1037/ccp0000163)] [Medline: [28221061](https://pubmed.ncbi.nlm.nih.gov/28221061/)]
38. Nash EL, Gilroy D, Sriksalanukul W, Abhayaratna WP, Stanton T, Mitchell G, et al. Facebook advertising for participant recruitment into a blood pressure clinical trial. *J Hypertens* 2017 Dec;35(12):2527-2531. [doi: [10.1097/HJH.0000000000001477](https://doi.org/10.1097/HJH.0000000000001477)] [Medline: [28704263](https://pubmed.ncbi.nlm.nih.gov/28704263/)]

39. Partridge SR, Balestracci K, Wong AT, Hebden L, McGeechan K, Denney-Wilson E, et al. Effective Strategies to Recruit Young Adults Into the TXT2BFiT mHealth Randomized Controlled Trial for Weight Gain Prevention. *JMIR Res Protoc* 2015 Jun 05;4(2):e66 [FREE Full text] [doi: [10.2196/resprot.4268](https://doi.org/10.2196/resprot.4268)] [Medline: [26048581](https://pubmed.ncbi.nlm.nih.gov/26048581/)]
40. Rabin C, Horowitz S, Marcus B. Recruiting young adult cancer survivors for behavioral research. *J Clin Psychol Med Settings* 2013 Mar;20(1):33-36 [FREE Full text] [doi: [10.1007/s10880-012-9317-0](https://doi.org/10.1007/s10880-012-9317-0)] [Medline: [22810954](https://pubmed.ncbi.nlm.nih.gov/22810954/)]
41. Raviotta JM, Nowalk MP, Lin CJ, Huang H, Zimmerman RK. Using Facebook™ to Recruit College-Age Men for a Human Papillomavirus Vaccine Trial. *Am J Mens Health* 2016 Mar;10(2):110-119 [FREE Full text] [doi: [10.1177/1557988314557563](https://doi.org/10.1177/1557988314557563)] [Medline: [25389213](https://pubmed.ncbi.nlm.nih.gov/25389213/)]
42. Rounds T, Harvey J. Enrollment Challenges: Recruiting Men to Weight Loss Interventions. *Am J Mens Health* 2019;13(1):1557988319832120 [FREE Full text] [doi: [10.1177/1557988319832120](https://doi.org/10.1177/1557988319832120)] [Medline: [30789079](https://pubmed.ncbi.nlm.nih.gov/30789079/)]
43. Sanchez VA, Hall DA, Millar B, Escabi CD, Sharman A, Watson J, et al. Recruiting ENT and Audiology patients into pharmaceutical trials: evaluating the multi-centre experience in the UK and USA. *Int J Audiol* 2018 Sep;57(sup4):S55-S66. [doi: [10.1080/14992027.2018.1425002](https://doi.org/10.1080/14992027.2018.1425002)] [Medline: [29355055](https://pubmed.ncbi.nlm.nih.gov/29355055/)]
44. Shere M, Zhao XY, Koren G. The role of social media in recruiting for clinical trials in pregnancy. *PLoS One* 2014 Mar;9(3):e92744 [FREE Full text] [doi: [10.1371/journal.pone.0092744](https://doi.org/10.1371/journal.pone.0092744)] [Medline: [24671210](https://pubmed.ncbi.nlm.nih.gov/24671210/)]
45. Usadi RS, Diamond MP, Legro RS, Schlaff WD, Hansen KR, Casson P, Reproductive Medicine Network. Recruitment strategies in two reproductive medicine network infertility trials. *Contemp Clin Trials* 2015 Nov;45(Pt B):196-200 [FREE Full text] [doi: [10.1016/j.cct.2015.09.010](https://doi.org/10.1016/j.cct.2015.09.010)] [Medline: [26386293](https://pubmed.ncbi.nlm.nih.gov/26386293/)]
46. Volkova E, Michie J, Corrigan C, Sundborn G, Eyles H, Jiang Y, et al. Effectiveness of recruitment to a smartphone-delivered nutrition intervention in New Zealand: analysis of a randomised controlled trial. *BMJ Open* 2017 Jul 02;7(6):e016198 [FREE Full text] [doi: [10.1136/bmjopen-2017-016198](https://doi.org/10.1136/bmjopen-2017-016198)] [Medline: [28674144](https://pubmed.ncbi.nlm.nih.gov/28674144/)]
47. Waltman NL, Smith KM, Kupzyk KA, Lappe JM, Mack LR, Bilek LD. Approaches to Recruitment of Postmenopausal Women for a Community-Based Study. *Nursing Research* 2019;68(4):307-316. [doi: [10.1097/nnr.0000000000000356](https://doi.org/10.1097/nnr.0000000000000356)]
48. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant Recruitment and Retention in Remote eHealth Intervention Trials: Methods and Lessons Learned From a Large Randomized Controlled Trial of Two Web-Based Smoking Interventions. *J Med Internet Res* 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: [10.2196/10351](https://doi.org/10.2196/10351)] [Medline: [30143479](https://pubmed.ncbi.nlm.nih.gov/30143479/)]
49. Glossary of Terms on Clinical Trials for Patient Engagement Advisory Committee Meeting. Food and Drug Administration. URL: <https://www.fda.gov/media/108378/download> [accessed 2020-10-02]
50. Reuter K. Social Media for Clinical Trial Recruitment: How Real is the Potential? *EMJ Innov* 2020;4(1):34-39 [FREE Full text]
51. Frampton GK, Shepherd J, Pickett K, Griffiths G, Wyatt JC. Digital tools for the recruitment and retention of participants in randomised controlled trials: a systematic map. *Trials* 2020 Jun 05;21(1):478 [FREE Full text] [doi: [10.1186/s13063-020-04358-3](https://doi.org/10.1186/s13063-020-04358-3)] [Medline: [32498690](https://pubmed.ncbi.nlm.nih.gov/32498690/)]

## Abbreviations

**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews

*Edited by G Eysenbach; submitted 31.07.20; peer-reviewed by K Bordenave, J Heffner; comments to author 26.08.20; revised version received 31.08.20; accepted 15.09.20; published 26.10.20.*

### *Please cite as:*

Darmawan I, Bakker C, Brockman TA, Patten CA, Eder M

*The Role of Social Media in Enhancing Clinical Trial Recruitment: Scoping Review*

*J Med Internet Res* 2020;22(10):e22810

URL: <http://www.jmir.org/2020/10/e22810/>

doi: [10.2196/22810](https://doi.org/10.2196/22810)

PMID: [33104015](https://pubmed.ncbi.nlm.nih.gov/33104015/)

©Ida Darmawan, Caitlin Bakker, Tabettha A Brockman, Christi A Patten, Milton Eder. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# The Abortion Web Ecosystem: Cross-Sectional Analysis of Trustworthiness and Bias

Leo Han<sup>1</sup>, MD, MPH; Emily R Boniface<sup>1</sup>, MPH; Lisa Yin Han<sup>2,3</sup>, PhD; Jonathan Albright<sup>4</sup>, PhD; Nora Doty<sup>5</sup>, MD, MPH; Blair G Darney<sup>1,6,7</sup>, PhD, MPH

<sup>1</sup>Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, United States

<sup>2</sup>Department of English, Arizona State University, Tempe, AZ, United States

<sup>3</sup>Department of Film and Media Studies, University of California–Santa Barbara, Santa Barbara, CA, United States

<sup>4</sup>Tow Center for Digital Journalism, Graduate School of Journalism, Columbia University, New York, NY, United States

<sup>5</sup>Jersey Shore University Medical Center, Department of Obstetrics and Gynecology, Neptune, NJ, United States

<sup>6</sup>Center for Population Health Research, National Institute of Public Health, Cuernavaca, Mexico

<sup>7</sup>School of Public Health, Oregon Health and Science University–Portland State University, Portland, OR, United States

**Corresponding Author:**

Lisa Yin Han, PhD

Department of English

Arizona State University

Ross-Blakley Hall

1102 S McAllister Ave

Tempe, AZ, 85281

United States

Phone: 1 720 201 1533

Email: [lyhan@asu.edu](mailto:lyhan@asu.edu)

## Abstract

**Background:** People use the internet as a primary source for learning about medical procedures and their associated safety profiles and risks. Although abortion is one of the most common procedures worldwide among women in their reproductive years, it is controversial and highly politicized. Substantial scientific evidence demonstrates that abortion is safe and does not increase a woman's future risk for depressive disorders or infertility. The extent to which information found on the internet reflects these medical facts in a trustworthy and unbiased manner is not known.

**Objective:** The purpose of this study was to collate and describe the trustworthiness and political slant or bias of web-based information about abortion safety and risks of depression and infertility following abortion.

**Methods:** We performed a cross-sectional study of internet websites using 3 search topics: (1) is abortion safe?, (2) does abortion cause depression?, and (3) does abortion cause infertility? We used the Google Adwords tool to identify the search terms most associated with those topics and Google's search engine to generate databases of websites related to each topic. We then classified and rated each website in terms of content slant (pro-choice, neutral, anti-choice), clarity of slant (obvious, in-between, or difficult/can't tell), trustworthiness (rating scale of 1-5, 5=most trustworthy), type (forum, feature, scholarly article, resource page, news article, blog, or video), and top-level domain (.com, .net, .org, .edu, .gov, or international domain). We compared website characteristics by search topic (safety, depression, or infertility) using bivariate tests. We summarized trustworthiness using the median and IQR, and we used box-and-whisker plots to visually compare trustworthiness by slant and domain type.

**Results:** Our search methods yielded a total of 111, 120, and 85 unique sites for safety, depression, and infertility, respectively. Of all the sites (n=316), 57.3% (181/316) were neutral, 35.4% (112/316) were anti-choice, and 7.3% (23/316) were pro-choice. The median trustworthiness score was 2.7 (IQR 1.7-3.7), which did not differ significantly across topics ( $P=.409$ ). Anti-choice sites were less trustworthy (median score 1.3, IQR 1.0-1.7) than neutral (median score 3.3, IQR 2.7-4.0) and pro-choice (median score 3.7, IQR 3.3-4.3) sites. Anti-choice sites were also more likely to have slant clarity that was "difficult to tell" (41/112, 36.6%) compared with neutral (25/181, 13.8%) or pro-choice (4/23, 17.4%;  $P<.001$ ) sites. A negative search term used for the topic of safety (eg, "risks") produced sites with lower trustworthiness scores than search terms with the word "safety" (median score 1.7 versus 3.7, respectively;  $P<.001$ ).



**Conclusions:** People seeking information about the safety and potential risks of abortion are likely to encounter a substantial amount of untrustworthy and slanted/biased abortion information. Anti-choice sites are prevalent, often difficult to identify as anti-choice, and less trustworthy than neutral or pro-choice sites. Web searches may lead the public to believe abortion is riskier than it is.

(*J Med Internet Res* 2020;22(10):e20619) doi:[10.2196/20619](https://doi.org/10.2196/20619)

## KEYWORDS

internet; abortion; media; websites; infodemiology; infodemic; quality of health information; bias in patient education

## Introduction

The internet is the first source the public turns to for medical information [1-3]. At least 70% of adults who use the internet use it for health information research, with 43% of them seeking information about specific treatments or procedures [4]. Although the internet is a vast repository of searchable information, there is often incorrect, deliberately misleading, conflicting, and/or hard to understand information [5]. Furthermore, public perception and knowledge is shaped by personalized internet experience, which is delimited by technology company algorithms [6-8]. Abortion is one of the most common medical procedures in the world. According to recent estimates, over 800,000 women in the United States choose an abortion each year [9]; even more consider abortion but do not obtain one [10]. Surgical procedures for abortion are referred to as aspiration, dilation, and curettage (D&C), or dilation and evacuation (D&E) if done later in pregnancy. For abortions in the first trimester of pregnancy (up to 10 weeks), medications are used [11]. The best scientific evidence clearly demonstrates that induced abortion is safe [12,13] and that abortion does not increase a woman's future risk for disorders such as depression, anxiety, or suicidality [14], or secondary infertility [15,16]. Only low-quality and/or discredited studies suggest otherwise [17,18]. The risks of abortion do increase with the gestational age of the pregnancy, but abortion at any gestational age is safer than childbirth [12].

Despite these facts, many antiabortion arguments rest on misinformation regarding the safety and health consequences of abortion [19]. Crisis pregnancy centers (CPCs)—organizations that try to intercept women who are considering an abortion—often describe abortion as dangerous or deadly in order to dissuade women from choosing to obtain an abortion [20]. In addition, false or misleading information on the internet about abortion continues to exert an influence on public debates and policy [21]. Previous studies have examined web-based abortion information, focusing on the quality of information available for self-referral (on CPC websites, specifically) and about D&E procedures [20,22,23]. These studies found frequent inaccuracies and a lack of comprehensive information about abortion on the internet. However, the studies were limited by small data samples [23,24] or narrow inquiries on specific websites [10]. It remains unclear exactly what a search about abortion safety or potential subsequent risks would yield in terms of websites, as well as the trustworthiness and accuracy of information contained on them. The purpose of this study was to collate and describe the web-based ecosystem about abortion in terms of the

trustworthiness and bias about abortion safety and risks. We focused on websites that provided information related to three questions: (1) is abortion safe?, (2) does abortion cause depression?, and (3) does abortion cause infertility? We chose safety, infertility, and depression for our case studies because the existing medical evidence is very clear; the mainstream scientific consensus is that abortion is safe and does not cause subsequent infertility or depression [25]. However, these questions continue to be disputed in public discourse and provide an opportunity to evaluate the quality of website information.

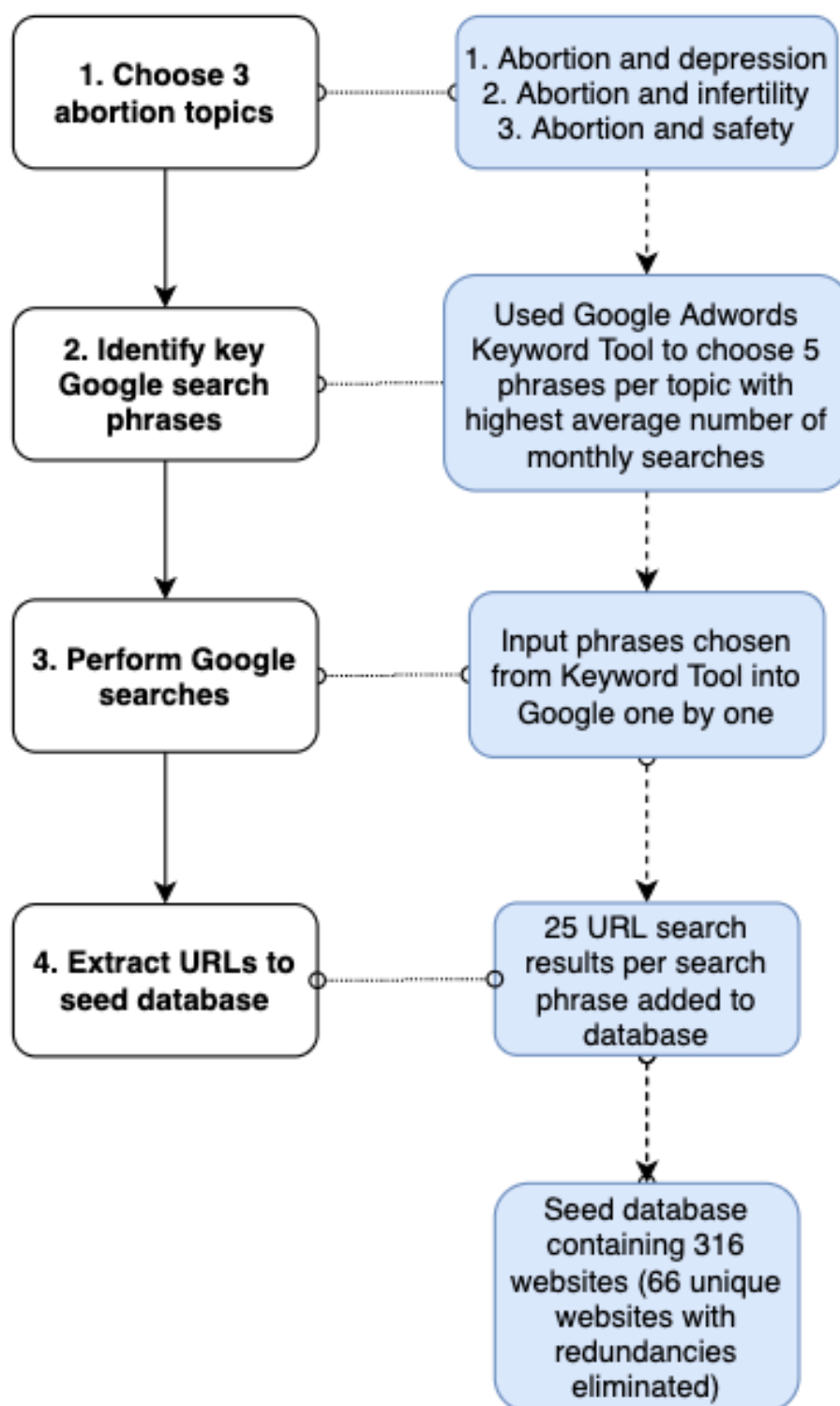
## Methods

This cross-sectional study had two phases: first, we created a database of websites from searches about abortion and safety, depression, and infertility; second, we used the ecosystem database that we created to classify websites in terms of content trustworthiness, political slant on abortion, type of site, and top-level domain.

### Website Database

We utilized several steps to create our website database in order to minimize bias and most broadly capture what the general public would encounter while searching on the web (Figure 1). First, we used Google's Adwords tool, a free Google advertising tool that allows advertisers to associate their ads with search terms based on the popularity of the term. For a given topic, the keyword tool identifies the most commonly used search terms with a topic based on historical search data [26]. We selected 5 search terms for each of our 3 keyword combinations (abortion safety, abortion depression, and abortion infertility) based on their frequency as identified by Google Adwords and relevance (Figure 2). Next, we conducted Google searches using these phrases on August 31, 2018, and compiled website results for each of our search terms. In order to remove search term personalization, each search was performed using a cookie-cleared Google Chrome (version 10.8) browser in incognito mode combined with a free virtual private network called Windscribe (version 1.82) in order to reduce locational bias as well as anonymize our internet protocol addresses. For each search phrase, we extracted the top 25 page-ranked results, noting multiple occurrences. Each link was then imported into a spreadsheet using MozBar (Moz Inc), a Chrome extension that also produced titles, brief excerpts, and a search position number for each link based on its proximity to the top of the Google page. We produced a database for each of our topics (safety, depression, infertility) based on this search methodology.



**Figure 1.** Workflow of website selection for the analysis.

[illegible]

We next classified and rated each website in the database. We rated websites on 5 metrics: (1) slant, (2) slant clarity, (3) trustworthiness, (4) type, and (5) domain. For slant, a website was determined to have a pro-choice or anti-choice bias if either (a) information was given in a biased or dramatic fashion (eg, an anti-choice site describing the procedure as “tearing the baby from the womb”), or (b) the website displayed an opinion regarding the provision of abortion and its legality (eg, a pro-choice site stating that abortion should be free and legal for all women); otherwise, the website was considered neutral. We assigned a slant clarity rating based on how easy it was to discern the website’s slant (obvious, in-between, or difficult/can’t tell). We scored trustworthiness on a rating scale from 1 to 5, with 5 being most trustworthy based on factors

<http://www.jmir.org/2020/10/e20619/>

## Data Analysis

First, we created a visualization of the websites and their parent search terms and topics using RAWGraphs. Next, we compared website characteristics overall and by search topic (safety, depression, or infertility) using Pearson chi-square test or Fisher exact test for all categorical classifications to compare characteristics across topics. We summarized trustworthiness, measured on a 5-point scale, using the median and IQR. For each search topic, we also assessed the distribution of websites based on the content's slant for the 5 query terms, and then compared the trustworthiness of query term results using Kruskal-Wallis and Wilcoxon rank sum tests. We used box-and-whisker plots to visually compare trustworthiness by slant and by domain type. Finally, we tabulated websites according to their slant and slant clarity, comparing distributions using Fisher exact test, and then used a bar graph to examine website counts graphically. The study was reviewed and approved by the Oregon Health and Science University's institutional review board and was deemed not human subjects research.

## Results

After removing duplicates, our search methods yielded a total of 111, 120, and 85 unique sites for safety, depression, and infertility, respectively. [Figure 2](#) provides a visualization of these websites and their parent search terms and topic (see [Multimedia Appendix 1](#) for full list of sites with URLs). While the majority of sites had a neutral slant (181/316, 57.3%) overall, the slant distribution differed by topic, with a higher proportion of sites about safety and depression having an anti-choice slant ([Table 1](#)). Approximately 40% of sites about safety and depression had an anti-choice slant (safety: 44/111, 39.6%;

depression: 51/120, 42.5%), compared with 15.3% (17/111) and 3.3% (4/120) of the safety and depression sites, respectively, being categorized as having a pro-choice slant. Infertility had the highest proportion of neutral sites (66/85, 77.7%). In terms of website type, we categorized the majority of websites as resource pages (158/316, 50.8%). Commercial (.com/.net) (154/316, 48.7%) and organization (.org) (101/316, 32.0%) sites accounted for the majority of domain types. While we did not specifically perform an analysis of slant by web address (URL), we noticed that many of the anti-choice sites came from generic-appearing addresses (eg, americanpregnancy.org, adviceandaid.com). One state health department site (Alaska) was rated by our researchers as being anti-choice.

We found that the median score for trustworthiness of the websites was 2.7 (IQR 1.7-3.7), which did not differ significantly across topics (safety: median score 3.0, IQR 1.7-3.7; depression: median score 2.3, IQR 1.7-3.7; infertility: median score 2.7, IQR 1.7-3.7;  $P=.409$ ). Agreement in trustworthiness scores among our coders was 71.8%, 70.0%, and 80.9% for safety, depression, and infertility, respectively. Of the 316 sites in our sample, only 59 (18.7%) sites received median trustworthiness scores of 4 or more (data not shown). The trustworthiness rating of sites was different by slant and by domain type ([Figure 3](#)). Overall, anti-choice sites had lower trustworthiness scores (median score 1.3, IQR 1.0-1.7) compared with neutral (median score 3.3, IQR 2.7-4.0) and pro-choice (median score 3.7, IQR 3.3-4.3) sites. Sites that came from educational institutions (.edu) were consistently given higher trustworthiness scores (median score 4.3, IQR 3.7-4.3), while other domains received overall lower and more widely distributed trustworthiness scores, with commercial sites (.com/.net) considered to be the least trustworthy (median score 2.3, IQR 1.3-3.3).

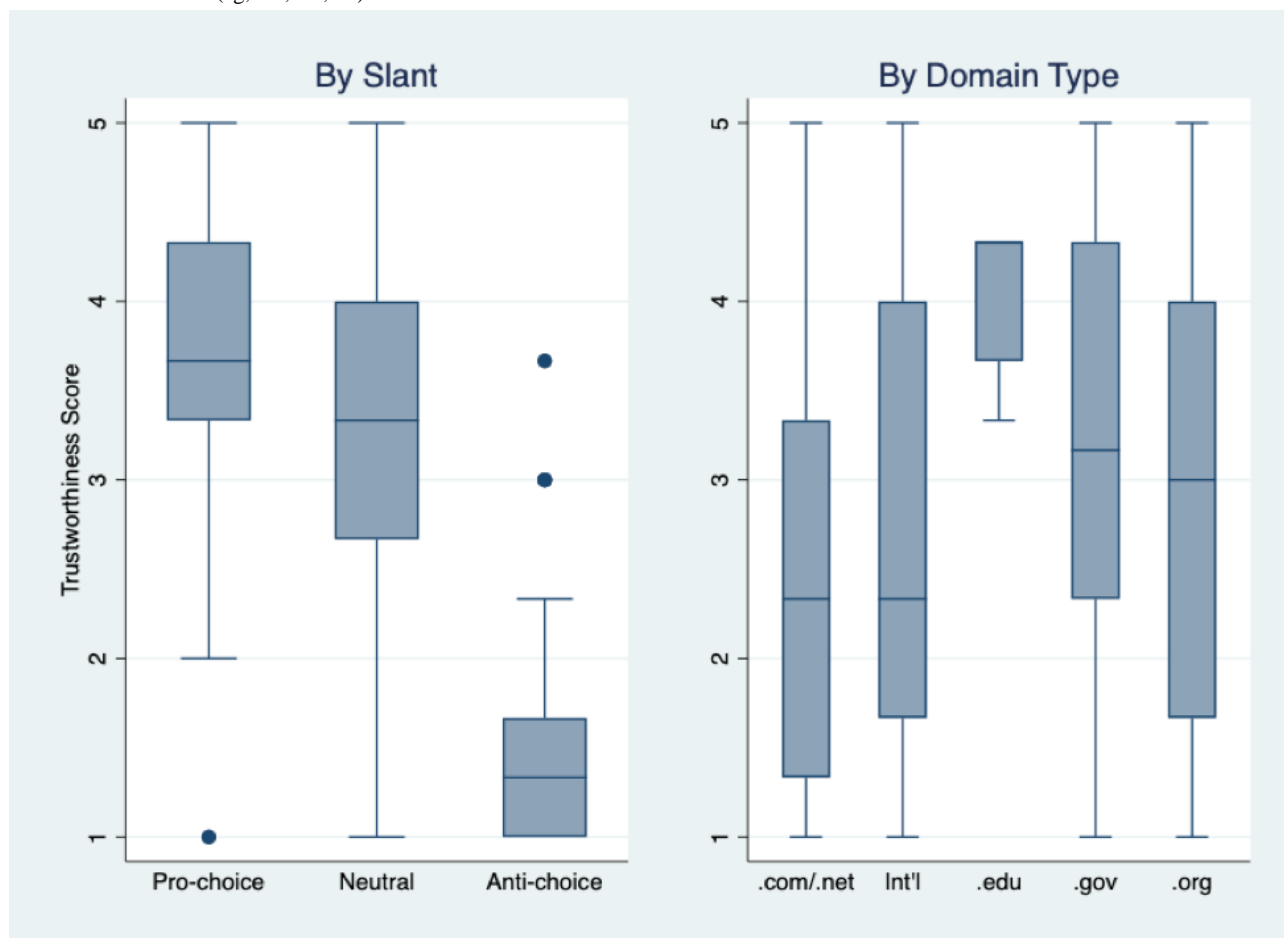
**Table 1.** Website characteristics by search topic (n=316).

Characteristic	Overall	Safety (n=111) <sup>a</sup>	Depression (n=120)	Infertility (n=85)	P value
<b>Slant, n (%)</b>					<.001
Pro-choice	23 (7.3)	17 (15.3)	4 (3.3)	2 (2.4)	
Neutral	181 (57.3)	50 (45.1)	65 (54.2)	66 (77.7)	
Anti-choice	112 (35.4)	44 (39.6)	51 (42.5)	17 (20.0)	
<b>Slant clarity, n (%)</b>					.002
Obvious	121 (38.3)	41 (36.9)	49 (40.8)	31 (36.5)	
In-between	125 (39.6)	49 (44.1)	33 (27.5)	43 (50.6)	
Difficult/can't tell	70 (22.1)	21 (18.9)	38 (31.7)	11 (12.9)	
<b>Domain, n (%)</b>					.002
Commercial (.com/.net)	154 (48.7)	45 (40.5)	60 (50.0)	49 (57.7)	
International	42 (13.3)	8 (7.2)	17 (14.2)	17 (20.0)	
Education (.edu)	5 (1.6)	2 (1.8)	1 (0.8)	2 (2.4)	
Government (.gov)	14 (4.4)	8 (7.2)	4 (3.3)	2 (2.4)	
Organization (.org)	101 (32.0)	48 (43.2)	38 (31.7)	15 (17.7)	
<b>Page type, n (%)</b>					.007
News article	30 (9.7)	12 (11.3)	13 (10.8)	5 (5.9)	
Forum	17 (5.5)	2 (1.9)	2 (1.7)	13 (15.3)	
Blog	16 (5.1)	5 (4.7)	8 (6.7)	3 (3.5)	
Feature	63 (20.3)	21 (19.8)	20 (16.7)	22 (25.9)	
Scholarly article	25 (8.0)	9 (8.5)	11 (9.2)	5 (5.9)	
Resource page	158 (50.8)	57 (53.8)	64 (53.3)	37 (43.5)	
Video	2 (0.6)	0 (0)	2 (1.7)	0 (0)	
Trustworthiness, median (IQR)	2.7 (1.7-3.7)	3.0 (1.7-3.7)	2.3 (1.7-3.7)	2.7 (1.7-3.7)	.409

<sup>a</sup>Five websites in the safety category were missing data about page type.



**Figure 3.** Box-and-whisker plots comparing trustworthiness of online abortion resources by slant and domain type (n=316). Trustworthiness was scored on a scale of 1 (least trustworthy) to 5 (most trustworthy). The central line in each box marks the median trustworthiness score; upper and lower box edges mark the 75th and 25th percentile, respectively; whiskers indicate 150% of the interquartile range and outliers are shown as individual points. Int'l: international domain (eg, .uk, .nz, .id).



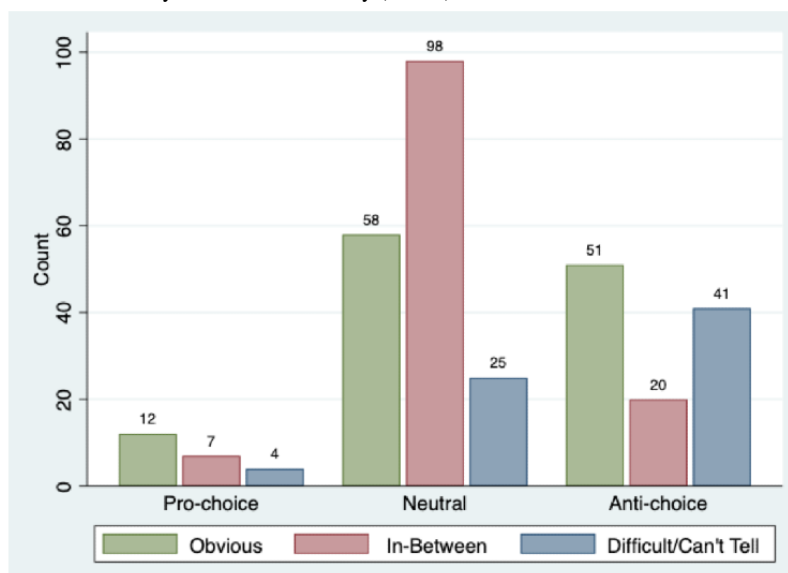
We also examined our results using specific search query terms and noted that for searches on safety and infertility, trustworthiness scores were significantly different across query search terms (Table 2). For searches focused on abortion safety, including the word “safe” resulted in sites with significantly higher trustworthiness scores (median score 3.7, IQR 3.3-4.3) compared with searches that included the more negative word “risks” (median 1.7, IQR 1.0-3.3; Wilcoxon rank sum  $P<.001$ ). Anti-choice sites appeared more frequently with negative search terms, such as “risks of abortion” (17/23, 73.9%), compared with when the word “safe” was included in the search term, such as, “How safe is abortion?” (1/27, 3.7%) (Table 2).

Finally, we found that slant clarity, or how difficult it was to discern the anti-choice, neutral, or pro-choice slant of a website, varied significantly both by topic and by slant. In regard to the topic, there were many more depression websites with a “difficult/can’t tell” slant compared with sites about safety and infertility (31.7% versus 18.9% and 12.9%, respectively;  $P=.002$ ). However, all 3 topic categories contained clarity ratings of “obvious” for fewer than one-half of the sites (Table 1). In terms of slant, anti-choice sites were more likely to have slant clarity that was “difficult/can’t tell” (41/112, 36.6%) compared with neutral (25/181, 13.8%) or pro-choice (4/23, 17.4%) sites ( $P<.001$ ) (Figure 4).



**Table 2.** Search query terms used and associated trustworthiness and website slant for each search topic.

Topic and search query terms	Total, n (%)	Pro-choice, n (%)	Neutral, n (%)	Anti-choice, n (%)	Trustworthiness, median (IQR)	P value
<b>Safety (n=111)</b>						<.001
Abortion health risks	19 (17.1)	1 (5.3)	7 (36.8)	11 (57.9)	2.3 (1.7-3.3)	
Effects of abortion	24 (21.6)	0 (0)	10 (41.7)	14 (58.3)	1.7 (1.0-3.0)	
How safe is abortion?	27 (24.4)	5 (18.5)	21 (77.8)	1 (3.7)	3.7 (3.0-4.3)	
Risks of abortion	23 (20.7)	0 (0)	6 (26.1)	17 (73.9)	1.3 (1.0-2.3)	
Safe abortion	18 (16.2)	11 (61.1)	6 (33.3)	1 (5.6)	3.7 (3.7-4.3)	
<b>Depression (n=120)</b>						.143
Abortion depression	28 (23.3)	0 (0)	17 (60.7)	11 (39.3)	3.2 (1.7-3.7)	
Abortion recovery	29 (24.2)	0 (0)	14 (48.3)	15 (51.7)	2.3 (1.7-3.7)	
Coping with abortion	26 (21.7)	0 (0)	21 (80.8)	5 (19.2)	3.2 (1.7-3.7)	
Post-abortion depression	10 (8.3)	0 (0)	6 (60.0)	4 (40.0)	2.2 (1.3-3.3)	
Post-abortion syndrome	27 (22.5)	4 (14.8)	7 (25.9)	16 (59.3)	1.7 (1.3-3.0)	
<b>Infertility (n=85)</b>						.005
Abortion pill infertility	19 (22.4)	1 (5.3)	15 (78.9)	3 (15.8)	3.7 (2.0-4.3)	
Can abortion cause infertility?	11 (12.9)	0 (0)	8 (72.7)	3 (27.3)	2.0 (1.7-3.3)	
Can multiple abortions cause infertility?	15 (17.6)	0 (0)	11 (73.3)	4 (26.7)	1.7 (1.0-2.7)	
Infertility after abortion	17 (20.0)	1 (5.9)	10 (58.8)	6 (35.3)	2.3 (1.0-3.3)	
Pregnancy after abortion	23 (27.1)	0 (0)	22 (95.7)	1 (4.3)	3.0 (2.3-3.7)	

**Figure 4.** Counts of online abortion resources by slant and slant clarity (n=316).

## Discussion

### Principal Findings

Our study of websites that provide information about safety of abortion or subsequent risks of depression or infertility found that overall, 35.4% (112/316) of sites had an anti-choice slant. Anti-choice sites were less trustworthy, and the clarity of their stance was also more difficult to determine. Our data also show that people seeking information about the safety and potential

risks of abortion are likely to find substantial amounts of untrustworthy abortion information; the overall median trustworthiness score of these sites was 2.7 (on a scale of 1-5), with a wide range (IQR 1.7-3.7).

Our results are consistent with other examples of intentionally disseminated abortion disinformation. CPCs have an antiabortion agenda but use neutral language and advertising to appear as if they are a normal medical clinic [20]. In our study, almost two-thirds (71/112, 63.4%) of anti-choice sites were categorized

as resource pages and many derived from web addresses that appeared generic and neutral. Moreover, one state health department site rated as anti-choice, which demonstrates that even official governmental sites can be biased sources of information. These findings suggest that anti-choice sites may be trying to obscure their biases or appear as neutral information brokers. Based on the high number of “difficult to tell” sites in our data sets, we need more research to understand the characteristics of anti-choice sites that create this effect.

We know that in the case of abortion, false or misleading information continues to play a role in public debates and government legislation. For example, Targeted Regulation of Abortion Providers (TRAP) laws, or costly and burdensome regulations aiming to restrict access to abortions by dictating how, by whom, and when abortions can be provided, are often based on false health claims but framed as protecting patient safety [31]. The internet is one of the most important propagators of false information. Previous research has shown that media is more influential among young people (aged 13 to 29 years) than friends, family, and health care providers when it comes to learning about abortion [32]. However, there is also evidence that exposing the public to more evidence-based information on abortion can change opinions about the provision and regulation of abortion services [33]. Thus, efforts to address the spread of false information and strengthen the presence of evidence-based information on abortion can have real impact. Our results can help inform clinicians and others who work in the abortion field about likely information gaps [34]. Moreover, health organizations that seek to disseminate evidence-based information may want to conduct regular search audits such as this in order to optimize their search positions to reach a broader audience.

### Limitations

Our study results should be interpreted with the following limitations in mind. We used expert ratings to gain insight into the abortion online media ecosystem. All of our experts are engaged in abortion-related research and two-thirds of our

experts are abortion providers. We recognize that non-experts may have rated and classified these same websites differently. However, our expert ratings provide a benchmark for future work that explores how non-experts perceive the information contained in these websites. We also acknowledge that trustworthiness scores could be confounded with slant and that sites with obvious anti-choice slants may have been rated with lower trustworthiness scores based on the impression of bias. However, it is difficult to blind these components from each other and we wanted to provide some kind of measure of quality for the information we encountered. We recognize that our study provides only a cross-sectional snapshot of the internet. While specific site rankings may be driven by news cycles, we believe the overall picture is unlikely to change [35]. We also acknowledge that while we took measures to anonymize the search results, individuals will receive different results based on their previous search history, and the majority will not scroll past the first page. However, our database sizes captured approximately the first 8 to 12 pages of search results for a given topic. Finally, while our focus was on websites, websites alone do not fully address the digital means through which abortion information is shared. In particular, social networking platforms are critical sources [36]. The strengths of our study include the use of a systematic and rigorous approach to create our website database, incorporation of multiple topic areas, and focus on content trustworthiness and bias.

Our results provide insight into the online abortion ecosystem. We find that anti-choice sites are prevalent, often hard to identify as anti-choice (difficult slant clarity), and less trustworthy than neutral and pro-choice sites. Additionally, the search terms a user chooses may play a substantial role in the quality and bias of websites they see. Our results help us understand how the internet may impact public perceptions and knowledge about the safety of abortion and potential risks of depression and infertility. Our findings suggest that web searches may lead people to perceive abortion procedures as more dangerous and riskier than they actually are.

---

### Acknowledgments

This study was funded by The Society for Family Planning.

---

### Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Full list of websites and site ratings.

[[XLSX File \(Microsoft Excel File\), 87 KB - jmir\\_v22i10e20619\\_app1.xlsx](#)]

---

### References

1. Rupert DJ, Gard Read J, Amoozegar JB, Moultrie RR, Taylor OM, O'Donoghue AC, et al. Peer-Generated Health Information: The Role of Online Communities in Patient and Caregiver Health Decisions. *J Health Commun* 2016 Nov;21(11):1187-1197 [[FREE Full text](#)] [doi: [10.1080/10810730.2016.1237592](https://doi.org/10.1080/10810730.2016.1237592)] [Medline: [27805496](https://pubmed.ncbi.nlm.nih.gov/27805496/)]
2. Ziebland S, Powell J, Briggs P, Jenkinson C, Wyke S, Sillence E, et al. Examining the role of patients' experiences as a resource for choice and decision-making in health care: a creative, interdisciplinary mixed-method study in digital health. *Programme Grants Appl Res* 2016 Nov;4(17):1-214. [doi: [10.3310/pgfar04170](https://doi.org/10.3310/pgfar04170)]

3. McMullan M. Patients using the Internet to obtain health information: how this affects the patient-health professional relationship. *Patient Educ Couns* 2006 Oct;63(1-2):24-28. [doi: [10.1016/j.pec.2005.10.006](https://doi.org/10.1016/j.pec.2005.10.006)] [Medline: [16406474](https://pubmed.ncbi.nlm.nih.gov/16406474/)]
4. Fox S, Duggan M. Health Online 2013 Internet. Pew Res Cent Internet Sci Tech. Washington, D.C.; 2013 Jan 15. URL: <http://www.pewinternet.org/2013/01/15/health-online-2013/> [accessed 2020-04-21]
5. Benigeri M. Shortcomings of health information on the Internet. *Health Promotion International* 2003 Dec 01;18(4):381-386. [doi: [10.1093/heapro/dag409](https://doi.org/10.1093/heapro/dag409)]
6. Bessi A, Ferrara E. Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday* 2016 Nov 03;NA-NA [FREE Full text] [doi: [10.5210/fm.v21i11.7090](https://doi.org/10.5210/fm.v21i11.7090)]
7. Gottfried J, Shearer E. News Use Across Social Media Platforms 2016 Internet. Pew Res Cent Journal Proj. 2016 May 26. URL: <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/> [accessed 2020-04-21]
8. Christen CT, Huberty KE. Media Reach, Media Influence? The Effects of Local, National, and Internet News on Public Opinion Inferences. 2016 Jun 25;84(2):315-334. [doi: [10.1177/107769900708400208](https://doi.org/10.1177/107769900708400208)]
9. Jones RK, Witwer E, Jerman J. Abortion Incidence and Service Availability in the United States, 2017. Guttmacher Institute. New York: Guttmacher Institute; 2019 Sep. URL: <https://www.guttmacher.org/report/abortion-incidence-service-availability-us-2017> [accessed 2020-04-21]
10. Roberts SCM, Johns NE, Williams V, Wingo E, Upadhyay UD. Estimating the proportion of Medicaid-eligible pregnant women in Louisiana who do not get abortions when Medicaid does not cover abortion. *BMC Women's Health* 2019 Jun 19;19(1):78. [doi: [10.1186/s12905-019-0775-5](https://doi.org/10.1186/s12905-019-0775-5)]
11. Committee on Health Care for Underserved Women. ACOG Committee Opinion No. 613: Increasing access to abortion. *Obstet Gynecol* 2014 Nov;124(5):1060-1065. [doi: [10.1097/01.AOG.0000456326.88857.31](https://doi.org/10.1097/01.AOG.0000456326.88857.31)] [Medline: [25437742](https://pubmed.ncbi.nlm.nih.gov/25437742/)]
12. Raymond EG, Grimes DA. The Comparative Safety of Legal Induced Abortion and Childbirth in the United States. 2012;119(2, Part 1):215-219. [doi: [10.1097/aog.0b013e31823fe923](https://doi.org/10.1097/aog.0b013e31823fe923)]
13. Jatlaoui TC, Ewing A, Mandel MG, Simmons KB, Suchdev DB, Jamieson DJ, et al. Abortion Surveillance — United States, 2013. *MMWR Surveill. Summ* 2016 Nov 25;65(12):1-44. [doi: [10.15585/mmwr.ss6512a1](https://doi.org/10.15585/mmwr.ss6512a1)]
14. Steinberg JR, McCulloch CE, Adler NE. Abortion and mental health: findings from The National Comorbidity Survey-Replication. *Obstet Gynecol* 2014 Feb;123(2 Pt 1):263-270 [FREE Full text] [doi: [10.1097/AOG.0000000000000092](https://doi.org/10.1097/AOG.0000000000000092)] [Medline: [24402590](https://pubmed.ncbi.nlm.nih.gov/24402590/)]
15. Frank P, McNamee R, Hannaford PC, Kay CR, Hirsch S. The effect of induced abortion on subsequent fertility. *Br J Obstet Gynaecol* 1993 Jun;100(6):575-580. [doi: [10.1111/j.1471-0528.1993.tb15313.x](https://doi.org/10.1111/j.1471-0528.1993.tb15313.x)] [Medline: [8334095](https://pubmed.ncbi.nlm.nih.gov/8334095/)]
16. Lurie S, Levy R, Katz Z, Appelman Z, Insler V. The influence of midtrimester termination of pregnancy on subsequent fertility: four to five years follow-up. *Contraception* 1994 Sep;50(3):239-241. [doi: [10.1016/0010-7824\(94\)90069-8](https://doi.org/10.1016/0010-7824(94)90069-8)]
17. Coleman PK. Abortion and mental health: quantitative synthesis and analysis of research published 1995-2009. *Br J Psychiatry* 2011 Sep;199(3):180-186. [doi: [10.1192/bjp.bp.110.077230](https://doi.org/10.1192/bjp.bp.110.077230)] [Medline: [21881096](https://pubmed.ncbi.nlm.nih.gov/21881096/)]
18. Koch E, Chireau M, Pliego F, Stanford J, Haddad S, Calhoun B, et al. Abortion legislation, maternal healthcare, fertility, female literacy, sanitation, violence against women and maternal deaths: a natural experiment in 32 Mexican states. *BMJ Open* 2015 Feb 23;5(2):e006013-e006013. [doi: [10.1136/bmjopen-2014-006013](https://doi.org/10.1136/bmjopen-2014-006013)]
19. AbortionFacts.com. URL: <http://www.abortionfacts.com/> [accessed 2018-01-05]
20. Bryant AG, Narasimhan S, Bryant-Comstock K, Levi EE. Crisis pregnancy center websites: Information, misinformation and disinformation. *Contraception* 2014 Dec;90(6):601-605. [doi: [10.1016/j.contraception.2014.07.003](https://doi.org/10.1016/j.contraception.2014.07.003)]
21. Conti JA, Cahill E. Abortion in the media. *Curr Opin Obstet Gynecol* 2017 Dec;29(6):427-430. [doi: [10.1097/GCO.0000000000000412](https://doi.org/10.1097/GCO.0000000000000412)] [Medline: [28915157](https://pubmed.ncbi.nlm.nih.gov/28915157/)]
22. Dodge LE, Phillips SJ, Neo DT, Nippita S, Paul ME, Hacker MR. Quality of Information Available Online for Abortion Self-Referral. 2018;132(6):1443-1452. [doi: [10.1097/aog.0000000000002950](https://doi.org/10.1097/aog.0000000000002950)]
23. Carlsson T, Axelsson O. Patient Information Websites About Medically Induced Second-Trimester Abortions: A Descriptive Study of Quality, Suitability, and Issues. *J Med Internet Res* 2017 Jan 10;19(1):e8. [doi: [10.2196/jmir.6380](https://doi.org/10.2196/jmir.6380)]
24. Han L, Rodriguez M, Han L. Tweeting PP: an analysis of the 2015–2016 Planned Parenthood controversy on Twitter. *Contraception* 2016 Oct;94(4):399. [doi: [10.1016/j.contraception.2016.07.064](https://doi.org/10.1016/j.contraception.2016.07.064)]
25. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Care Services, Board on Population Health and Public Health Practice, Committee on Reproductive Health Services: Assessing the Safety and Quality of Abortion Care in the U.S.. *The Safety and Quality of Abortion Care in the United States*. Washington, DC: National Academies Press; Mar 16, 2018:1-187.
26. Use Keyword Planner. Google Ads Help. URL: <https://support.google.com/google-ads/answer/7337243?hl=en> [accessed 2019-04-09]
27. Ahmad R, Komlodi A, Wang J, Hercegf K. The impact of user experience levels on web credibility judgments. *Proc. Am. Soc. Info. Sci. Tech* 2011 Feb 03;47(1):1-4. [doi: [10.1002/meet.14504701180](https://doi.org/10.1002/meet.14504701180)]
28. Song H, Omori K, Kim J, Tenzek KE, Morey HJ, Lin W, et al. Trusting Social Media as a Source of Health Information: Online Surveys Comparing the United States, Korea, and Hong Kong. *J Med Internet Res* 2016 Mar 14;18(3):e25 [FREE Full text] [doi: [10.2196/jmir.4193](https://doi.org/10.2196/jmir.4193)] [Medline: [26976273](https://pubmed.ncbi.nlm.nih.gov/26976273/)]

29. Pereira JL, Koski S, Hanson J, Bruera ED, Mackey JR. Internet Usage Among Women with Breast Cancer: An Exploratory Study. *Clinical Breast Cancer* 2000 Jul;1(2):148-153. [doi: [10.3816/cbc.2000.n.013](https://doi.org/10.3816/cbc.2000.n.013)]
30. Zimdars M. False, Misleading, Clickbait-y, and/or Satirical “News” Sources. Google document. 2016 Nov 17. URL: [https://docs.google.com/document/d/10eA5-mCZLSS4MQY5QG5ewC3VAL6pLkT53V\\_81ZyitM/preview](https://docs.google.com/document/d/10eA5-mCZLSS4MQY5QG5ewC3VAL6pLkT53V_81ZyitM/preview) [accessed 2020-04-21]
31. Targeted Regulation of Abortion Providers. Guttmacher Institute. Washington, D.C: Guttmacher Institute; 2020 Oct 01. URL: <https://www.guttmacher.org/state-policy/explore/targeted-regulation-abortion-providers> [accessed 2020-10-01]
32. Altshuler AL, Gerns Storey HL, Prager SW. Exploring abortion attitudes of US adolescents and young adults using social media. *Contraception* 2015 Mar;91(3):226-233. [doi: [10.1016/j.contraception.2014.11.009](https://doi.org/10.1016/j.contraception.2014.11.009)] [Medline: [25537853](https://pubmed.ncbi.nlm.nih.gov/25537853/)]
33. White K, Grossman D, Stevenson AJ, Hopkins K, Potter JE. Does information about abortion safety affect Texas voters' opinions about restrictive laws? A randomized study. *Contraception* 2017 Dec;96(6):381-387. [doi: [10.1016/j.contraception.2017.08.007](https://doi.org/10.1016/j.contraception.2017.08.007)]
34. Tan SS, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. *J Med Internet Res* 2017 Jan 19;19(1):e9 [FREE Full text] [doi: [10.2196/jmir.5729](https://doi.org/10.2196/jmir.5729)] [Medline: [28104579](https://pubmed.ncbi.nlm.nih.gov/28104579/)]
35. Vaughan L. New measurements for search engine evaluation proposed and tested. 2004 Jul;40(4):677-691. [doi: [10.1016/s0306-4573\(03\)00043-8](https://doi.org/10.1016/s0306-4573(03)00043-8)]
36. Majchrzak A, Faraj S, Kane GC, Azad B. The Contradictory Influence of Social Media Affordances on Online Communal Knowledge Sharing. *J Comput-Mediat Comm* 2013 Oct 18;19(1):38-55. [doi: [10.1111/jcc4.12030](https://doi.org/10.1111/jcc4.12030)]

## Abbreviations

**CPCs:** crisis pregnancy centers

**D&C:** aspiration, dilation, and curettage

**D&E:** dilation and evacuation

**TRAP:** Targeted Regulation of Abortion Providers

*Edited by G Eysenbach; submitted 24.05.20; peer-reviewed by Q Wu, M Jordan-Marsh; comments to author 11.08.20; revised version received 28.08.20; accepted 22.09.20; published 26.10.20.*

*Please cite as:*

Han L, Boniface ER, Han LY, Albright J, Doty N, Darney BG

*The Abortion Web Ecosystem: Cross-Sectional Analysis of Trustworthiness and Bias*

*J Med Internet Res* 2020;22(10):e20619

URL: <http://www.jmir.org/2020/10/e20619/>

doi: [10.2196/20619](https://doi.org/10.2196/20619)

PMID: [33104002](https://pubmed.ncbi.nlm.nih.gov/33104002/)

©Leo Han, Emily R Boniface, Lisa Yin Han, Jonathan Albright, Nora Doty, Blair G Darney. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Parental Online Information Access and Childhood Vaccination Decisions in North America: Scoping Review

Sarah Ashfield<sup>1\*</sup>, MSCN; Lorie Donelle<sup>1\*</sup>, PhD

Western University, London, ON, Canada

\* all authors contributed equally

**Corresponding Author:**

Lorie Donelle, PhD

Western University

FNB Room 2356 Faculty of Health Science

1115 Richmond Street

London, ON, N6A 5B9

Canada

Phone: 1 519 661 2111 ext 86565

Email: [ldonelle@uwo.ca](mailto:ldonelle@uwo.ca)

## Abstract

**Background:** Immunizing children throughout their early years prevents the spread of communicable disease and decreases the morbidity and mortality associated with many vaccine-preventable diseases. Searching online allows individuals rapid access to health information.

**Objective:** The purpose of this review was to develop an understanding of the existing literature of parents' online health information-seeking behaviors to inform their vaccination choices for their children and to identify gaps in the literature around parents' use of online health information and their vaccination choices.

**Methods:** A scoping review of peer-reviewed literature from Canada and the United States was performed. The following databases were utilized to perform the search: PubMed, CINAHL, Nursing & Allied Health Database, Scopus, and PsycINFO. The purpose of this review was to examine parents' use of online information seeking related to vaccine information and to understand how parents utilize this information to inform decisions about vaccinating their children. Of the 34 papers included in the review, 4 relevant themes and subthemes were identified: information seeking, online information resources, online vaccine content, and trust in health care providers.

**Results:** Examination of the literature revealed conflicting information regarding parents' use of social media and online resources to inform decisions around vaccinating their children. There is evidence of significant misinformation regarding vaccine risks online. Parents' digital health literacy levels are unknown and may affect their ability to appraise online vaccination information.

**Conclusions:** Parents are seeking vaccine information from online sources. However, the influence of online vaccine information on parental vaccine practices remains uncertain.

(*J Med Internet Res* 2020;22(10):e20002) doi:[10.2196/20002](https://doi.org/10.2196/20002)

**KEYWORDS**

childhood vaccination; parental vaccine decisions; online vaccine information; social media; vaccine hesitancy; digital health literacy

## Introduction

Vaccination programs are a vital contribution to public health practice in North America [1,2]. Immunizing children throughout their early years prevents the spread of communicable disease and decreases the morbidity and mortality associated with many vaccine-preventable diseases. Sustaining

vaccination rates above 95% maintains community immunity and prevents outbreaks of vaccine-preventable diseases [3]. Despite ongoing public health efforts in North America, childhood vaccination rates are not meeting the established goals for effective disease prevention [4,5]. In 2019, outbreaks of measles, a highly contagious, vaccine-preventable disease, occurred in both the United States and Canada [6]. Health Canada has highlighted the importance of understanding factors



associated with under-vaccinated and unvaccinated children, as well as vaccine hesitancy among parents and guardians, as key to improving vaccination rates in Canada [4]. Public health officials in the United States have also identified that parental delay and refusal of vaccinations threaten community immunity and identify this issue as a significant research priority [7]. The health implications for under-vaccinated children are concerning as the prevalence of communicable disease outbreaks become increasingly common. The use of digital information, including the prevalence of social media use among Canadian parents, highlights a need to understand the impact of online information on parents' vaccine choices.

National immunization targets for Canadian children aim for 95% vaccination coverage by a child's seventh birthday for the following diseases: diphtheria, tetanus, pertussis, polio, measles, mumps, and rubella [8]. Recently reported statistics demonstrate that vaccination rates fall short of the established immunization targets across Canada; immunization rates for 7-year-olds are 71% for diphtheria, tetanus, and pertussis; 90% for polio; and 86% for measles, mumps, and rubella [8]. American statistics also demonstrate immunization rates lower than 95%, with 83.2% of children aged 35 months having received at least 4 doses of diphtheria, tetanus, and pertussis vaccine; 92.7% having at least 3 doses of polio vaccine; and 91.5% having at least 1 dose of the measles, mumps, and rubella vaccine [9].

Vaccine hesitancy among parents or guardians is a growing public health issue reflected by the increased number of medical and nonmedical vaccine exemptions in both Canada and the United States [10]. Vaccine hesitancy is defined by the World Health Organization as the reluctance or refusal to vaccinate despite the availability of vaccines [11]. Despite continued efforts to improve childhood vaccination rates, both Canada and the United States are not meeting national goals, and nonmedical exemptions continue to proliferate [4,9].

Accessing information about the benefits and risks of childhood vaccines helps parents make informed decisions regarding vaccinating their children. Searching online via the internet allows individuals rapid access to health information. A 2016 survey of Canadians' online activity demonstrated that 96% of Canadians aged 15-34 years and 93% of those aged 35-44 years use the internet on a daily basis [12], and as many as 79% of American internet users have searched online for health information [13]. While accessing health information online is important, digital health literacy or having the skills to seek, find, understand, and appraise online health information and then apply that knowledge to making an informed decision is critical [14]. While population-based assessments of Canadian digital health literacy levels are unknown, the health literacy of Canadians is concerning. Over 60% of Canadians have low health literacy skills that place them at higher risk of poor health [15,16]. Canadian adults' health literacy skills were measured utilizing the International Adult Literacy and Skills Survey that assesses prose literacy, document literacy, numeracy, and problem-solving skills in different languages and cultures focused on broadly defined health content in the following areas: health promotion, health protection and accident prevention, disease prevention, and health care activities [16].

Vaccination information is available to most Canadians and Americans, although understanding and applying this information to one's individual health can be challenging. Online digital health information has evolved from static information retrievable from online websites to include interactive and collaborative sites where there is no central authority [17]. Within the context of childhood immunization information, parents are able to retrieve information but also contribute their personal knowledge and experience through interactive online social media platforms. Given the high prevalence of online information seeking among parents [18-20], investigating what information exists online related to childhood vaccinations and how parents use this information may provide insight into vaccination decision making. This scoping review examined research regarding parents' use of online resources regarding primary schedule vaccinations, to understand where parents are searching online and how they utilize online information to inform vaccination choices for their infants and children.

A scoping review was undertaken for 2 reasons: to grasp an understanding of the existing literature of parents' online health information-seeking behavior to inform vaccination choices for their children and to identify gaps in the literature around parents' use of online health information resources and their vaccination choices.

The following research questions informed this review: "What are parents' online information-seeking patterns and behaviors related to childhood primary immunization series?" and "How did parents use online resources to inform their decision regarding vaccination of their children?"

## Methods

A scoping review of the research literature was an appropriate method to examine this issue that is inclusive of qualitative, quantitative, and mixed methods literature to achieve a breadth of knowledge in this subject area. Researchers followed the collective guidelines of Colquhoun et al [21], Arksey and O'Malley [22], and Levac et al [23] to conduct this scoping review study. The steps involved in this review were identification of the research question; identification of relevant studies; study selection; charting the data; and collating, summarizing, and reporting the results [22].

### Identification and Study Selection

The following databases were searched: PubMed, CINAHL, Nursing & Allied Health Database, Scopus, and PsycINFO. The following search terms were used: vaccine, vaccines, vaccination, immunization, vaccinated, vaccinate, vaccine hesitancy, parent, parents, mother, mothers, father, fathers, parental, social media, digital health information, facebook, twitter, pinterest, snapchat, tumblr, Instagram, linkedin, google plus, youtube, reddit, flickr, vine, quora, periscope, whatsapp, and internet. Search terms were combined using AND or OR in the database search. A research librarian was consulted to assist with the search strategy. A justification search was completed with the Allied and Alternative Medicine (AMED) database; this search revealed no further articles in the subject area. Grey literature was located by searching Proquest

Dissertations & Theses Global. Reference lists of articles were hand searched to identify any further literature that met the inclusion criteria.

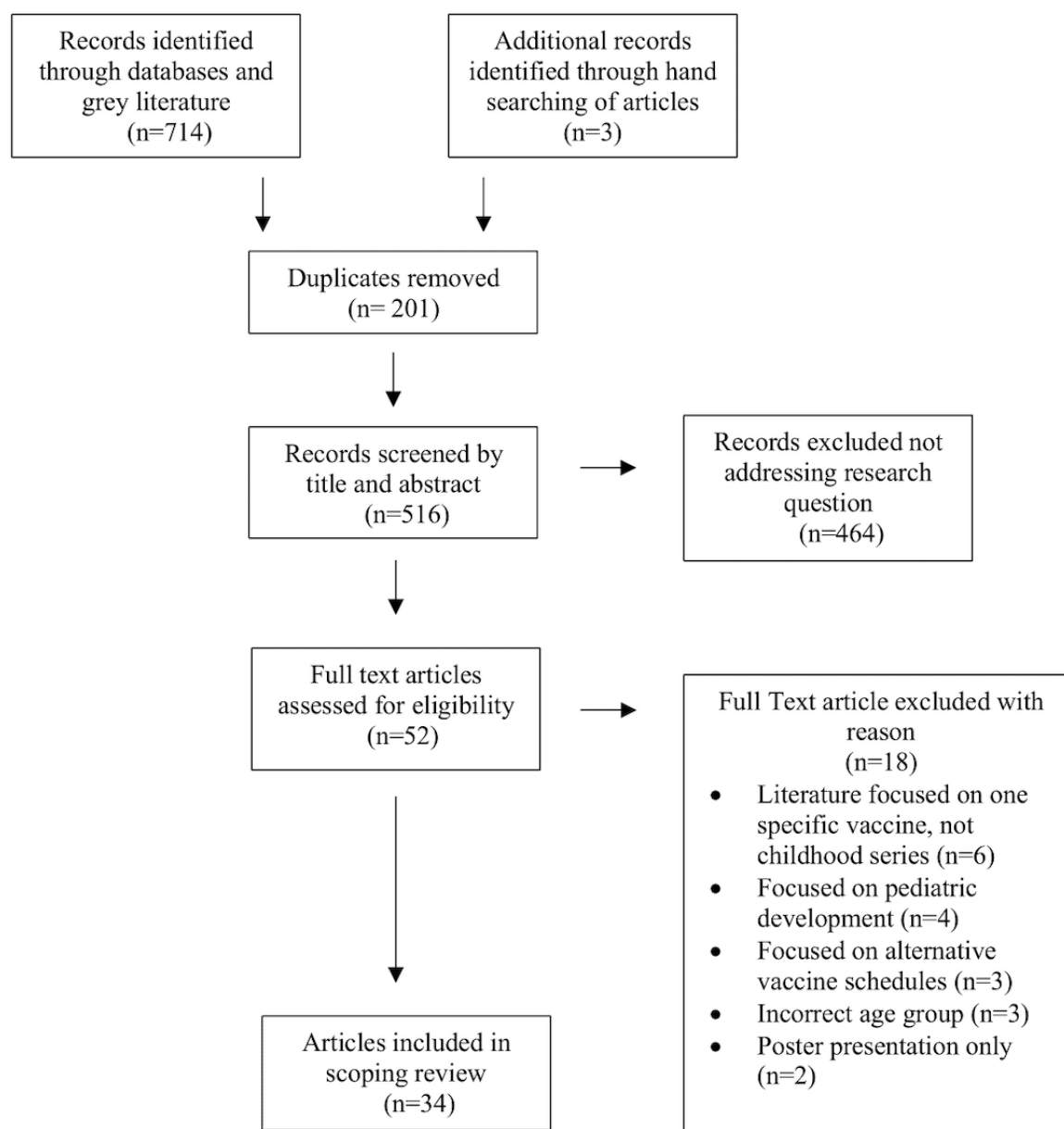
### Inclusion Criteria

Articles published from January 1, 2010 to December 31, 2019 in the English language from Canada and the United States were included. The aim was to focus on the past decade to reflect changes in information seeking that have occurred with widespread internet access in North America. The selected literature focused solely on the primary immunization series of children, parental decision relating to childhood vaccinations, online vaccination information seeking, and social media and childhood vaccinations.

### Exclusion Criteria

Articles published prior to January 1, 2010 were excluded as it is the authors' intent to determine the use of current social media and internet. Articles that focused on vaccination outside of the primary childhood series such as human papillomavirus and influenza and adult and adolescent hepatitis vaccinations were excluded. Any articles that focused on general childhood development and on adolescent vaccination decision making were excluded, as the intention is to focus the topic on parents' choices regarding vaccinations. Literature involving children who were able to consent to their own vaccines was also excluded from this review. Articles published in languages other than English were excluded, and articles where research was conducted outside of North America were excluded. [Figure 1](#) illustrates the PRISMA selection process [24].

**Figure 1.** PRISMA process to select articles.



## Analysis

A title review of all articles was conducted for subject eligibility, followed by an abstract review and finally a full article review. A second reviewer read 15% of the articles to ensure consistency in data analysis to address the research question and purpose [22,23]. Discrepancies between inclusion of articles were discussed between the authors until consensus was reached. After eligible articles were identified, each was read several times for detail and to identify relevant categories and themes. Articles were identified by a numbering system and collated in a spreadsheet identifying country of origin, methodologies, limitations, instruments, methods, and key findings from an iterative and inductive analysis of all articles. Thematic analysis allowed for identification of relevant themes grounded in the data and gaps in the literature. An iterative analysis process was utilized by the authors; themes were identified through discussion and immersion in the literature. Discrepancies were reconciled through discussion until all themes were mutually agreed upon.

## Results

The eligibility criteria for inclusion in the study were met by 34 articles. The majority (28/34, 82%) of the literature is from the United States, and 18% (6/34) is from Canada. The majority (20/34, 58%) of the literature was qualitative in nature, 32% (11/34) was quantitative, and 9% (3/34) used mixed method studies. Of the mixed method studies, 2 studies were concurrent, and 1 was sequential. Literature on social media was focused on the second half of the decade, with all but one article published between 2015 and 2019. The following 4 broad themes were identified: information seeking, online information resources, online vaccination information content, and trust in health care providers. A narrative review of these themes follows.

### Information Seeking

One of the most prevalent themes identified throughout the literature was one of information seeking; 50% (17/34) of articles discussed this theme [18,19,25-39]. Parents are looking online for vaccination information [18,26,27,30,31,39].

Online information seeking may have implications for the way that parents perceive the health and safety consequences of childhood vaccinations [20]. Canadian researchers investigated the impact of parental online information seeking through surveying parents and found that parents who searched the internet for childhood vaccine information (in 2011 and 2014) are more likely to perceive vaccines as less safe than those who did not search on the internet [20]. Yet, in another study also performed in 2011 that surveyed American parents, researchers found that 95% of parents of school-aged children who chose not to have their children vaccinated listed their health care provider as their main source of information, with only 34.5% reporting the internet as a source of information [25].

Mossey et al [34] evaluated Canadian parents' experiences in making vaccination decisions. Parents identified that searching for information was an important part of their decision-making process and that locating and interpreting online information

was difficult at times [34]. Reportedly, some parents looked for information online to confirm information provided by their health care provider; parents expressed concerns about their health care provider's lack of knowledge regarding childhood vaccines [35]. There is some evidence that parents who delayed or declined vaccinating their children specifically sought information through online social media platforms, such as blogs and videos, appreciating the personal experiences from other parents found on these platforms [34]. Jones et al [18] examined the information sources of parents of school-aged children in the United States who refused at least one vaccine for their child(ren). Researchers assessed the impact of diverse information sources on vaccine attitudes, beliefs, and medical exemptions. In this study, 40% of all respondents reported that the internet was a good source of vaccine information; however, only 19.9% of all parents in this study reported using the internet as a source of vaccine information [18]. Those who reported using the internet to obtain information about vaccines were less likely to agree that their children needed or would benefit from vaccines and were more likely to have obtained a nonmedical vaccine exemption for their child [18].

Hwang and Shah [33] examined the associations between health information sources, parental perceptions of vaccine benefits, and maintenance of vaccine schedules. They evaluated magazines, newspapers, television, the internet (search engines, general websites, drug company websites, and other health websites), interpersonal communication (doctors, nurses or physician assistants, pharmacists, and friends), and social media (social networking sites, blogs or vlogs) as health information sources [33]. Parents that valued social media as a source of health information reported fewer perceived benefits (identified as vaccine benefits outweighing vaccine risks) of vaccinations [33]. Positive parental perceptions of vaccine benefits were strongly associated with the maintenance of vaccination schedules [33].

Berthel [25] assessed sources of information among parents of school-aged children in the United States. They found that 95.4% of parents sought vaccine advice from their health care provider, 51.1% from family and friends, 34.5% from the internet, 25.3% from the news, and 15.5% from television and radio. Parents who cited the internet as a source of information listed the sites they most frequently used as webmd.com (34.4%), mayoclinic.com (27.5%), and government or other medically endorsed sites (68.9%) [25]. Researchers compared information sources of parents who vaccinated their child with those who exempted their children from vaccines. While 83.3% of immunizers versus 77.8% of vaccine abstainers identified health care providers as sources of vaccine information, only 25% of parents who immunized their children utilized the internet as a source of information compared to 44.4% of parents who did not immunize [25]. Deas et al [29] interviewed parents of school-aged children in 3 counties in the US state of Oregon with low immunization rates (combined pediatric immunization rate of 65%). Researchers found that all parents, regardless of vaccine acceptance or hesitancy, dismissed social media as they found it an unreliable source of vaccine information [29].

However, when exposed to evidence-informed online information, parental attitudes regarding childhood vaccinations

shifted; researchers found that, after exposure to a vaccine information website developed by experts in the field, vaccine-hesitant parents had a more positive attitude toward vaccines compared to vaccine-hesitant parents with no exposure to the evidence-informed website [28]. Glanz et al [32] evaluated vaccine information-seeking behaviors in parents who accessed information from an online website. The website was developed and mediated by several experts (pediatrician, vaccine safety researcher, and risk communication specialist). Parents in this study preferred to engage with online experts about vaccinations rather than interacting with other parents within this online site [32].

Researchers who focused on the composition of search terms used by parents seeking online vaccine information found that parents who utilized positive search terms (eg, “vaccine benefits”) when looking for childhood vaccine information encountered few myths about vaccine safety and effectiveness [40]. In contrast, parents who used negative search terms (eg, “vaccine risks”) found 4.8 times more misinformation or myths than a person using positive terminology [40]. The nature of the search terms used to find online vaccination information can alter the results and the information made available to parents [40].

### Online Information Resources

Parents searched online for vaccine information using common search engines such as Google and Yahoo, and many used popular social media sites such as Facebook, Instagram, Pinterest, Twitter, and YouTube [27,38,41-43]. One study found that the internet was listed within the top 3 most common sources for vaccine information among those who vaccinated their children, and 46% of parents who used the internet reported using search engines such as Google and Yahoo to search online for vaccine information [27].

The popularity of social media has given rise to prosumers — individuals who not only access online information but produce online content. Researchers found that websites that allow users to post online content without verification of information promoted antivaccination messages through antigovernment views, celebrities, personal stories, and naturalist arguments [43]. A Canadian assessment of online vaccine information websites (identified through searching Google, Facebook, Twitter, and YouTube) targeted to parents found that the majority of websites offered poor-quality information regarding childhood vaccination [42]. Researchers rated the websites with a communication index tool developed by the Centers for Disease Control and reported that 5% of materials (3 websites) met the standards for clear communication [42]. However, websites that monitored user-generated content and required academic references demonstrated a balance between openness and credibility [43].

### Online Vaccination Information Content

Researchers assessed the content of online vaccine information, and there is evidence that content was conflicting and inconsistent [20,41,43-47].

Regarding online websites, American researchers investigating online information content generated from Google searching

found that 59% of the first 100 Google sources were provaccine and 41% were antivaccine [43]. Similarly, Kata [46] sought antivaccination websites to assess their content and accuracy. Researchers utilized neutral vaccine information-seeking search terms (“vaccine,” “vaccination,” and “immunization OR immunisations”) to search American and Canadian versions of Google and found that, within the first 10 results, 24% of American and 13% of Canadian results were antivaccine [46]. A content analysis revealed that all of the assessed antivaccine websites claimed that vaccines are poisonous and cause idiopathic illnesses [46]. As well, 88% of the websites contained information that challenged the evidence about the efficacy of vaccines and whether vaccines conferred immunity, 88% endorsed alternative treatments (homeopathy, chiropractic care, naturopathy, or acupuncture) as superior to vaccination, 75% of the websites made accusations that regulatory bodies have information about vaccines that they are hiding from the public, and 75% of the websites suggested that vaccine promotion is solely motivated by profit seeking [46].

The content developers of websites advocating antivaccine messages used sophisticated strategies to communicate their perspective [48,49]. Researchers examined the content of antivaccine websites and found that these websites used persuasive tactics, attacked the credibility of vaccine advocates, expressed mistrust about scientific evidence, and used pseudoscientific evidence to support claims in favor of antivaccination [49]. Similarly, Getman et al [48] performed a network analysis of over 50,000 websites that contained vaccine-relevant content to determine the structure and influence of the online vaccine-hesitant community. They discovered effective use of hierarchical scientific language by the vaccine-hesitant community to enforce their online authority [48].

Regarding social media vaccine resources, researchers evaluated literature on vaccine information on the following social media platforms: YouTube, Twitter, Pinterest, Facebook, and various blogs. Facebook was identified as a vaccine information resource for parents in several studies [20,38,42,50]. The majority of research literature that focused on social media use among parents was published between 2015 and 2019, reflecting parents’ recent use of social media for seeking vaccine information. Parents also accessed social media sites to understand other parents’ experiences with vaccination processes; peer-to-peer information was a valued source of information accessible on social media sites [50]. Across social media platforms, there was a mix of positive and negative vaccine messaging. Our findings indicate widespread mistrust of government institutions and skepticism towards the vaccine industry across all social media platforms except for pediatrician-authored blogs [37,41,44,45,47,50-53]. Vaccine-hesitant online communities tend to leverage scientific and academic language to enforce their antivaccination narrative [48]. Consistent messaging about the dangers of vaccines was present on Twitter, Facebook, YouTube, parenting blogs, and Pinterest. The dominant message across social media was one of elevated health risks among vaccinated children — perceived as information concealed by government and industry.



## Trust in Health Care Providers

The concept of trusted health care providers was prevalent in many of the reviewed studies [18,20,27,30,31,34,35,45,46,49,50]. Parents often listed their physician or health care provider as a trusted source of vaccination information [18,27,30,31]. Eller et al [30] examined the association between the level of trust a mother had with their pediatrician and vaccine information sources. Mothers who trusted their pediatrician were 2.47 times more likely to list their pediatrician as their main source of vaccine information compared to mothers who had not established a trusting relationship with their pediatrician [30]. Parents who chose to vaccinate their children reported their physician as the biggest influence in their decision to vaccinate [27].

However, parents who identified as vaccine-hesitant or those who did not vaccinate their children were less trusting of information conveyed by their health care provider [27]. Parents who chose to delay or declined vaccinating their children tended to seek information from a diverse group of health disciplines such as cardiologists, health researchers, health care students, and homeopathic practitioners [34]. Kata [46] identified mistrust in the medical system as a strong message on antivaccine websites and on an antivaccination Facebook group [46,50].

## Discussion

This scoping review of 34 articles investigated parents' use of online information to inform vaccination choices for their children. Overall, research studies were broadly focused on understanding the content of online vaccine information and parents' online information seeking to inform their vaccine decisions. Our analysis of the current literature indicates that parents are actively looking online for vaccination information. Vaccination information was found on the social media platforms of YouTube, Twitter, Pinterest, and Facebook as well as various types of blogs. However, there is conflicting information within the published research regarding parental trust in the information found online and utilization of online information. Google was reported as the main online search engine among parents seeking vaccine information in both countries. The search terms that parents used in their online information seeking significantly impacted the disposition of the vaccine information and exposure to the number of vaccine myths retrieved online. Trust in health care providers plays an important role in information seeking. Parents who trust their health care provider tend to value them as an accurate source of vaccine information while those who do not trust their health care provider often seek information online. This is consistent with literature that demonstrates higher quality of life, more beneficial health behaviors, and higher treatment satisfaction when patients trust their health care provider [54].

The challenge for parents seeking vaccine information was the conflicting information found online. The influence of online vaccine information on parental vaccination practices (provaccination vs antivaccination) remains uncertain. A continuum of online information seeking among parents ranged from a complete dismissal of online information to regular online information access to inform their vaccine decision

making. Due to the limited amount of Canadian research available, direct comparison between Canada and the United States was not plausible. Existing Canadian research findings highlight the relationship between online information seeking and parental antivaccination sentiment; parents seeking information online were more likely to perceive vaccines as less safe and be less inclined to adhere to the recommended vaccine schedule [20]. However, a greater number of studies in this area is needed to substantiate these findings. There is a need for further research performed in Canada on parents' use of online vaccine information to determine how it informs their vaccine decisions.

Essentially absent in the published research was an understanding of parents' digital health literacy skills. Only one of the studies in this review considered parents' digital health literacy. The ability to interact with health information becomes more complex within the digital health context. Digital health literacy refers to the knowledge and skills inclusive of the ability to read and understand general information (traditional literacy and numeracy); effectively use digital devices, which includes awareness of data privacy, security, and ownership (computer literacy); critically understand and assess sophisticated media messages (media literacy); discern what is reliable and valid health information (scientific literacy); source information (information literacy); and navigate the health care system (health literacy) [14]. Although parents are looking online for information to inform their decision, the impact of parents' health literacy and digital health literacy skills that support access to credible information sources and an ability to critically assess the vaccination information is a significant gap in the research literature [55,56]. In fact, research exploring parents' information ecosystems is warranted to fully understand their information-seeking practices, preferred resources, and ability to critically evaluate vaccination-related information.

There was no research in this review that focused on parents' use of Instagram as a source for vaccine information. Research focused on parental use of Instagram and other emerging popular social media networks as a source for vaccine information is warranted.

The influence of misinformation on parental vaccine choices is an issue that may have significant implications for maintaining community immunity. Parents encounter inaccurate and false vaccine information, vaccine conspiracy theories, and vaccine myths propagated online, especially on social media sites [42,52]. Parents, in their search for information, may be exposed to persuasive tactics that perpetuate myths and fear mongering. Persuasive tactics combined with misinformation and myths may cause parents to believe that vaccines themselves are a threat to their child's health. Similar findings on the harms of social media rumors and misinformation surrounding COVID-19 also demonstrate the detrimental effects of online myth propagation. Improper use of pharmacological drugs and panic buying have resulted from online COVID-19 myths [57], as well as concern over the disease being spread through meat consumption and Chinese biological military laboratories [58].

Given the diversity of health information sources, health care providers along with public health organizations need to work



together with popular social media platforms to ensure information accuracy. Recently, some social media platforms have implemented measures to prevent the propagation of antivaccine messaging and misinformation. Twitter recently (2019) integrated a search tool into their platform that directs users to a sanctioned US government vaccine website [59]. Pinterest also implemented community guidelines in 2019 that limit misinformation; Pinterest now removes pins that promote antivaccine advice and redirects individuals to reputable vaccine information sources such as the World Health Organization [60]. Facebook and Instagram also implemented similar policies in 2019. Facebook suggests users visit a public health website for vaccine information, and Instagram blocks some false information posts and reports using third-party fact checkers to help reduce false information [61]. However, the prevalence of online vaccination myths including the misconception that vaccination causes autism persists [18,20,53], demonstrating that further work needs to be done to dispel unsubstantiated vaccine myths.

Given the prevalence of online antivaccine information and the use of hierarchical and authoritative language among the online vaccine-hesitant community to promote the antivaccine sentiment, [48] the composition of information-seeking factors has the potential to influence the nature of information that parents are accessing. Algorithms are used by online search engines such as Google to tailor information and influence the outcome of individuals' online information seeking [62]. The algorithms determine the information outcome based on factors such as the search terms used, country, and type of digital device [62]. One of the algorithms used by Google determines if content is reliable and demonstrates "expertise, authoritativeness, and trustworthiness" on a given topic [62]. Ruiz and Bell [40] determined that the character or disposition of a Google search for vaccine information significantly affects the outcome of the search. Parents who are concerned about and search for risk information regarding vaccines will encounter more vaccine myths than parents who use neutral search terms to search for vaccine benefits [48]. Understanding that Google is the most popular online search engine, health care providers may want to consult with parents regarding their search terms for online vaccination-related information. Further research exploring the impact of search algorithms and information-seeking behaviors regarding online vaccination information would be beneficial. It may be that clinicians not only "prescribe" evidence-informed online resources to parents but in addition will need to consider the composition of search strategies (eg, search terms or strings) to mitigate parents' access to vaccine misinformation and myths. The use of information prescriptions has been successful in the past for online searching in the pediatric parent population [63]. Educating parents about the benefits and risks of searching online and prescribing search terms may allow parents to access

evidence-informed information to facilitate informed decision making and also mitigate the potential harms of online vaccine myths. Research on the use of a search term prescription in this population is required.

### Limitations

This review evaluated parental information seeking in those who have internet access; it did not capture parents who have limited or no access to online health information. The online environment is fluid due to its interactive user-driven features with websites and social media evolving and changing from day to day. Findings involving online information should be viewed cautiously as digital information changes rapidly. This review included English-only research reports, and research literature in languages other than English may have findings that are different than reported here.

### Conclusion and Implications

This review identified that parents are looking online on major search engines and social media platforms for vaccine information. It was identified that locating accurate information online regarding the benefits and risks of vaccines is challenging for parents given the low number of sources that contain accurate information [42]. There was conflicting evidence about how parents utilize information found online to inform their vaccine choices. However, vaccine-hesitant parents who have access to accurate online vaccine information have significant improvement in attitudes regarding vaccination benefits and reduction in parental concerns about vaccination risks [28]. Given the plethora of misinformation perpetuated online, clinicians may want to provide "information prescriptions" to parents regarding the search terms they use and encourage parents to access websites moderated by health care experts. The interactive component to the websites would provide an opportunity for parents to ask questions of vaccination experts. Health care providers should discuss with parents the nature of online vaccine discussions. Reviewing with parents the utilization of hierarchical and scientific language utilized by some to promote antivaccine messaging. Further discussion focused on search terms and even providing parents suggestions for vaccine-positive or neutral terminology that will allow them access to a more balanced discussion of vaccine benefits and risks online.

Parents identify trust as a fundamental part of the vaccine decision-making process. This importance placed upon perceived trust in the source of information reinforces the importance of relational care practices, and a trusting relationship with a health care provider is a priority. Developing and fostering trust between primary care providers and parents may be a strategy to increase vaccine uptake by parents.

### Conflicts of Interest

None declared.

### References

1. Centers for Disease Control and Prevention. Vaccinate Your Baby for Best Protection. Centers for Disease Control and Prevention. 2019. URL: <http://www.cdc.gov/features/infantimmunization/> [accessed 2020-10-02]
2. Public Health Agency of Canada. Canadian Immunization Guide. Public Health Agency of Canada. 2014. URL: <https://www.canada.ca/en/public-health/services/canadian-immunization-guide.html> [accessed 2019-09-12]
3. Andre FE, Booy R, Bock HL, Clemens J, Datta SK, John TJ, et al. Vaccination greatly reduces disease, disability, death and inequity worldwide. *Bull World Health Organ* 2008 Feb;86(2):140-146 [FREE Full text] [doi: [10.2471/blt.07.040089](https://doi.org/10.2471/blt.07.040089)] [Medline: [18297169](https://pubmed.ncbi.nlm.nih.gov/18297169/)]
4. Government of Canada. Highlights from the 2017 childhood National Immunization Coverage Survey (cNICS). Government of Canada. 2019. URL: <https://www.canada.ca/en/services/health/publications/vaccines-immunization/vaccine-uptake-canadian-children-preliminary-results-2017-childhood-national-immunization-coverage-survey.html> [accessed 2019-09-12]
5. Hill HA, Elam-Evans LD, Yankey D, Singleton JA, Kang Y. Vaccination Coverage Among Children Aged 19-35 Months - United States, 2017. *MMWR Morb Mortal Wkly Rep* 2018 Oct 12;67(40):1123-1128 [FREE Full text] [doi: [10.15585/mmwr.mm6740a4](https://doi.org/10.15585/mmwr.mm6740a4)] [Medline: [30307907](https://pubmed.ncbi.nlm.nih.gov/30307907/)]
6. World Health Organization. New measles surveillance data for 2019. World Health Organization. 2019. URL: <http://www.who.int/immunization/newsroom/measles-data-2019/en/> [accessed 2019-12-03]
7. Messonnier N. U.S. public health response to the Measles outbreak. Centres for Disease Control and Prevention. 2019 Feb 07. URL: <https://www.cdc.gov/washington/testimony/2019/t20190227.htm> [accessed 2020-01-07]
8. Government of Canada. Vaccination Coverage Goals and Vaccine Preventable Disease Reduction Targets by 2025. Government of Canada. 2019. URL: <https://www.canada.ca/en/public-health/services/immunization-vaccine-priorities/national-immunization-strategy/vaccination-coverage-goals-vaccine-preventable-diseases-reduction-targets-2025.html#1.0> [accessed 2019-10-20]
9. Centres for Disease Control and Prevention. Immunization. Centres for Disease Control and Prevention. 2017. URL: <https://www.cdc.gov/nchs/fastats/immunize.htm> [accessed 2020-02-26]
10. Wilson SE, Seo CY, Lim GH, Fediurek J, Crowcroft NS, Deeks SL. Trends in medical and nonmedical immunization exemptions to measles-containing vaccine in Ontario: an annual cross-sectional assessment of students from school years 2002/03 to 2012/13. *CMAJ Open* 2015;3(3):E317-E323. [doi: [10.9778/cmajo.20140088](https://doi.org/10.9778/cmajo.20140088)] [Medline: [26457292](https://pubmed.ncbi.nlm.nih.gov/26457292/)]
11. World Health Organization. Ten threats to global health in 2019. World Health Organization. 2019. URL: <https://www.who.int/vietnam/news/feature-stories/detail/ten-threats-to-global-health-in-2019> [accessed 2019-11-06]
12. Statistics Canada. The Internet and Digital Technology. Statistics Canada. 2017. URL: <https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2017032-eng.htm> [accessed 2019-09-16]
13. Pew Research Center. Health Information Online. Pew Research Center. 2005 May 17. URL: <https://www.pewresearch.org/internet/2005/05/17/health-information-online/> [accessed 2019-11-17]
14. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. *J Med Internet Res* 2006 Jun 16;8(2):e9 [FREE Full text] [doi: [10.2196/jmir.8.2.e9](https://doi.org/10.2196/jmir.8.2.e9)] [Medline: [16867972](https://pubmed.ncbi.nlm.nih.gov/16867972/)]
15. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011 Jul 19;155(2):97-107. [doi: [10.7326/0003-4819-155-2-201107190-00005](https://doi.org/10.7326/0003-4819-155-2-201107190-00005)] [Medline: [21768583](https://pubmed.ncbi.nlm.nih.gov/21768583/)]
16. Canadian Counsel on Learning. Health literacy in Canada: A healthy understanding. Canadian Counsel on Learning. Ottawa: Canadian Council on Learning; 2008. URL: <http://www.en.copian.ca/library/research/ccl/health/health.pdf> [accessed 2019-12-04]
17. Lo B, Parham L. The impact of web 2.0 on the doctor-patient relationship. *J Law Med Ethics* 2010;38(1):17-26. [doi: [10.1111/j.1748-720X.2010.00462.x](https://doi.org/10.1111/j.1748-720X.2010.00462.x)] [Medline: [20446980](https://pubmed.ncbi.nlm.nih.gov/20446980/)]
18. Jones AM, Omer SB, Bednarczyk RA, Halsey NA, Moulton LH, Salmon DA. Parents' source of vaccine information and impact on vaccine attitudes, beliefs, and nonmedical exemptions. *Adv Prev Med* 2012;2012:932741 [FREE Full text] [doi: [10.1155/2012/932741](https://doi.org/10.1155/2012/932741)] [Medline: [23082253](https://pubmed.ncbi.nlm.nih.gov/23082253/)]
19. Jones DR. Vaccination decision-making of family members of children with autism. Ann Arbor: Capella University; 2015. URL: <https://www.lib.uwo.ca/cgi-bin/ezpauthn.cgi?url=http://search.proquest.com/docview/1709469954?accountid=15115> [accessed 2019-08-20]
20. Tustin JL, Crowcroft NS, Gesink D, Johnson I, Keelan J. Internet Exposure Associated With Canadian Parents' Perception of Risk on Childhood Immunization: Cross-Sectional Study. *JMIR Public Health Surveill* 2018 Jan 19;4(1):e7 [FREE Full text] [doi: [10.2196/publichealth.8921](https://doi.org/10.2196/publichealth.8921)] [Medline: [29351896](https://pubmed.ncbi.nlm.nih.gov/29351896/)]
21. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. *J Clin Epidemiol* 2014 Dec;67(12):1291-1294. [doi: [10.1016/j.jclinepi.2014.03.013](https://doi.org/10.1016/j.jclinepi.2014.03.013)] [Medline: [25034198](https://pubmed.ncbi.nlm.nih.gov/25034198/)]
22. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* 2005 Feb;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
23. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci* 2010 Sep 20;5:69 [FREE Full text] [doi: [10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69)] [Medline: [20854677](https://pubmed.ncbi.nlm.nih.gov/20854677/)]

24. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
25. Berreth T. Reasons behind immunization exemptions at school entry in Idaho: Parents' attitudes and beliefs. Ann Arbor: University of Idaho; 2011. URL: <https://www.lib.uwo.ca/cgi-bin/ezpauthn.cgi?url=http://search.proquest.com/docview/929133794?accountid=15115> [accessed 2019-08-20]
26. Brunson EK. The impact of social networks on parents' vaccination decisions. *Pediatrics* 2013 May;131(5):e1397-e1404. [doi: [10.1542/peds.2012-2452](https://doi.org/10.1542/peds.2012-2452)] [Medline: [23589813](https://pubmed.ncbi.nlm.nih.gov/23589813/)]
27. Chung Y, Schamel J, Fisher A, Frew PM. Influences on Immunization Decision-Making among US Parents of Young Children. *Matern Child Health J* 2017 Dec;21(12):2178-2187 [FREE Full text] [doi: [10.1007/s10995-017-2336-6](https://doi.org/10.1007/s10995-017-2336-6)] [Medline: [28755045](https://pubmed.ncbi.nlm.nih.gov/28755045/)]
28. Daley MF, Narwaney KJ, Shoup JA, Wagner NM, Glanz JM. Addressing Parents' Vaccine Concerns: A Randomized Trial of a Social Media Intervention. *Am J Prev Med* 2018 Jul;55(1):44-54. [doi: [10.1016/j.amepre.2018.04.010](https://doi.org/10.1016/j.amepre.2018.04.010)] [Medline: [29773490](https://pubmed.ncbi.nlm.nih.gov/29773490/)]
29. Deas J, Bean SJ, Sokolovska I, Fautin C. Childhood Vaccine Attitudes and Information Sources Among Oregon Parents and Guardians. *Health Promot Pract* 2019 Jul;20(4):529-538. [doi: [10.1177/1524839918778830](https://doi.org/10.1177/1524839918778830)] [Medline: [29884085](https://pubmed.ncbi.nlm.nih.gov/29884085/)]
30. Eller NM, Henrikson NB, Opel DJ. Vaccine Information Sources and Parental Trust in Their Child's Health Care Provider. *Health Educ Behav* 2019 Jun;46(3):445-453. [doi: [10.1177/1090198118819716](https://doi.org/10.1177/1090198118819716)] [Medline: [30616381](https://pubmed.ncbi.nlm.nih.gov/30616381/)]
31. Freed GL, Clark SJ, Butchart AT, Singer DC, Davis MM. Sources and perceived credibility of vaccine-safety information for parents. *Pediatrics* 2011 May;127 Suppl 1:S107-S112. [doi: [10.1542/peds.2010-1722P](https://doi.org/10.1542/peds.2010-1722P)] [Medline: [21502236](https://pubmed.ncbi.nlm.nih.gov/21502236/)]
32. Glanz JM, Wagner NM, Narwaney KJ, Kraus CR, Shoup JA, Xu S, et al. Web-based Social Media Intervention to Increase Vaccine Acceptance: A Randomized Controlled Trial. *Pediatrics* 2017 Dec;140(6):1-9. [doi: [10.1542/peds.2017-1117](https://doi.org/10.1542/peds.2017-1117)] [Medline: [29109107](https://pubmed.ncbi.nlm.nih.gov/29109107/)]
33. Hwang J, Shah DV. Health Information Sources, Perceived Vaccination Benefits, and Maintenance of Childhood Vaccination Schedules. *Health Commun* 2019 Oct;34(11):1279-1288. [doi: [10.1080/10410236.2018.1481707](https://doi.org/10.1080/10410236.2018.1481707)] [Medline: [29869894](https://pubmed.ncbi.nlm.nih.gov/29869894/)]
34. Mossey S, Hosman S, Montgomery P, McCauley K. Parents' Experiences and Nurses' Perceptions of Decision-Making About Childhood Immunization. *Can J Nurs Res* 2019 Apr 30;844562119847343. [doi: [10.1177/0844562119847343](https://doi.org/10.1177/0844562119847343)] [Medline: [31039630](https://pubmed.ncbi.nlm.nih.gov/31039630/)]
35. Sobo EJ, Huhn A, Sannwald A, Thurman L. Information Curation among Vaccine Cautious Parents: Web 2.0, Pinterest Thinking, and Pediatric Vaccination Choice. *Med Anthropol* 2016;35(6):529-546. [doi: [10.1080/01459740.2016.1145219](https://doi.org/10.1080/01459740.2016.1145219)] [Medline: [26814029](https://pubmed.ncbi.nlm.nih.gov/26814029/)]
36. Tustin JL. Doctoral Thesis: The Internet and Childhood Immunizations in Canada. 2016. URL: <https://tspace.library.utoronto.ca/handle/1807/73209> [accessed 2019-08-20]
37. Tustin JL, Crowcroft NS, Gesink D, Johnson I, Keelan J, Lachapelle B. User-Driven Comments on a Facebook Advertisement Recruiting Canadian Parents in a Study on Immunization: Content Analysis. *JMIR Public Health Surveill* 2018 Sep 20;4(3):e10090 [FREE Full text] [doi: [10.2196/10090](https://doi.org/10.2196/10090)] [Medline: [30249585](https://pubmed.ncbi.nlm.nih.gov/30249585/)]
38. Wachob DA, Boldy A. Social Media's Influence on Parents' Decision-Making Process of Child Vaccinations. *Epidemiol Biostat Public Health* 2019;16(1):5. [doi: [10.2427/13056](https://doi.org/10.2427/13056)]
39. Whyte MD, Whyte IV J, Cormier E, Eccles DW. Factors influencing parental decision making when parents choose to deviate from the standard pediatric immunization schedule. *J Community Health Nurs* 2011 Oct;28(4):204-214. [doi: [10.1080/07370016.2011.615178](https://doi.org/10.1080/07370016.2011.615178)] [Medline: [22053765](https://pubmed.ncbi.nlm.nih.gov/22053765/)]
40. Ruiz JB, Bell RA. Understanding vaccination resistance: vaccine search term selection bias and the valence of retrieved information. *Vaccine* 2014 Oct 07;32(44):5776-5780. [doi: [10.1016/j.vaccine.2014.08.042](https://doi.org/10.1016/j.vaccine.2014.08.042)] [Medline: [25176640](https://pubmed.ncbi.nlm.nih.gov/25176640/)]
41. Basch CH, Zybert P, Reeves R, Basch CE. What do popular YouTube videos say about vaccines? *Child Care Health Dev* 2017 Jul;43(4):499-503. [doi: [10.1111/cch.12442](https://doi.org/10.1111/cch.12442)] [Medline: [28105642](https://pubmed.ncbi.nlm.nih.gov/28105642/)]
42. Modanloo S, Stacey D, Dunn S, Choueiry J, Harrison D. Parent resources for early childhood vaccination: An online environmental scan. *Vaccine* 2019 Dec 03;37(51):7493-7500. [doi: [10.1016/j.vaccine.2019.09.075](https://doi.org/10.1016/j.vaccine.2019.09.075)] [Medline: [31590931](https://pubmed.ncbi.nlm.nih.gov/31590931/)]
43. Venkatraman A, Garg N, Kumar N. Greater freedom of speech on Web 2.0 correlates with dominance of views linking vaccines to autism. *Vaccine* 2015 Mar 17;33(12):1422-1425. [doi: [10.1016/j.vaccine.2015.01.078](https://doi.org/10.1016/j.vaccine.2015.01.078)] [Medline: [25665960](https://pubmed.ncbi.nlm.nih.gov/25665960/)]
44. Guidry JPD, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. *Vaccine* 2015 Sep 22;33(39):5051-5056. [doi: [10.1016/j.vaccine.2015.08.064](https://doi.org/10.1016/j.vaccine.2015.08.064)] [Medline: [26319742](https://pubmed.ncbi.nlm.nih.gov/26319742/)]
45. Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine* 2017 Jun 22;35(29):3621-3638 [FREE Full text] [doi: [10.1016/j.vaccine.2017.05.052](https://doi.org/10.1016/j.vaccine.2017.05.052)] [Medline: [28554500](https://pubmed.ncbi.nlm.nih.gov/28554500/)]
46. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine* 2010 Feb 17;28(7):1709-1716. [doi: [10.1016/j.vaccine.2009.12.022](https://doi.org/10.1016/j.vaccine.2009.12.022)] [Medline: [20045099](https://pubmed.ncbi.nlm.nih.gov/20045099/)]
47. Love B, Himelboim I, Holton A, Stewart K. Twitter as a source of vaccination information: content drivers and what they are saying. *Am J Infect Control* 2013 Jun;41(6):568-570. [doi: [10.1016/j.ajic.2012.10.016](https://doi.org/10.1016/j.ajic.2012.10.016)] [Medline: [23726548](https://pubmed.ncbi.nlm.nih.gov/23726548/)]

48. Getman R, Helmi M, Roberts H, Yansane A, Cutler D, Seymour B. Vaccine Hesitancy and Online Information: The Influence of Digital Networks. *Health Educ Behav* 2018 Aug;45(4):599-606. [doi: [10.1177/1090198117739673](https://doi.org/10.1177/1090198117739673)] [Medline: [29267129](https://pubmed.ncbi.nlm.nih.gov/29267129/)]
49. Moran MB, Lucas M, Everhart K, Morgan A, Prickett E. What makes anti-vaccine websites persuasive? A content analysis of techniques used by anti-vaccine websites to engender anti-vaccine sentiment. *Journal of Communication in Healthcare* 2016 Oct 03;9(3):151-163. [doi: [10.1080/17538068.2016.1235531](https://doi.org/10.1080/17538068.2016.1235531)]
50. Ma J, Stahl L. A multimodal critical discourse analysis of anti-vaccination information on Facebook. 2017 Oct;39(4):303-310. [doi: [10.1016/j.lisr.2017.11.005](https://doi.org/10.1016/j.lisr.2017.11.005)]
51. Meleo-Erwin Z, Basch C, MacLean SA, Scheibner C, Cadoret V. "To each his own": Discussions of vaccine decision-making in top parenting blogs. *Hum Vaccin Immunother* 2017 Aug 03;13(8):1895-1901 [FREE Full text] [doi: [10.1080/21645515.2017.1321182](https://doi.org/10.1080/21645515.2017.1321182)] [Medline: [28481675](https://pubmed.ncbi.nlm.nih.gov/28481675/)]
52. Shoup JA, Narwaney KJ, Wagner NM, Kraus CR, Gleason KS, Albright K, et al. Social Media Vaccine Websites: A Comparative Analysis of Public and Moderated Websites. *Health Educ Behav* 2019 Jun;46(3):454-462. [doi: [10.1177/1090198118818253](https://doi.org/10.1177/1090198118818253)] [Medline: [30596265](https://pubmed.ncbi.nlm.nih.gov/30596265/)]
53. Tangherlini TR, Roychowdhury V, Glenn B, Crespi CM, Bandari R, Wadia A, et al. "Mommy Blogs" and the Vaccination Exemption Narrative: Results From A Machine-Learning Approach for Story Aggregation on Parenting Social Media Sites. *JMIR Public Health Surveill* 2016 Nov 22;2(2):e166 [FREE Full text] [doi: [10.2196/publichealth.6586](https://doi.org/10.2196/publichealth.6586)] [Medline: [27876690](https://pubmed.ncbi.nlm.nih.gov/27876690/)]
54. Birkhäuser J, Gaab J, Kossowsky J, Hasler S, Krummenacher P, Werner C, et al. Trust in the health care professional and health outcome: A meta-analysis. *PLoS One* 2017;12(2):e0170988 [FREE Full text] [doi: [10.1371/journal.pone.0170988](https://doi.org/10.1371/journal.pone.0170988)] [Medline: [28170443](https://pubmed.ncbi.nlm.nih.gov/28170443/)]
55. Brigham EL, Goldenberg L, Stolfi A, Mueller GA, Forbis SG. Associations Between Parental Health Literacy, Use of Asthma Management Plans, and Child's Asthma Control. *Clin Pediatr (Phila)* 2016 Feb;55(2):111-117. [doi: [10.1177/0009922815587089](https://doi.org/10.1177/0009922815587089)] [Medline: [25994320](https://pubmed.ncbi.nlm.nih.gov/25994320/)]
56. Jimenez ME, Barg FK, Guevara JP, Gerdes M, Fiks AG. The impact of parental health literacy on the early intervention referral process. *J Health Care Poor Underserved* 2013 Aug;24(3):1053-1062. [doi: [10.1353/hpu.2013.0141](https://doi.org/10.1353/hpu.2013.0141)] [Medline: [23974380](https://pubmed.ncbi.nlm.nih.gov/23974380/)]
57. Tasnim S, Hossain MM, Mazumder H. Impact of Rumors and Misinformation on COVID-19 in Social Media. *J Prev Med Public Health* 2020 May;53(3):171-174 [FREE Full text] [doi: [10.3961/jpmph.20.094](https://doi.org/10.3961/jpmph.20.094)] [Medline: [32498140](https://pubmed.ncbi.nlm.nih.gov/32498140/)]
58. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *J Med Internet Res* 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: [10.2196/19016](https://doi.org/10.2196/19016)] [Medline: [32287039](https://pubmed.ncbi.nlm.nih.gov/32287039/)]
59. Harvey D. Helping you find reliable public health information on Twitter. Twitter. 2019 May 10. URL: [https://blog.twitter.com/en\\_us/topics/company/2019/helping-you-find-reliable-public-health-information-on-twitter.html](https://blog.twitter.com/en_us/topics/company/2019/helping-you-find-reliable-public-health-information-on-twitter.html) [accessed 2020-08-02]
60. Pinterest. Community guidelines. Pinterest Policy. 2019. URL: <https://policy.pinterest.com/en/community-guidelines> [accessed 2019-11-14]
61. Instagram. Reducing the Spread of False Information on Instagram. Instagram Help Center. 2020. URL: <https://help.instagram.com/1735798276553028> [accessed 2020-08-08]
62. Google. Search Algorithms. Google. 2020. URL: <https://www.google.com/search/howsearchworks/algorithms/> [accessed 2020-03-05]
63. D'Alessandro DM, Kreiter CD, Kinzer SL, Peterson MW. A randomized controlled trial of an information prescription for pediatric patient education on the Internet. *Arch Pediatr Adolesc Med* 2004 Sep;158(9):857-862. [doi: [10.1001/archpedi.158.9.857](https://doi.org/10.1001/archpedi.158.9.857)] [Medline: [15351750](https://pubmed.ncbi.nlm.nih.gov/15351750/)]

*Edited by G Eysenbach; submitted 08.05.20; peer-reviewed by N MacDonald, J Cataldi, TV Duong; comments to author 29.06.20; revised version received 14.08.20; accepted 02.09.20; published 13.10.20.*

*Please cite as:*

Ashfield S, Donelle L

Parental Online Information Access and Childhood Vaccination Decisions in North America: Scoping Review

*J Med Internet Res* 2020;22(10):e20002

URL: <http://www.jmir.org/2020/10/e20002/>

doi: [10.2196/20002](https://doi.org/10.2196/20002)

PMID: [33048055](https://pubmed.ncbi.nlm.nih.gov/33048055/)

(<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Original Paper

# Risk-Taking Behaviors and Adherence to HIV Pre-Exposure Prophylaxis in Users of Geosocial Networking Apps: Real-World, Multicenter Study

Hongyi Wang<sup>1,2,3,4</sup>, PhD; Jing Zhang<sup>1,2,3,4</sup>, PhD; Zhenxing Chu<sup>1,2,3,4</sup>; Qinghai Hu<sup>1,2,3,4</sup>, Prof Dr; Willa Dong<sup>5</sup>, MSPH; Xiaojie Huang<sup>6</sup>, Prof Dr; Yaokai Chen<sup>7</sup>, Prof Dr; Hui Wang<sup>8</sup>, Prof Dr; Xiaoqing He<sup>7</sup>; Lukun Zhang<sup>8</sup>, PhD; Zhili Hu<sup>1,2,3,4</sup>; Rantong Bao<sup>1,2,3,4</sup>; Shangcao Li<sup>1,2,3,4</sup>; Hang Li<sup>1,2,3,4</sup>; Sitong Cui<sup>1,2,3,4</sup>; Xia Jin<sup>1,2,3,4</sup>; Haibo Ding<sup>1,2,3,4</sup>, PhD; Wenqing Geng<sup>1,2,3,4</sup>, Prof Dr; Yongjun Jiang<sup>1,2,3,4</sup>, Prof Dr; Junjie Xu<sup>1,2,3,4\*</sup>, Prof Dr; Hong Shang<sup>1,2,3,4\*</sup>, Prof Dr

<sup>1</sup>NHC Key Laboratory of AIDS Immunology (China Medical University), National Clinical Research Center for Laboratory Medicine, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning Province, China

<sup>2</sup>Key Laboratory of AIDS Immunology, Chinese Academy of Medical Sciences, Shenyang, China

<sup>3</sup>Key Laboratory of AIDS Immunology of Liaoning Province, Shenyang, China

<sup>4</sup>Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou, China

<sup>5</sup>Department of Health Behavior, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

<sup>6</sup>Center for Infectious Diseases, Beijing Youan Hospital, Capital Medical University, Beijing, China

<sup>7</sup>Chongqing Public Health Medical Center, Chongqing, China

<sup>8</sup>Shenzhen Third People's Hospital, Shenzhen, China

\* these authors contributed equally

**Corresponding Author:**

Hong Shang, Prof Dr

NHC Key Laboratory of AIDS Immunology (China Medical University), National Clinical Research Center for Laboratory Medicine, The First Affiliated Hospital of China Medical University

No 155, Nanjing North Street, Heping District

Shenyang, Liaoning Province, 110001

China

Phone: 86 8328 2634

Email: [hongshang100@hotmail.com](mailto:hongshang100@hotmail.com)

**Related Article:**

This is a corrected version. See correction statement: <http://www.jmir.org/2020/11/e25565/>

## Abstract

**Background:** Over half of men who have sex with men (MSM) use geosocial networking (GSN) apps to encounter sex partners. GSN apps' users have become a unique large subpopulation among MSM for interventions concerning HIV prevention and control. Pre-exposure prophylaxis (PrEP) is a promising measure for HIV prevention, especially for MSM, but its effectiveness largely depends on medication adherence. However, little is known about PrEP adherence among GSN apps' users, which is critical to addressing the overall optimization of PrEP compliance outside of clinical trials in the context of large-scale implementation.

**Objective:** The objective of this study is to understand the correlation between GSN apps' use and medication adherence among MSM receiving PrEP, with the aim to increase their awareness about PrEP use in order to increase adherence.

**Methods:** This study based on the China Real-world Oral intake of PrEP (CROPrEP) project, a multicenter, real-world study of Chinese MSM on daily and event-driven PrEP. Eligible participants completed a detailed computer-assisted self-interview on sociodemographic, GSN apps' use, and sexual behavior. Then participants were followed up for 12 months and assessed for various characteristics (eg, PrEP delivery, adherence assessment, PrEP coverage of sexual activities, and regimens switch). A

generalized estimation equation was used to analyze the predictors of medication adherence and regimen conversion among GSN apps' users and nonusers.

**Results:** At baseline, 756 of the 1023 eligible participants (73.90%) reported primarily using GSN apps to seek sexual partners, and GSN apps' users are more likely to have high-risk behaviors such as multiple sex partners and condomless anal intercourse than other nonusers (all  $P < .05$ ). During follow-up, GSN apps' users had a significantly low level of pill-counting adherence than nonusers (adjusted odds ratio [aOR] 0.8, 95% CI 0.6-1.0,  $P = .038$ ). In the event-driven group, GSN apps' users had marginally lower levels of self-reported adherence (aOR 0.7, 95% CI 0.4-1.0,  $P = .060$ ) and lower PrEP coverage of sexual practices (aOR 0.6, 95% CI 0.4-1.0,  $P = .038$ ). Additionally, GSN apps' users seemed more likely to switch from event-driven to daily regimen (aOR 1.8, 95% CI 0.9-3.3,  $P = .084$ ).

**Conclusions:** GSN apps' users are highly prevalent among MSM, despite their higher sexual risk and lower adherence levels, suggesting that eHealth needs to be introduced to the GSN platform to promote PrEP adherence.

**Trial Registration:** Chinese Clinical Trial Registry ChiCTR-IIN-17013762; <https://tinyurl.com/yy2mhrv4>.

**International Registered Report Identifier (IRRID):** RR2-10.1186/s12879-019-4355-y

(*J Med Internet Res* 2020;22(10):e22388) doi:[10.2196/22388](https://doi.org/10.2196/22388)

## KEYWORDS

men who have sex with men (MSM); MSM; HIV/AIDS prevention; Pre-exposure prophylaxis (PrEP); geosocial networking app; dating app; adherence; regimen switch

## Introduction

Men who have sex with men (MSM) are estimated to have a 20-fold higher risk of new HIV infections than the general population on average, with MSM contributing to more than half of new HIV infections in western and central Europe and North America, and 30% in Asia and the Pacific [1]. Digital technology has changed the way MSM lives and seeks sexual partners, and a meta-analysis estimated that 50%-80% of MSM reported using geosocial networking (GSN) apps for sexual encounters [2,3]. Recent studies indicate an increase in GSN app users with high-risk sexual behaviors (ie, condomless and multisex partners); however, effective interventions for HIV prevention among this group remain an underexplored topic [2-7].

Pre-exposure prophylaxis (PrEP) is a promising novel HIV prevention method, and its coverage continues to expand globally [8]. Adherence is probably the strongest determinant of effectiveness for PrEP based on evidence from existing clinical trials and real-world studies [9-12], and optimizing adherence has become a common goal across various settings globally. Recent studies have found that GSN apps may facilitate the roll out of PrEP programs [13-15], though more information is needed on the role of GSN apps in PrEP initiation and adherence. In 2017, a large-scale survey of GSN apps' users among MSM in Europe and Central Asia reported that 10% of MSM were using PrEP in the last 3 months, and 33% were planning to start PrEP in the next 6 months [16]. A study on Grindr use among MSM from the United States found that those with recent GSN apps' use had a higher percentage of initiating PrEP compared with nonusers (24.6% vs 14%,  $P < .001$ ) [15]. The above evidence suggests that MSM with GSN apps' use experience may comprise a sizeable proportion of PrEP users, but we do not know exactly the size and risk level they occupy among PrEP users in China. In addition, despite the importance of adherence for PrEP effectiveness based on existing clinical trials and real-world studies among the overall MSM population,

there is little information on these outcomes for MSM who are using GSN apps. In 2019, the updated WHO guidelines for PrEP recommended both daily and event-driven regimens for PrEP among MSM [17], raising concerns about regimen switches. The limited available research evidence suggests that the switch between daily and event regimens is more common when 2 regimens are available [18]. Understanding adherence and regimens switch among GSN apps' users and addressing the barriers preventing adherence will be crucial to the long-term success of PrEP interventions in the Chinese MSM population. Discontinuation of PrEP is another concern and may hinder its prevention effect [19,20]. However, research has only scarcely explored the effect of app use behavior on PrEP discontinuation.

Therefore, this study aims to understand the behavioral characteristics of GSN apps' users among PrEP users and assess their adherence, regimen conversion, and discontinuation in a real-world study in China with options for daily and event-driven PrEP regimens, to explore new directions for improving adherence.

## Methods

### Study Design and Participants

This study was based on the China Real-world Oral intake of PrEP (CROPrEP) project [21], a multicenter, real-world study of daily or event-driven emtricitabine/tenofovir disoproxil fumarate (TDF/FTC) PrEP regimens among HIV-negative MSM at high risk of HIV infection in China. From December 2018 to October 2019, participants were recruited from 4 cities including Shenyang (First Affiliated Hospital of China Medical University), Beijing (Beijing You An Hospital of Capital Medical University), Shenzhen (Shenzhen Third People's Hospital), and Chongqing (Chongqing Public Health Medical Treatment Center). Eligible participants reported GSN apps' use and sexual characteristics at baseline and then began quarterly follow-up visits (eg, free PrEP distribution, medication adherence assessments, and computer-assisted self-interview [CASI]) over 12 months. The expected follow-up time was from

October 2019 to October 2020. Other methodological considerations, such as inclusion and exclusion criteria for eligible participants, are detailed in this study protocol paper [21].

## Procedures

Data analysis for this study was conducted from December 2018 to January 2020. Eligible participants completed 6 follow-up visits at baseline, 1, 3, 6, 9, and 12 months. At enrollment, after receiving professional medication counseling from doctors on daily and event-driven regimens, participants received free TDF/FTC as PrEP and chose between taking PrEP on a daily basis, or following an event-driven regimen (consisting of 2 pills 2-24 hours before sexual intercourse or 1 pill if the last medication was taken 1-6 days ago, and a pill every 24 h from the first drug intake during the period of sexual activity) [18]. Participants were permitted to change regimens during the study. A detailed computer-assisted self-administered questionnaire on GSN apps' use (referring to using GSN apps as the main venue for seeking sexual partners during the previous 3 months), sociodemographic information, sexual behavior, and substance use was also filled out by eligible participants to understand the distribution and characteristics of GSN apps' users among MSM on PrEP.

At each clinical follow-up visit, participants completed assessments of adherence through pill counts and CASI. Regimen switching events were measured in the CASI. Participants were required to return the remaining pills at each visit in exchange for a dose of TDF/FTC that would cover the next visit, and the pills were counted by study staff. Participants who tested positive for HIV discontinued TDF/FTC and were offered counseling and referral services. A unique 6-digit number was assigned to participants in place of their real names to protect privacy throughout the study. For each follow-up visit, participants receive 50 RMB (about US \$7.07) as compensation for travel and missed work.

## Outcomes and Core Variables

### Primary Outcomes

In our study, the primary outcomes were pill-counting adherence and regimen switches for PrEP. Participants were asked to return surplus pills at each follow-up, and adherence was calculated by counting the number of pills actually taken and dividing by the number of theoretical pills to be taken during that period. For the event-driven group, the theoretical number of pills to be taken is calculated according to the medication regimen: 2 pills of TDF/FTC orally 2-24 hours before sex, 1 pill 24 and 48 hours after the first dose (if the interval between the last dose is within 1-6 days, take 1 pill before sex), and 1 pill every 24 hours during continuous sex. The total number of pills taken in a week does not exceed 7. The numbers of drug taking days and sex days of the drug users were collected via the questionnaire; pill-counting adherence greater than 90% was defined as an adherence to PrEP. Participants were allowed to change regimens, and the date and direction of medication regimens switching from an event-driven regimen to a daily regimen or vice versa were collected through self-administered questionnaires during follow-up.

### Secondary Outcomes

Secondary outcomes included self-reported adherence, PrEP coverage of last sex events, and PrEP discontinuation. Self-reported adherence was assessed by a questionnaire on whether the medication was missed. We calculated recent sexual behavior with PrEP based on self-administered questionnaires and adjusted the data based on weekly records of pill taking and sexual practices. PrEP discontinuation excludes those who discontinue PrEP due to seroconversion to HIV positivity.

### GSN Apps' Use, Sexual Behaviors, and Other Observation Variables

The GSN app platforms mentioned in the questionnaire included *Blued*, *Jacked*, *Momo*, and other common gay social apps with geographical networking. Other adherence-related variables were also gathered via questionnaires. Specifically, questionnaires collected data on sociodemographics (eg, age, occupation, education), sexual behavior (ie, sexual role, number of sexual partners, type of sexual partners, condomless anal intercourse, and rectal douching), and substance use during the past 3 years. For sexual partners, casual partners were defined as one-night stands, as opposed to regular partners. Substance use was defined as the use of rush (poppers or alkyl nitrites), MDMA (3,4-methylenedioxymethamphetamine; ecstasy), ice, amphetamines, tramadol, or ketamine.

### Statistical Analysis

The distribution of demographic, behavioral, and PrEP-related characteristics of GSN apps' users and nonusers among MSM on PrEP (such as sexual behavior, substance use, and the choice of regimen) was compared using chi-square analysis or Fisher exact test, as appropriate. To clarify the determinants of PrEP adherence and regimen switches, generalized estimation equation with logistic models was used for analysis to obtain robust results by controlling for the correlation between observed variables [22,23], given that adherence, PrEP coverage of last sex events, and regimen switches were evaluated at each follow-up among GSN apps' users and nonusers. In this model, at each follow-up, data on time-dependent variables such as sexual behavior, drug use, and psychological factors were analyzed, and only variables statistically significant at the  $P < .20$  level were entered into the multivariate analysis with adjustment for demographic confounding factors including ethnicity, monthly income, and residence. All analyses were performed using SPSS version 24.0 (IBM). All  $P$ -values and confidence intervals were 2-sided.

### Ethics Statement

This study was approved by the Institutional Review Board at the 4 research centers and research procedures were carried out strictly in accordance with relevant guidelines ([2018]2015-139-5) and the Declaration of Helsinki. Written informed consents were obtained from participants prior to starting any study procedure. This study is registered with the Chinese Clinical Trial Registry (Trial registration number ChiCTR-IIN-17013762).

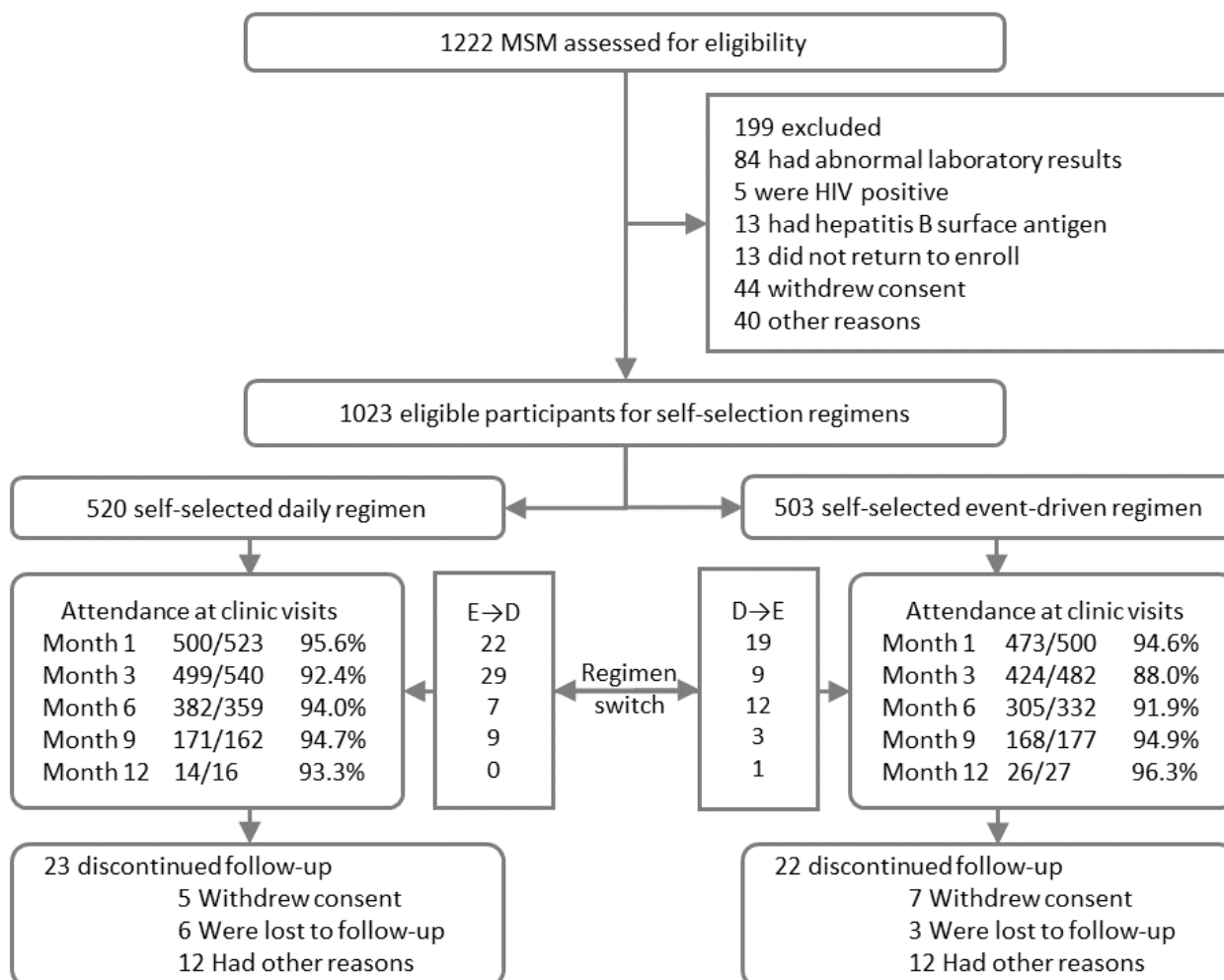
## Results

### Participant Demographics

Of the 1222 MSM screened, 1023 eligible participants were enrolled, including 520 MSM who chose daily PrEP and 503

MSM who chose event-driven PrEP at baseline (Figure 1). At PrEP initiation, the median age was 29 years (IQR 25-35), 81.23% (831/1023) of participants self-reported having a university degree or above, and 54.55% (558/1023) of participants self-identified as single.

**Figure 1.** Enrollment and follow up of the study participants. D→E: switching from daily to event-driven regimen; E→D: switching from event-driven to daily regimen; MSM: men who have sex with men.



### GSN Apps' Usage and Characteristics of GSN App Users

As much as 73.90% (756/1023) of eligible participants reported primarily using GSN apps to seek sexual partners in the past 3 months. GSN apps' users were younger and had higher education. A detailed distribution of sociodemographic characteristics is shown in Table 1.

Participants who use GSN apps were more likely to engage in risk behaviors than those not using GSN apps. Specifically, GSN apps' users had more male casual sexual partners ( $\geq 4$ : 34.9% [264/756] vs 21.7% [58/267],  $P<.001$ ) and more condomless anal intercourse (CAI; 1-2: 19.6% [148/756] vs 14.2% [38/267],  $\geq 3$ : 42.2% [319/756] vs 39.3% [105/267],  $P=.034$ ) than nonusers. However, there was no statistical difference between GSN apps' users and nonusers in the choice of medication regimens.

**Table 1.** Baseline characteristics of MSM on PrEP who did and did not report mainly seeking partners through geosocial networking apps during the past 3 months (N=1023).

Characteristics	GSN apps <sup>a</sup> nonuser (N=267), n (%)	GSN apps user (N=756), n (%)	$\chi^2$ (df)	P-value
<b>Age (years)</b>			6.33 (2)	.042
18-24	53 (19.9)	183 (24.2)		
25-34	128 (47.9)	387 (51.2)		
35-65	86 (32.2)	186 (24.6)		
Han Ethnic	239 (89.5)	680 (89.9)	0.04 (1)	.840
Local residents of research centers	249 (93.3)	707 (93.5)	0.02 (1)	.883
Having a university degree or above	200 (74.9)	631 (83.5)	9.48 (1)	.002
Average annual income $\geq$ 567 (US \$)	177 (66.3)	510 (67.5)	0.12 (1)	.727
<b>Marital status</b>			18.06 (1)	<.001
Married/Cohabitation with a woman	21 (7.9)	46 (6.1)		
Cohabitation with a man	121 (45.3)	251 (33.2)		
Single	116 (43.4)	442 (58.5)		
Separated, divorced, or widowed	9 (3.4)	17 (2.2)		
Ever received HIV PEP <sup>b</sup> in the past 6 months	12 (4.5)	53 (7.0)	2.10 (1)	.147
<b>PrEP<sup>c</sup> dosing regimens</b>			0.00 (1)	.968
Event-driven regimen	136 (50.9)	384 (50.8)		
Daily regimen	131 (49.1)	372 (49.2)		
Homosexual	198 (74.2)	605 (80.0)	4.03 (1)	.045
<b>Behaviors in the past 3 months</b>				
<b>Sexual roles with male</b>			6.69 (3)	.082
Oral	1 (0.4)	6 (0.8)		
Top	81 (30.3)	206 (27.2)		
Bottom	58 (21.7)	223 (29.5)		
Versatile	127 (47.6)	321 (42.5)		
<b>Number of male sex partners</b>			31.73 (2)	<.001
0-1	76 (28.5)	105 (13.9)		
2-3	86 (32.2)	246 (32.5)		
$\geq$ 4	105 (39.3)	405 (53.6)		
<b>Number of male casual sex partners</b>			38.70 (2)	<.001
0	130 (48.7)	213 (28.2)		
1-3	79 (29.6)	279 (36.9)		
$\geq$ 4	58 (21.7)	264 (34.9)		
<b>Number of CAI<sup>d</sup></b>			6.79 (2)	.034
0	124 (46.4)	289 (38.2)		
1-2	38 (14.2)	148 (19.6)		
$\geq$ 3	105 (39.3)	319 (42.2)		
Substance use	94 (35.2)	389 (51.5)	20.90 (1)	<.001
STI <sup>e</sup> symptoms	18 (6.7)	69 (9.1)	1.41 (1)	.236
Ever tested HIV in the past 12 months	245 (91.8)	718 (95.0)	3.69 (1)	.055



<sup>a</sup>GSN: geosocial networking.

<sup>b</sup>PEP: postexposure prophylaxis.

<sup>c</sup>PrEP: pre-exposure prophylaxis.

<sup>d</sup>CAI: condomless anal intercourse.

<sup>e</sup>STI: sexually transmitted infection.

### GSN Apps' Use, Adherence, and Regimen Switches for PrEP

By January 2020, 97.07% (993/1023) of participants completed at least one clinical visit, with a total of 2962 follow-up visits, and the median follow-up time was 0.50 person-years, equivalent to 513.07 person-years.

As shown in [Figure 2](#), according to adherence assessments by pill count and questionnaires at each visit, the proportion of adherence greater than 90% was lower and the rate of switching from event-driven to daily regimen was higher for GSN apps' users than nonusers. As shown in [Table 2](#), in the multivariable analysis of generalized estimation equation with logistic regression, GSN apps' use was also found to be a significantly independent predictor of adherence over 90% with adjusted odds ratio (aOR) of 0.8 (95% CI 0.6-1.0;  $P=.037$ ). In terms of the medication regimen, the adherence to the daily regimen was higher than that to the event-driven regimen (aOR 3.8, 95% CI 3.1-4.6). The subgroup analysis of different regimens showed that the adherence of GSN apps' users to the event-driven PrEP was lower than that of nonusers (aOR 0.7, 95% CI 0.6-0.9). Other independent predictors of adherence included age (vs 18-24 years; 25-34: aOR 1.3, 95% CI 1.0-1.7; 35-65 years: aOR 1.4, 95% CI 1.0-1.8), college education and above (aOR 1.3, 95% CI 1.0-1.7), daily regimen (vs event-driven regimen; aOR 3.8, 95% CI 3.1-4.6), having multiple male sex partners (vs 0-1

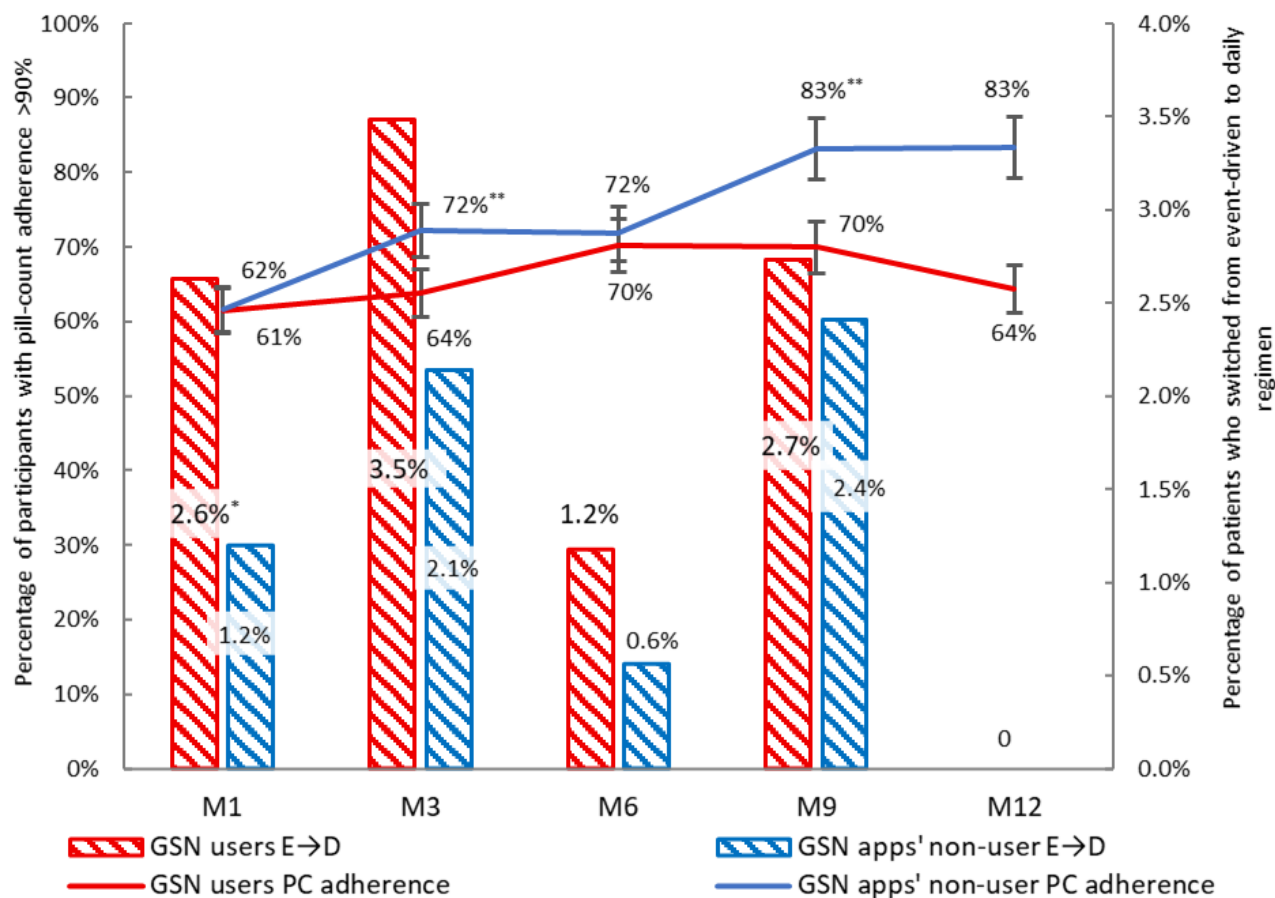
male sex partners; 2-3: aOR 1.2, 95% CI 1.0-1.5;  $\geq 4$ : aOR 1.3, 95% CI 1.1-1.7), multiple casual male sex partners (vs 0; 1-3: aOR 1.1, 95% CI 0.9-1.3;  $\geq 4$ : aOR 1.3, 95% CI 1.0-1.6), number of CAI instances of 3 or more (vs no CAI; aOR 1.2, 95% CI 1.0-1.5), and substance use (vs no substance use; aOR 0.7, 95% CI 0.5-0.9; [Table 2](#)).

For self-reported adherence measures, GSN apps' users showed a statistically marginally lower level of adherence than nonusers in the event-driven regimen (aOR 0.7, 95% CI 0.4-1.0,  $P=.060$ ). PrEP coverage of last sex events was significantly lower with GSN apps' users than with nonusers on the event-driven regimen (aOR 0.6, 95% CI 0.4-1.0,  $P=.038$ ; [Table 3](#)).

As much as 9.7% (96/993) of PrEP users switched regimens, including 82 who switched once and 14 who switched more than once. Of the 111 regimen-switching events, 67 (60.4%) were from event-driven to the daily regimen, and 44 (39.6%) were from event-driven to the daily regimen. Interestingly, based on the above multivariable analysis, GSN apps' users were more likely to switch regimens from event-driven to daily use, compared with nonusers (aOR 1.8, 95% CI 0.9-3.3,  $P=.084$ ; [Table 3](#)).

Additionally, 4.40% (45/1023) of participants discontinued PrEP and dropped out of the study, and there were no significant differences in rates of discontinuation between GSN apps' users and nonusers (4.4% [23/520] vs 4.4% [22/503],  $P=.796$ ).

**Figure 2.** Pill-counting adherence and regimen switches for PrEP among GSN users and nonusers during follow-up. The difference in the rate for pill counting adherence and regimen switching between GSN users and nonusers at each follow-up visit was marked as  $*P<.20$ ,  $**P<.05$ . E→D: Switching from event-driven to daily regimen; GSN: geosocial networking; PC: pill counting; PrEP: pre-exposure prophylaxis.



**Table 2.** Generalized estimated equation analysis of behaviors during follow-up as predictors of over 90% adherence to pre-exposure prophylaxis by pill counting (N=993).<sup>a</sup>

Factors	Total, n (%)	Adherence >90%, n (%)	Multivariate analysis aOR <sup>b</sup> (95% CI)	P-value
<b>Age (years)</b>				
18-24	230 (23.2)	146 (63.5)	Reference	
25-34	501 (50.5)	333 (66.5)	1.3 (1.0 to 1.7)	.028
35-65	262 (26.4)	179 (68.3)	1.3 (1.0 to 1.8)	.036
<b>Education</b>				
High school or below	175 (17.6)	113 (64.6)	Reference	
College and above	818 (82.4)	545 (66.6)	1.3 (1.0 to 1.7)	.037
<b>Sexual role with male</b>				
Oral	57 (5.7)	30 (52.6)	Reference	
Top	341 (34.3)	233 (68.3)	2.0 (1.4 to 2.8)	<.001
Bottom	279 (28.1)	193 (69.2)	2.1 (1.5 to 3.0)	<.001
Versatile	316 (31.8)	202 (63.9)	1.8 (1.2 to 2.6)	.002
<b>Main venue to seek male sex partners</b>				
Others	256 (25.8)	181 (70.7)	Reference	
GSN apps <sup>c</sup>	737 (74.2)	477 (64.7)	0.8 (0.6 to 1.0)	.037
<b>Sexual behaviors in past 3 months</b>				
<b>Number of male sex partners</b>				
0-1	303 (30.5)	188 (62.0)	Reference	
2-3	336 (33.8)	237 (70.5)	1.2 (1.0 to 1.5)	.050
≥4	354 (35.6)	233 (65.8)	1.3 (1.1 to 1.7)	.012
<b>Number of male casual sex partners</b>				
0	474 (47.7)	309 (65.2)	Reference	
1-3	283 (28.5)	198 (70.0)	1.1 (0.9 to 1.3)	.289
≥4	236 (23.8)	151 (64.0)	1.3 (1.0 to 1.6)	.054
<b>Number of CAI<sup>d</sup></b>				
0	522 (52.6)	343 (65.7)	Reference	
1-2	135 (13.6)	85 (63.0)	1.0 (0.8 to 1.2)	.816
≥3	336 (33.8)	230 (68.5)	1.2 (1.0 to 1.5)	.041
<b>Substance use</b>				
No	559 (56.3)	514 (91.9)	Reference	
Yes	434 (43.7)	401 (92.4)	0.7 (0.5 to 0.9)	.014
<b>PrEP<sup>e</sup> dosing regimens</b>				
Event-driven regimen	466 (46.9)	245 (52.6)	Reference	
Daily regimen	527 (53.1)	413 (78.4)	3.8 (3.1 to 4.6)	<.001

<sup>a</sup>All estimates were from generalized estimating equation logistic models, adjusted for ethnic, local residents, and annual income.<sup>b</sup>aOR: adjusted odds ratio.<sup>c</sup>GSN: geosocial networking.<sup>d</sup>CAI: condomless anal intercourse.<sup>e</sup>PrEP: pre-exposure prophylaxis.

**Table 3.** Adherence to pre-exposure prophylaxis and change of regimen among men who have sex with men that are geosocial networking app users and nonusers (N=993).<sup>a</sup>

	Total (GSN apps <sup>b</sup> user vs nonuser)		Daily regimen (GSN apps' user vs nonuser)		Event-driven regimen (GSN apps' user vs nonuser)	
	aOR <sup>c</sup> (95% CI)	P-value	aOR (95% CI)	P-value	aOR (95% CI)	P-value
Pill counting (vs adherence ≤90%)	0.8 (0.6 to 1.0)	.037	0.8 (0.6 to 1.1)	.205	0.7 (0.6 to 0.9)	.019
Self-report pills used as recommended (vs No)	0.9 (0.7 to 1.1)	.224	1.0 (0.7 to 1.3)	.968	0.7 (0.4 to 1.0)	.060
Self-report last sex events fully covered by PrEP <sup>d</sup> (vs No)	0.7 (0.5 to 1.0)	.057	1.1 (0.6 to 2.0)	.869	0.6 (0.4 to 1.0)	.038
Ever switched regimens (vs continuing initial regimen)	1.5 (0.9 to 2.6)	.161	1.7 (0.9 to 3.3)	.104	1.2 (0.5 to 2.6)	.681
Event-driven to daily PrEP	1.8 (0.9 to 3.3)	.084	1.7 (0.9 to 3.3)	.104	—	—
Daily to event-driven PrEP	1.2 (0.6 to 2.6)	.633	—	—	1.2 (.5 to 2.9)	.681

<sup>a</sup>All estimates are from generalized estimating equation logistic models, adjusted for ethnic, local residents, and annual income.

<sup>b</sup>GSN: geosocial networking.

<sup>c</sup>aOR: adjusted odds ratio.

<sup>d</sup>PrEP: pre-exposure prophylaxis.

## Discussion

### Principal Findings

These study results represented, to our best knowledge, the first evidence on the correlation between GSN apps' use and PrEP adherence in a real-world setting with options for daily or event-driven regimens to address bottlenecks in order to optimize PrEP adherence. Our results revealed that a high proportion of PrEP users mainly seek sexual partners through GSN apps, and GSN apps' users are at a higher sexual risk than other PrEP users. Worryingly, GSN apps' users had lower levels of medication adherence and higher rates of regimen switching during follow-up than GSN apps' nonusers. These data suggest the importance of reaching MSM at high risk through GSN platforms and providing them with education and interventions to improve PrEP adherence.

In this study, nearly three-quarters of MSM on PrEP reported seeking sexual partners mainly via GSN apps at baseline, which was higher than those reported in the general MSM population (56.0%) [2]. A study on American MSM found that those using GSN apps seemed more likely to start PrEP (aOR 1.6) compared with non-GSN apps' users, which seems to partly explain the high level of GSN apps' usage among MSM on PrEP [15]. Additionally, our results are similar to those from previous studies that found a higher proportion of HIV-related high-risk sexual behaviors among GSN apps' users compared to nonusers [4,7,24,25]. GSN apps offer many advantages not available with other traditional ways of encountering a sexual partner and making it easier to locate and screen via the app itself [4,7,24].

Medication adherence and regimen switching are important factors affecting the effectiveness of PrEP; however, there has been a lack of evidence on the effect of GSN apps' use on the above outcome variables. Our results found that although GSN apps' users were at a higher sexual risk than non-GSN apps' users, they had lower level of medication adherence to PrEP,

and higher rate of regimen switches during follow-up. The observed decrease in PrEP adherence among GSN apps' users could be attributed to their higher frequent substance using behaviors than nonusers. Previous research had demonstrated that substance use with sex is strongly associated with poor adherence to HIV treatments [26]. Additionally, this study observed a potential decline in the adherence of GSN users during the 6-12-month study period, which may be explained by the higher proportion of younger MSM among GSN users compared with nonusers. Our results also found a lower PrEP coverage of sex events among GSN apps' users compared with nonusers to further support this hypothesis. Furthermore, this study found that GSN apps' users were prone to transition from event-driven PrEP to daily dosing regimen than nonusers, although there was no statistically significant difference between the 2 groups in the choice of regimen at initiation. Additional work is needed to address the high rate of regimen switching among GSN apps' users, and response recommendations should be added to the WHO and country PrEP guidelines that recommend MSM currently using GSN apps to choose a more appropriate daily regimen based on their active sexual behavior. Daily dosing regimens are recommended by guidelines and some trial evidence for MSM with more frequent and irregular sexual behaviors due to the regimen's simplicity and ease of operation [17,27,28]. Given that GSN apps' users have a high conversion rate for medication regimens, they should be provided with timely education on new medication regimens, not only when PrEP is initiated, but also when they switch regimens. Because of the low dropout rate in this study, no significant difference was found in the termination rate of PrEP between GSN users and nonusers, and the outcome remains to be further observed and confirmed in a larger population with low cohort retention rate in the future.

Given that the GSN platforms cover most MSM using PrEP with high risk levels but low adherence, adherence intervention through eHealth on the GSN platforms may be an ideal way to

optimize adherence in the large-scale implementation of PrEP. The soaring number of dating app users offers unique opportunities to reach out to at-risk MSM for a broad range of health education and interventions [20,29]. GSN app-based advocacy has proven effective in expanding HIV testing and prevention interventions recruitment [30-35]. However, how to effectively deliver eHealth of PrEP adherence on GSN apps needs to be further explored [36]. Although banner ads in apps may be expensive [29,37,38], they can be used to effectively educate the public about PrEP and raise awareness, such as about different PrEP regimens. Pop-up ads in apps appear to be suitable for use as a just-in-time adherence intervention that can be set up on the dating apps to improve PrEP medication adherence [39]. A study conducted in the United States suggests that detailed user profiles on social platforms could help increase sexual health awareness of MSM [40]. This experience could be extended to reduce HIV high-risk behaviors of PrEP users by adding this setting to the GSN platform in China and other countries with similar HIV rates. Offline adherence support also must provide specific assistance to MSM with GSN apps' usage by providing counseling on medication regimens and switching regimens during PrEP initiation to minimize regimen switching.

### Strengths and Limitations

A major strength of this study was the assessment of adherence and regimen switches among MSM who reported using GSN apps, which increased our awareness of the correlation between GSN apps' use and adherence to PrEP. Furthermore, this study was based on a national multicenter real-world study that simulated real-world scenarios in which participants were able to choose their regimens and to switch between daily and event-driven dosing regimens if desired. However, there were also several limitations in this study. The follow-up time was

relatively short, and long-term outcomes may not be observed. Medication adherence may be overestimated to some extent due to the free availability of the PrEP and certain incentives in this demonstration study. This study did not further analyze the effect of GSN apps on HIV-prevention efficacy and drug concentration for PrEP, which future studies should examine. Although this study tries to simulate real-world scenarios as much as possible, compliance may be overestimated to some extent due to the free drug supply and the limited amount of incentive costs used to motivate PrEP use, as PrEP is not yet available in China. This study used a variety of methods to supervise the return of medicines, and pill counting remains inevitably affected by uncontrollable reasons such as forgetting to take them. Therefore, our study also uses questionnaires to comprehensively evaluate PrEP adherence from multiple perspectives. Although this study used flexible online and offline recruitment methods to reach all levels of MSM, similar to previous studies, the very high proportion of MSM with university degrees who start PrEP may not be representative of the entire MSM population in China. Therefore, further research is necessary to examine the causality between app use and the PrEP adherence of MSM.

### Conclusions

This study found that GSN apps' users who account for a large proportion of daily and event-driven oral PrEP users have lower adherence and are more likely to switch regimens despite their higher sexual risk than GSN apps' nonusers. Based on these results, GSN platforms provide a unique opportunity to reach high-risk MSM for promoting PrEP outreach, and it is also necessary to customize medication adherence interventions for GSN apps' users to improve the preventive effect of PrEP.

### Acknowledgments

This work was supported by the Mega-Projects of National Science Research (13th Five-Year Plan [2017ZX10201101]) and the National Natural Science Foundation of China (Grant No. 81872674). The authors are very grateful to all members of the CROPrEP study team for their contribution. Members not included in the authors' list include Yao Li, Fang Zhao, Yonghui Zhang, Zhu Mei, Yueru Jia, Rui Li, Yanni Ma, Xin Ma, and Xiaoyun Shi. We also thank the MSM who participated in this study and the staff of community-based organizations who contributed to its success. Gilead Sciences donated the research drug and is gratefully acknowledged.

### Authors' Contributions

HW, JX, and HS conceived and designed the study; HW, JX, JZ, ZC, QH, XH, YC, HW, XH, LZ, ZH, RB, SL, HL, XJ, HD, WG, and YJ performed the study and experiments; HW and JX analyzed the data; HW, JX, and SC drew the figures and tables; HW, JX, HS, and WD wrote and revised the manuscript; and JX and HS contributed equally as corresponding authors. All authors reviewed and approved the final manuscript.

### Conflicts of Interest

None declared.

### References

1. UNAIDS DATA 2019. 2019. URL: <https://www.unaids.org/en/resources/documents/2019/> [accessed 2020-06-05]
2. Duncan DT, Park SH, Hambrick HR, Dangerfield Ii DT, Goedel WC, Brewer R, et al. Characterizing Geosocial-Networking App Use Among Young Black Men Who Have Sex With Men: A Multi-City Cross-Sectional Survey in the Southern United States. *JMIR Mhealth Uhealth* 2018 Jun 14;6(6):e10316 [FREE Full text] [doi: [10.2196/10316](https://doi.org/10.2196/10316)] [Medline: [29903702](https://pubmed.ncbi.nlm.nih.gov/29903702/)]



3. Wang H, Zhang L, Zhou Y, Wang K, Zhang X, Wu J, et al. The use of geosocial networking smartphone applications and the risk of sexually transmitted infections among men who have sex with men: a systematic review and meta-analysis. *BMC Public Health* 2018 Oct 16;18(1):1178 [FREE Full text] [doi: [10.1186/s12889-018-6092-3](https://doi.org/10.1186/s12889-018-6092-3)] [Medline: [30326887](#)]
4. Beymer MR, Weiss RE, Bolan RK, Rudy ET, Bourque LB, Rodriguez JP, et al. Sex on demand: geosocial networking phone apps and risk of sexually transmitted infections among a cross-sectional sample of men who have sex with men in Los Angeles County. *Sex Transm Infect* 2014 Nov;90(7):567-572 [FREE Full text] [doi: [10.1136/sextrans-2013-051494](https://doi.org/10.1136/sextrans-2013-051494)] [Medline: [24926041](#)]
5. Contesse MG, Fredericksen RJ, Wohlfeiler D, Hecht J, Kachur R, Strona FV, et al. Attitudes About the Use of Geosocial Networking Applications for HIV/STD Partner Notification: A Qualitative Study. *AIDS Educ Prev* 2019 Jun;31(3):273-285. [doi: [10.1521/aeap.2019.31.3.273](https://doi.org/10.1521/aeap.2019.31.3.273)] [Medline: [31145001](#)]
6. Macapagal K, Moskowitz DA, Li DH, Carrión A, Bettin E, Fisher CB, et al. Hookup App Use, Sexual Behavior, and Sexual Health Among Adolescent Men Who Have Sex With Men in the United States. *Journal of Adolescent Health* 2018 Jun;62(6):708-715. [doi: [10.1016/j.jadohealth.2018.01.001](https://doi.org/10.1016/j.jadohealth.2018.01.001)]
7. Xu J, Yu H, Tang W, Leuba SI, Zhang J, Mao X, et al. The Effect of Using Geosocial Networking Apps on the HIV Incidence Rate Among Men Who Have Sex With Men: Eighteen-Month Prospective Cohort Study in Shenyang, China. *J Med Internet Res* 2018 Dec 21;20(12):e11303 [FREE Full text] [doi: [10.2196/11303](https://doi.org/10.2196/11303)] [Medline: [30578225](#)]
8. Regulatory Status of TDF/FTC for PrEP. AVAC. 2019. URL: <https://www.avac.org/infographic/> [accessed 2020-06-05]
9. Desai M, Field N, Grant R, McCormack S. Recent advances in pre-exposure prophylaxis for HIV. *BMJ* 2017 Dec 11;359:j5011 [FREE Full text] [doi: [10.1136/bmj.j5011](https://doi.org/10.1136/bmj.j5011)] [Medline: [29229609](#)]
10. Huang X, Hou J, Song A, Liu X, Yang X, Xu J, et al. Efficacy and Safety of Oral TDF-Based Pre-exposure Prophylaxis for Men Who Have Sex With Men: A Systematic Review and Meta-Analysis. *Front Pharmacol* 2018;9:799 [FREE Full text] [doi: [10.3389/fphar.2018.00799](https://doi.org/10.3389/fphar.2018.00799)] [Medline: [30233355](#)]
11. Riddell J, Amico KR, Mayer KH. HIV Preexposure Prophylaxis: A Review. *JAMA* 2018 Mar 27;319(12):1261-1268. [doi: [10.1001/jama.2018.1917](https://doi.org/10.1001/jama.2018.1917)] [Medline: [29584848](#)]
12. Anderson PL, Glidden DV, Liu A, Buchbinder S, Lama JR, Guanira JV, et al. Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men. *Sci Transl Med* 2012 Sep 12;4(151):151ra125. [doi: [10.1126/scitranslmed.3004006](https://doi.org/10.1126/scitranslmed.3004006)] [Medline: [22972843](#)]
13. Han J, Bouey JZ, Wang L, Mi G, Chen Z, He Y, et al. PrEP uptake preferences among men who have sex with men in China: results from a National Internet Survey. *J Int AIDS Soc* 2019 Feb;22(2):e25242 [FREE Full text] [doi: [10.1002/jia2.25242](https://doi.org/10.1002/jia2.25242)] [Medline: [30724470](#)]
14. Hampel B, Kusejko K, Braun DL, Harrison-Quintana J, Kouyos R, Fehr J. Assessing the need for a pre-exposure prophylaxis programme using the social media app Grindr®. *HIV Med* 2017 Nov;18(10):772-776 [FREE Full text] [doi: [10.1111/hiv.12521](https://doi.org/10.1111/hiv.12521)] [Medline: [28544123](#)]
15. Hoenigl M, Little SJ, Grelotti D, Skaathun B, Wagner GA, Weibel N, et al. Grindr™ Users Take more Risks, but are more Open to HIV Pre-Exposure Prophylaxis: Could this Dating App Provide Platform for HIV Prevention Outreach? *Clin Infect Dis* 2019 Nov 02;ciz1093 [FREE Full text] [doi: [10.1093/cid/ciz1093](https://doi.org/10.1093/cid/ciz1093)] [Medline: [31677383](#)]
16. Bourne A, Alba B, Garner A, Spiteri G, Pharris A, Noori T. Use of, and likelihood of using, HIV pre-exposure prophylaxis among men who have sex with men in Europe and Central Asia: findings from a 2017 large geosocial networking application survey. *Sex Transm Infect* 2019 May;95(3):187-192. [doi: [10.1136/sextrans-2018-053705](https://doi.org/10.1136/sextrans-2018-053705)] [Medline: [30612107](#)]
17. What's the 2+1+1? Technical Brief Event-Driven Oral Pre-Exposure Prophylaxis to Prevent HIV for Men Who Have Sex With Men: Update to WHO'S Recommendation on Oral Prep. 2019. URL: <https://apps.who.int/iris/bitstream/handle/10665/325955/WHO-CDS-HIV-198-engpdf?ua=1> [accessed 2020-06-05]
18. Hoornenborg E, Coyer L, Achterbergh RCA, Matser A, Schim van der Loeff MF, Boyd A, Amsterdam PrEP Project team in the HIV Transmission Elimination AMsterdam (H-TEAM) Initiative. Sexual behaviour and incidence of HIV and sexually transmitted infections among men who have sex with men using daily and event-driven pre-exposure prophylaxis in AMPREP: 2 year results from a demonstration study. *Lancet HIV* 2019 Jul;6(7):e447-e455. [doi: [10.1016/S2352-3018\(19\)30136-5](https://doi.org/10.1016/S2352-3018(19)30136-5)] [Medline: [31178284](#)]
19. Marshall BDL, Goedel WC, King MRF, Singleton A, Durham DP, Chan PA, et al. Potential effectiveness of long-acting injectable pre-exposure prophylaxis for HIV prevention in men who have sex with men: a modelling study. *Lancet HIV* 2018 Sep;5(9):e498-e505 [FREE Full text] [doi: [10.1016/S2352-3018\(18\)30097-3](https://doi.org/10.1016/S2352-3018(18)30097-3)] [Medline: [29908917](#)]
20. Holloway IW, Dougherty R, Gildner J, Beougher SC, Pulsipher C, Montoya JA, et al. Brief Report. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2017;74(1):15-20. [doi: [10.1097/qai.0000000000001164](https://doi.org/10.1097/qai.0000000000001164)]
21. Wang H, Zhang Y, Mei Z, Jia Y, Leuba SI, Zhang J, et al. Protocol for a multicenter, real-world study of HIV pre-exposure prophylaxis among men who have sex with men in China (CROPrEP). *BMC Infect Dis* 2019 Aug 15;19(1):721 [FREE Full text] [doi: [10.1186/s12879-019-4355-y](https://doi.org/10.1186/s12879-019-4355-y)] [Medline: [31416439](#)]
22. Hanley J, Negassa A, Edwardes M, Forrester J. Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am J Epidemiol* 2003 Feb 15;157(4):364-375. [doi: [10.1093/aje/kwf215](https://doi.org/10.1093/aje/kwf215)] [Medline: [12578807](#)]
23. Zeger SL, Liang K, Albert PS. Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics* 1988 Dec;44(4):1049-1060. [doi: [10.2307/2531734](https://doi.org/10.2307/2531734)] [Medline: [3233245](#)]

24. Eaton LA, Maksut JL, Gamarel KE, Siembida EJ, Driffin DD, Baldwin R. Online Sex Partner Meeting Venues as a Risk Factor for Testing HIV Positive Among a Community-Based Sample of Black Men Who Have Sex With Men. *Sex Transm Dis* 2016 Jun;43(6):360-364. [doi: [10.1097/OLQ.0000000000000454](https://doi.org/10.1097/OLQ.0000000000000454)] [Medline: [27200520](https://pubmed.ncbi.nlm.nih.gov/27200520/)]
25. Mustanski BS. Are sexual partners met online associated with HIV/STI risk behaviours? Retrospective and daily diary data in conflict. *AIDS Care* 2007 Jun 14;19(6):822-827. [doi: [10.1080/09540120701237244](https://doi.org/10.1080/09540120701237244)]
26. Zhang Y, Wilson TE, Adedimeji A, Merenstein D, Milam J, Cohen J, et al. The Impact of Substance Use on Adherence to Antiretroviral Therapy Among HIV-Infected Women in the United States. *AIDS Behav* 2017 May 30;22(3):896-908. [doi: [10.1007/s10461-017-1808-4](https://doi.org/10.1007/s10461-017-1808-4)] [Medline: [28560499](https://pubmed.ncbi.nlm.nih.gov/28560499/)]
27. Antoni G, Tremblay C, Delaunay C, Charreau I, Cua E, Rojas Castro D, ANRS IPERGAY study group. On-demand pre-exposure prophylaxis with tenofovir disoproxil fumarate plus emtricitabine among men who have sex with men with less frequent sexual intercourse: a post-hoc analysis of the ANRS IPERGAY trial. *Lancet HIV* 2020 Feb;7(2):e113-e120. [doi: [10.1016/S2352-3018\(19\)30341-8](https://doi.org/10.1016/S2352-3018(19)30341-8)] [Medline: [31784343](https://pubmed.ncbi.nlm.nih.gov/31784343/)]
28. Bekker L, Roux S, Sebastien E, Yola N, Amico KR, Hughes JP, HPTN 067 (ADAPT) study team. Daily and non-daily pre-exposure prophylaxis in African women (HPTN 067/ADAPT Cape Town Trial): a randomised, open-label, phase 2 trial. *Lancet HIV* 2018 Feb;5(2):e68-e78 [FREE Full text] [doi: [10.1016/S2352-3018\(17\)30156-X](https://doi.org/10.1016/S2352-3018(17)30156-X)] [Medline: [28986029](https://pubmed.ncbi.nlm.nih.gov/28986029/)]
29. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of Smartphone Application-Based HIV Prevention Among Young Men Who Have Sex With Men. *AIDS Behav* 2014 Feb 18;18(2):285-296. [doi: [10.1007/s10461-013-0671-1](https://doi.org/10.1007/s10461-013-0671-1)] [Medline: [24292281](https://pubmed.ncbi.nlm.nih.gov/24292281/)]
30. Nikolopoulos GK, Pavlitina E, Muth SQ, Schneider J, Psychogiou M, Williams LD, et al. A network intervention that locates and intervenes with recently HIV-infected persons: The Transmission Reduction Intervention Project (TRIP). *Sci Rep* 2016 Dec 05;6:38100 [FREE Full text] [doi: [10.1038/srep38100](https://doi.org/10.1038/srep38100)] [Medline: [27917890](https://pubmed.ncbi.nlm.nih.gov/27917890/)]
31. Smyrnov P, Williams LD, Korobchuk A, Sazonova Y, Nikolopoulos GK, Skaathun B, et al. Risk network approaches to locating undiagnosed HIV cases in Odessa, Ukraine. *J Int AIDS Soc* 2018 Jan;21(1):e25040 [FREE Full text] [doi: [10.1002/jia2.25040](https://doi.org/10.1002/jia2.25040)] [Medline: [29356365](https://pubmed.ncbi.nlm.nih.gov/29356365/)]
32. Sun CJ, Stowers J, Miller C, Bachmann LH, Rhodes SD. Acceptability and feasibility of using established geosocial and sexual networking mobile applications to promote HIV and STD testing among men who have sex with men. *AIDS Behav* 2015 Mar;19(3):543-552 [FREE Full text] [doi: [10.1007/s10461-014-0942-5](https://doi.org/10.1007/s10461-014-0942-5)] [Medline: [25381563](https://pubmed.ncbi.nlm.nih.gov/25381563/)]
33. Burrell ER, Pines HA, Robbie E, Coleman L, Murphy RD, Hess KL, et al. Use of the location-based social networking application GRINDR as a recruitment tool in rectal microbicide development research. *AIDS Behav* 2012 Oct;16(7):1816-1820 [FREE Full text] [doi: [10.1007/s10461-012-0277-z](https://doi.org/10.1007/s10461-012-0277-z)] [Medline: [22851153](https://pubmed.ncbi.nlm.nih.gov/22851153/)]
34. Alarcón Gutiérrez M, Fernández Quevedo M, Martín Valle S, Jacques-Aviñó C, Díez David E, Caylà JA, et al. Acceptability and effectiveness of using mobile applications to promote HIV and other STI testing among men who have sex with men in Barcelona, Spain. *Sex Transm Infect* 2018 Sep;94(6):443-448. [doi: [10.1136/sextrans-2017-053348](https://doi.org/10.1136/sextrans-2017-053348)] [Medline: [29626174](https://pubmed.ncbi.nlm.nih.gov/29626174/)]
35. Buttram ME, Kurtz SP. Representations of Pre-exposure Prophylaxis, Informal Antiretroviral Medication Use for HIV Prevention, and Diversion on Geosocial Networking Apps among Men Who Have Sex with Men. *AIDS Behav* 2019 Jul;23(7):1790-1796. [doi: [10.1007/s10461-018-2281-4](https://doi.org/10.1007/s10461-018-2281-4)] [Medline: [30255385](https://pubmed.ncbi.nlm.nih.gov/30255385/)]
36. Winchester III WW, Abel TD, Bauermeister J. The use of partner-seeking computer-mediated communication applications by young men that have sex with men (YMSM): uncovering human-computer interaction (HCI) design opportunities in HIV prevention. *Health Systems* 2017 Dec 19;1(1):26-35. [doi: [10.1057/hs.2012.3](https://doi.org/10.1057/hs.2012.3)]
37. Huang E, Marlin RW, Young SD, Medline A, Klausner JD. Using Grindr, a Smartphone Social-Networking Application, to Increase HIV Self-Testing Among Black and Latino Men Who Have Sex With Men in Los Angeles, 2014. *AIDS Educ Prev* 2016 Aug;28(4):341-350. [doi: [10.1521/aeap.2016.28.4.341](https://doi.org/10.1521/aeap.2016.28.4.341)] [Medline: [27427928](https://pubmed.ncbi.nlm.nih.gov/27427928/)]
38. Badal HJ, Stryker JE, DeLuca N, Purcell DW. Swipe Right: Dating Website and App Use Among Men Who Have Sex With Men. *AIDS Behav* 2017 Sep 7;22(4):1265-1272. [doi: [10.1007/s10461-017-1882-7](https://doi.org/10.1007/s10461-017-1882-7)] [Medline: [28884248](https://pubmed.ncbi.nlm.nih.gov/28884248/)]
39. Grov C, Rendina HJ, Jimenez R, Parsons JT. Using Online Settings to Identify Gay and Bisexual Men Willing to Take or With Experience Taking PrEP: Implications for Researchers and Providers. *AIDS Educ Prev* 2016 Oct;28(5):378-392 [FREE Full text] [doi: [10.1521/aeap.2016.28.5.378](https://doi.org/10.1521/aeap.2016.28.5.378)] [Medline: [27710085](https://pubmed.ncbi.nlm.nih.gov/27710085/)]
40. Hecht J. Detailed Profile Options on Dating Apps Increase Awareness about Sexual Health. International AIDS Conference 2020. 2020. URL: <https://cattendee.abstractsonline.com/meeting/9289/> [accessed 2020-08-06]

## Abbreviations

**aOR:** adjusted odds ratio  
**CAI:** condomless anal intercourse  
**CASI:** computer-assisted self-interview  
**GSN:** geosocial networking  
**MSM:** men who have sex with men  
**PrEP:** pre-exposure prophylaxis  
**STI:** sexually transmitted infection

**TDF/FTC:** tenofovir; disoproxil fumarate plus emtricitabine

*Edited by G Eysenbach; submitted 20.07.20; peer-reviewed by W Tang, J Hecht; comments to author 01.08.20; revised version received 12.08.20; accepted 15.09.20; published 14.10.20.*

*Please cite as:*

Wang H, Zhang J, Chu Z, Hu Q, Dong W, Huang X, Chen Y, Wang H, He X, Zhang L, Hu Z, Bao R, Li S, Li H, Cui S, Jin X, Ding H, Geng W, Jiang Y, Xu J, Shang H

*Risk-Taking Behaviors and Adherence to HIV Pre-Exposure Prophylaxis in Users of Geosocial Networking Apps: Real-World, Multicenter Study*

*J Med Internet Res* 2020;22(10):e22388

URL: <http://www.jmir.org/2020/10/e22388/>

doi: [10.2196/22388](https://doi.org/10.2196/22388)

PMID: [33052132](https://pubmed.ncbi.nlm.nih.gov/33052132/)

©Hongyi Wang, Jing Zhang, Zhenxing Chu, Qinghai Hu, Willa Dong, Xiaojie Huang, Yaokai Chen, Hui Wang, Xiaoqing He, Lukun Zhang, Zhili Hu, Rantong Bao, Shangcao Li, Hang Li, Sitong Cui, Xia Jin, Haibo Ding, Wenqing Geng, Yongjun Jiang, Junjie Xu, Hong Shang. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# A Data Visualization and Dissemination Resource to Support HIV Prevention and Care at the Local Level: Analysis and Uses of the AIDSVu Public Data Resource

Patrick Sean Sullivan<sup>1</sup>, DVM, PhD; Cory Woodyatt<sup>2</sup>, BSN, RN, CEN; Chelsea Koski<sup>3</sup>, BA; Elizabeth Pembleton<sup>1</sup>, MPH; Pema McGuinness<sup>4</sup>, MPH; Jennifer Taussig<sup>1</sup>, MPH; Alexandra Ricca<sup>1</sup>, MPH; Nicole Luisi<sup>1</sup>, MPH; Eve Mokotoff<sup>5</sup>, MPH; Nanette Benbow<sup>6</sup>, MAS; Amanda D Castel<sup>7</sup>, MD, MPH; Ann N Do<sup>1</sup>, MD, MPH; Ronald O Valdiserri<sup>1</sup>, MD, MPH; Heather Bradley<sup>8</sup>, PhD; Chandni Jaggi<sup>1</sup>, MPH; Daniel O'Farrell<sup>9</sup>, MS; Rebecca Filipowicz<sup>1</sup>, MPH, MS; Aaron J Siegler<sup>10</sup>, PhD; James Curran<sup>1</sup>, MD, MPH; Travis H Sanchez<sup>1</sup>, MPH, DVM

<sup>1</sup>Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States

<sup>2</sup>Oregon Health & Science University, Portland, OR, United States

<sup>3</sup>Signal Group, Washington, DC, United States

<sup>4</sup>Gilead Sciences, Foster City, CA, United States

<sup>5</sup>HIV Counts, Ann Arbor, MI, United States

<sup>6</sup>Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

<sup>7</sup>Department of Epidemiology, Milken Institute for Public Health, George Washington University, Washington, DC, United States

<sup>8</sup>Department of Epidemiology, School of Public Health, Georgia State University, Atlanta, GA, United States

<sup>9</sup>Maimonides Medical Center, New York, NY, United States

<sup>10</sup>Department of Behavioral Sciences and Health Education, Rollins School of Public Health, Emory University, Atlanta, GA, United States

## Corresponding Author:

Patrick Sean Sullivan, DVM, PhD

Department of Epidemiology

Rollins School of Public Health

Emory University

1518 Clifton Road NE

Atlanta, GA, 30322

United States

Phone: 1 4047272038

Email: [pssulli@emory.edu](mailto:pssulli@emory.edu)

## Abstract

**Background:** AIDSVu is a public resource for visualizing HIV surveillance data and other population-based information relevant to HIV prevention, care, policy, and impact assessment.

**Objective:** The site, AIDSVu.org, aims to make data about the US HIV epidemic widely available, easily accessible, and locally relevant to inform public health decision making.

**Methods:** AIDSVu develops visualizations, maps, and downloadable datasets using results from HIV surveillance systems, other population-based sources of information (eg, US Census and national probability surveys), and other data developed specifically for display and dissemination through the website (eg, pre-exposure prophylaxis [PrEP] prescriptions). Other types of content are developed to translate surveillance data into summarized content for diverse audiences using infographic panels, interactive maps, local and state fact sheets, and narrative blog posts.

**Results:** Over 10 years, AIDSVu.org has used an expanded number of data sources and has progressively provided HIV surveillance and related data at finer geographic levels, with current data resources providing HIV prevalence data down to the census tract level in many of the largest US cities. Data are available at the county level in 48 US states and at the ZIP Code level in more than 50 US cities. In 2019, over 500,000 unique users consumed AIDSVu data and resources, and HIV-related data and insights were disseminated through nearly 4,000,000 social media posts. Since AIDSVu's inception, at least 249 peer-reviewed publications have used AIDSVu data for analyses or referenced AIDSVu resources. Data uses have included targeting of HIV



testing programs, identifying areas with inequitable PrEP uptake, including maps and data in academic and community grant applications, and strategically selecting locations for new HIV treatment and care facilities to serve high-need areas.

**Conclusions:** Surveillance data should be actively used to guide and evaluate public health programs; AIDSvu translates high-quality, population-based data about the US HIV epidemic and makes that information available in formats that are not consistently available in surveillance reports. Bringing public health surveillance data to an online resource is a democratization of data, and presenting information about the HIV epidemic in more visual formats allows diverse stakeholders to engage with, understand, and use these important public health data to inform public health decision making.

(*J Med Internet Res* 2020;22(10):e23173) doi:[10.2196/23173](https://doi.org/10.2196/23173)

## KEYWORDS

HIV; surveillance; infodemiology; data visualization; infectious disease; health policy; data dashboard; health department data; dashboard; data

## Introduction

Public health surveillance is the basis of effective public health decision making [1], and surveillance data have been called “the conscience of an epidemic” [2]. Defined as “the ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice” [3], surveillance data have a critical role to play in monitoring and protecting the health of individuals and populations. More specifically, surveillance is critical to delineate patterns of disease and to evaluate population-level health data and the impact of public health interventions and policies.

Dissemination of data is a critical aspect of surveillance practice: making public health data available broadly helps to raise awareness of public health threats, promotes the efficient targeting of resources and testing programs, supports evidence-based legislation to provide for prevention and care services, and enables the monitoring of progress in disease prevention and care [4]. Historically, surveillance data have been disseminated through reports, published at least annually [5], and at varying geographic levels (ie, national, state, and local) [4]. More recently, the broad availability of the internet and increasing data transfer speeds have enabled the online dissemination of reports; in many health jurisdictions, surveillance reports are now only available online.

The US case-based surveillance system for HIV is one of the largest, most well-resourced, and comprehensive infectious disease surveillance systems in the world [6]. The US Centers for Disease Control and Prevention (CDC) administers funding for the system, which is implemented by all US states and selected US cities. The CDC is responsible for processing surveillance data and deduplicating records and for producing national surveillance reports; the CDC does not report or map HIV surveillance data at geographic levels finer than at the county level. State and local health departments are responsible for producing local surveillance reports for jurisdictions smaller than counties. Surveillance reports have traditionally largely comprised tabular presentations of data [7]. Typical surveillance reports include only summary data and are not necessarily geared to lay audiences without technical backgrounds. Data visualizations are generally lacking from surveillance reporting, although more recently, some state and local health departments have incorporated choropleth maps to illustrate variations in

levels of disease by smaller geographic areas within states [8] or cities [9,10]. Despite these welcome additions, the types of data released and formats of visualizations vary across states, and the methods used to develop them may also vary among states.

We describe the development, implementation, methods, governance, and impact of AIDSvu.org, a free public data source created to visualize and disseminate HIV surveillance data. Launched in 2009, the resource is a cooperative private-public-academic partnership. We also summarize the reach and utilization of AIDSvu and its service and data resources.

## Methods

### Overview of AIDSvu and Statement of Purpose

AIDSvu was created in 2009 as an interactive online mapping platform to visualize the impact of the HIV epidemic on communities across the United States and to improve access to surveillance data at granular geographic levels. The goals of AIDSvu are to make HIV-related data widely available, easily accessible, and locally relevant to inform public health decision making. Improved accessibility and use of HIV surveillance data promotes disease awareness among the general public, media professionals, and medical practitioners and informs data-driven public health decision making. Local relevance refers to publishing data at the finest possible geographic levels, so that public health decisions can be made with awareness of heterogeneity of needs within states or cities. The platform also promotes the wide availability of multiple types of data, including surveillance data, census data on social determinants, locations of HIV-related service providers, and data on the uptake of key HIV services (eg, pre-exposure prophylaxis [PrEP] and HIV testing). Researchers have used AIDSvu’s data and tools to conduct analyses and evaluate public health interventions; in addition, advocates and community leaders have used AIDSvu’s maps and other tools to educate their audiences and support funding applications.

### Principles of Data Inclusion

With assistance from an advisory committee of key stakeholders in HIV and a technical advisory group that includes HIV surveillance technical experts, AIDSvu’s scientists have established firm principles of data inclusion to guide the scope



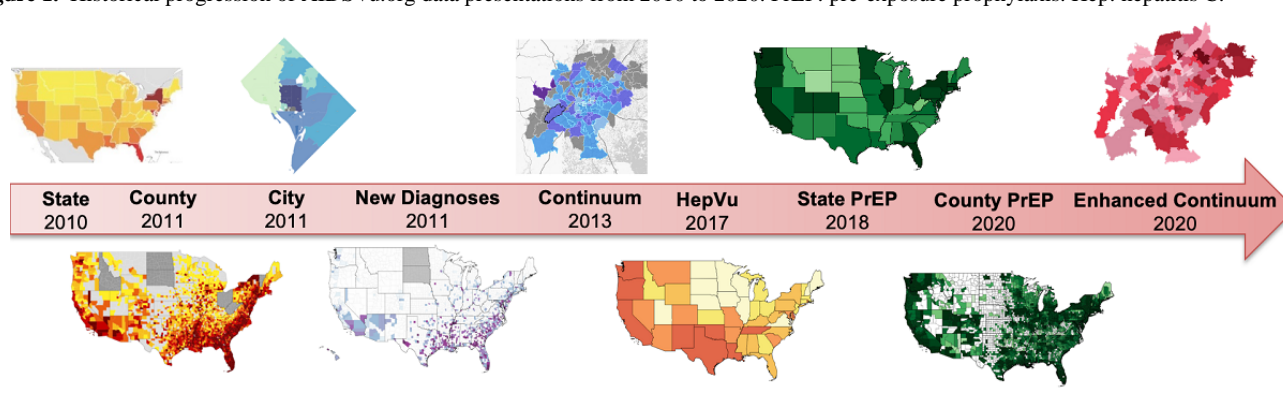
of data to be included on the site. First, all sources must be publicly available or fully transparent to allow replication. Second, all metrics must be consistently defined and standardized to enable meaningful comparisons across geographies. Third, data presented on the site are required to be substantially population based; for example, case-based surveillance data, other data sources that represent a census of outcomes (eg, PrEP utilization data from pharmacy sources), and population-based survey data. Fourth, data must be timely; AIDSvu's scientific team aims to provide data on the site as quickly as possible after they have been released and reviewed for data quality. Fifth, in recognition of privacy concerns and data suppression requirements set by health departments and the CDC, data are suppressed when indicated by either small numbers of cases or small populations (see [Multimedia](#)

[Appendix 1](#) for details). Within these criteria, the AIDSvu team continuously aims to release finer geographic levels of data to increase relevance to prevention and care activities in local communities.

### Overview of Data and Sources

AIDSvu presents HIV-related metrics from a variety of sources, including the US CDC, state and local health departments, health care claims databases, and public data such as US Census data. The elements presented on AIDSvu have expanded since the platform's inception. AIDSvu's first HIV surveillance map released in 2010 included only state-level data, and over time AIDSvu's team has added additional data elements at finer geographic levels, including county- and ZIP Code-level data for selected states and jurisdictions beginning in 2011 (see [Figure 1](#)).

**Figure 1.** Historical progression of AIDSvu.org data presentations from 2010 to 2020. PrEP: pre-exposure prophylaxis. Hep: hepatitis C.



The HIV surveillance data presented on AIDSvu are updated annually or as the data become publicly available from the data sources. HIV surveillance maps can be viewed by demographic breakdowns and can be overlaid with three types of service locations—PrEP, HIV testing, and HIV treatment—to enable examination of these resources relative to HIV prevalence and diagnosis data at different geographic levels [11].

Infographics and city and state profile pages provide supplementary visualizations and information, and complement

the data presented in the maps and existing surveillance reports. Other public data, such as US Census data on social determinants of health (ie, poverty, high school education, median household income, income inequality, health insurance, etc) can be viewed alongside interactive maps to provide context. [Table 1](#) displays the availability of data by visualization type [12,13]. [Table 2](#) displays the availability of data by geographic level.

**Table 1.** Availability of data by visualization type from the AIDSvu Project, July 2020.

Geographic level	Profiles	Prevalence	Diagnoses	Mortality	PrEP <sup>a</sup> use [12]	PrEP-to-need ratio [13]	SDH <sup>b</sup>	Continuum	Data sources
National	X <sup>c</sup>	X	X	X	X	X	X	X	CDC <sup>d</sup> and Symphony Health
Regional	X	X	X	X	X	X	X		CDC
State	X	X	X	X	X	X	X	Select states	CDC
County	X	X	X	X		X	X	Select counties	CDC
ZIP Code	X	X	X				X	Select cities	City and state health departments
3-digit ZIP Code					X				Symphony Health
Census tract, neighborhood, and community area		Select cities	Select cities						City and state health departments

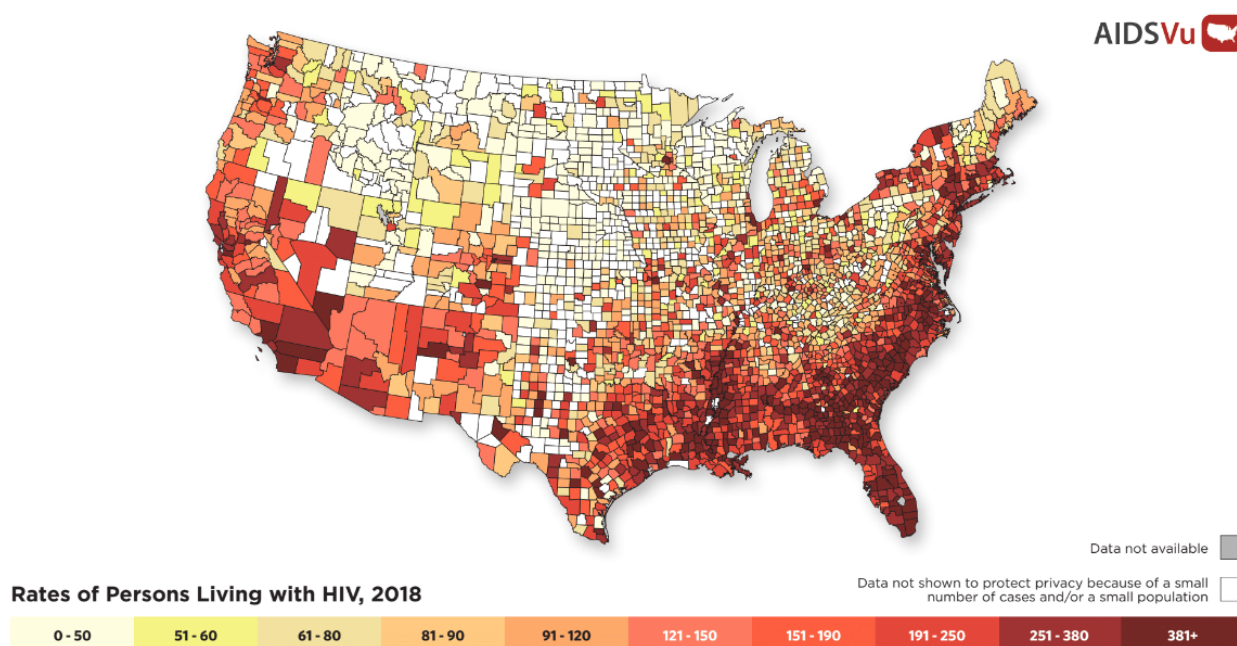
<sup>a</sup>PrEP: pre-exposure prophylaxis.<sup>b</sup>SDH: social determinants of health.<sup>c</sup>X: Data are presented on AIDSvu. Empty cells: data are not presented on AIDSvu.<sup>d</sup>CDC: Centers for Disease Control and Prevention.**Table 2.** Availability of data by geographic level from the AIDSvu Project, July 2020.

Available data	State	Region	County	City	ZIP Code
HIV prevalence	X <sup>a</sup>	X	X	Select cities	Select cities
HIV diagnoses	X	X	X	Select cities	Select cities
PrEP <sup>b</sup> use	X	X	X		
PrEP-to-need ratio	X	X	X		
Mortality	X	X	X	Select cities	
Social determinants of health	X	X	X	Select cities	
Testing	X		Select counties		
Late diagnoses	X	X		Select cities	Select cities
Linkage to care	X			Select cities	Select cities
Viral suppression	X			Select cities	Select cities
Receipt of care	X			Select cities	Select cities
Federal funding	X	X			
Sexually transmitted diseases	X	X	Select counties	X	

<sup>a</sup>X: Data are presented on AIDSvu. Empty cells: data are not presented on AIDSvu.<sup>b</sup>PrEP: pre-exposure prophylaxis.

AIDSvu currently maps HIV prevalence and new diagnoses data at state and county levels in addition to the ZIP Code level for 50 cities and community areas and wards, and maps them at the census tract level for a subset of three of those cities. At the county level, AIDSvu maps data on prevalence (see [Figure 2](#)) and new diagnoses plus PrEP use [12] and PrEP-to-need ratio [13]; at the state level, AIDSvu maps all of the prior indicators

plus HIV mortality data and HIV testing data (see [Table 1](#)). For US cities, data are available at the ZIP Code level for HIV prevalence and 5-year risk of HIV diagnosis for 45 cities; data are available for the HIV continuum (eg, late diagnosis, prompt entry to care, retention in care, and viral suppression) for 35 cities.

**Figure 2.** AIDSvu map of HIV prevalence at the county level, 2018.

Regarding the intended audiences, AIDSvu's content was developed to meet the needs of several types of users: high-interest users (eg, health department staff, academic researchers, advocacy organizations, and community-based organizations), medical providers (eg, practicing clinicians and clinic staff), policy makers, the media, and the public. Different areas of the site are targeted primarily at different combinations of user audiences; for example, state profile pages might be relevant for both high-interest users and the public, data downloads are of most relevance to researchers, and infographic panels are of high relevance to all groups.

### Governance and Organizational Structure

AIDSvu is the result of a unique partnership between the academic community; governmental organizations, including federal, state, and local health agencies; and private industry. AIDSvu is led by Emory University's Rollins School of Public Health in partnership with Gilead Sciences, Inc, and the Center for AIDS Research at Emory University. The leadership team of AIDSvu is composed of the Advisory Committee, the Technical Advisory Group (TAG), and the Prevention and Treatment Advisory Committee (PTAC), as well as a core team of staff who manage daily activities. The Advisory Committee meets annually and consists of key stakeholders who provide oversight and guidance for the project, including members from federal government agencies such as the US Department of Health and Human Services and the CDC, as well as state departments of health, academic research institutions, health-focused not-for-profit organizations, and industry. AIDSvu's technical advisory teams (ie, TAG and PTAC) consist of HIV surveillance experts who provide input regarding data and other technical issues and HIV prevention professionals who advise AIDSvu's team on how the tool can be utilized in

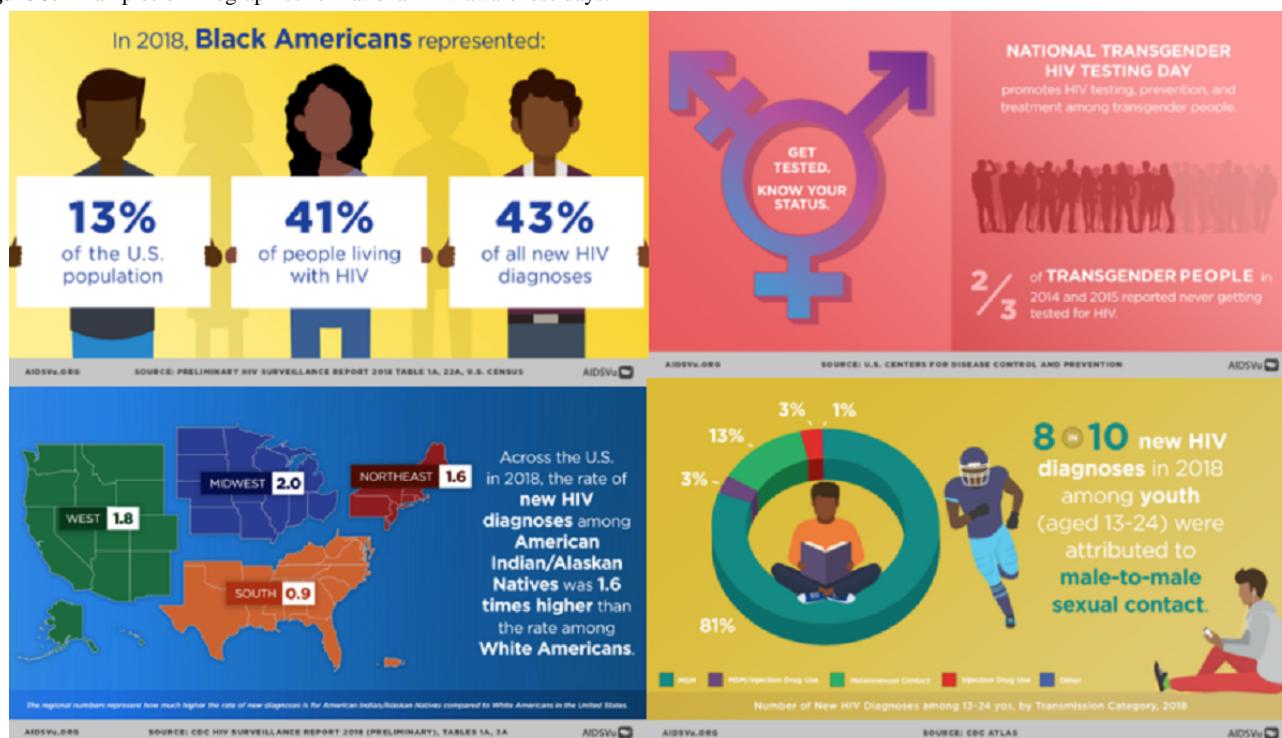
HIV prevention efforts. TAG and PTAC members meet at least annually and include representatives from federal government agencies, state and city departments of health, and health-focused not-for-profit organizations. This shared responsibility of AIDSvu among public, private, and academic institutions improves ongoing sustainability, agility, and relevance to various audiences.

### Other Content Types

AIDSvu data are presented to optimize utility for the intended audiences. To do this, AIDSvu's team provides a variety of content and visualizations, including infographics to support HIV awareness days, blog posts, downloadable map images, HIV service and clinic locators, side-by-side maps of social determinants of health data, state and local profile pages, and downloadable datasets. These have been used to support user presentations [14], research [15,16], grant applications [17], and the development of educational and advocacy materials that have been presented to stakeholders at conferences, such as the American Public Health Association, the US Conference on AIDS, the Texas HIV Prevention Conference, Fast-Track Cities, Adherence, and the International AIDS Conference.

### Awareness Day Infographics

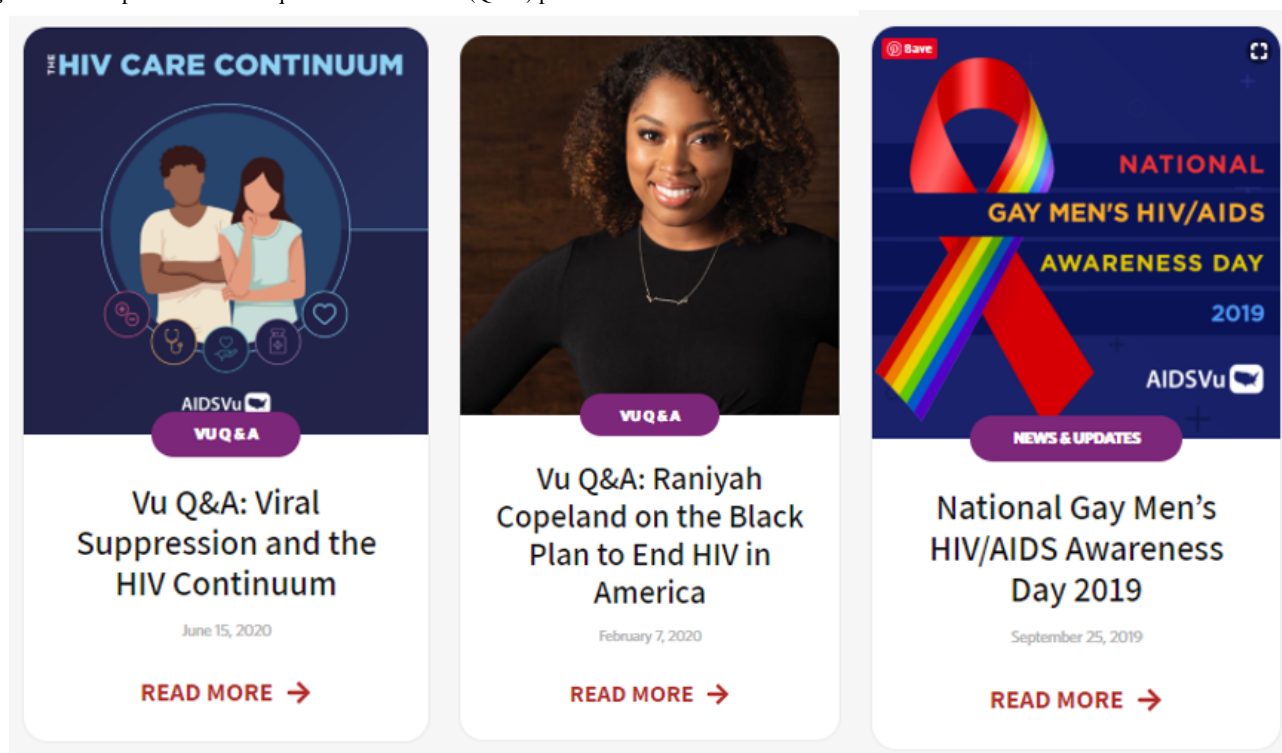
Annually, in observance of national HIV awareness days, AIDSvu releases a series of infographics highlighting key statistics for various populations, including gay men, women and girls, Black people, Asian and Pacific Islander populations, Native people, Latinx people, Youth, and Transgender people. These are made available to the public via Facebook, Twitter, and AIDSvu.org. The data that inform these graphics are calculated from data available on AIDSvu and from the CDC (see Figure 3).

**Figure 3.** Examples of infographics for national HIV awareness days.

## Blog Posts

Throughout the course of the year, AIDSVu develops and publishes blog posts on various topics, including HIV awareness days, the release of new data or a manuscript, a question-and-answer session on a particular issue (eg, Medicaid

expansion and health inequities in the South), and years in review. Since 2010, AIDSVu has published over 160 blog posts to the website, many of which have featured some of the leading experts in HIV surveillance, research, and advocacy (see [Figure 4](#)).

**Figure 4.** Examples of AIDSVu question-and-answer (Q&A) posts.

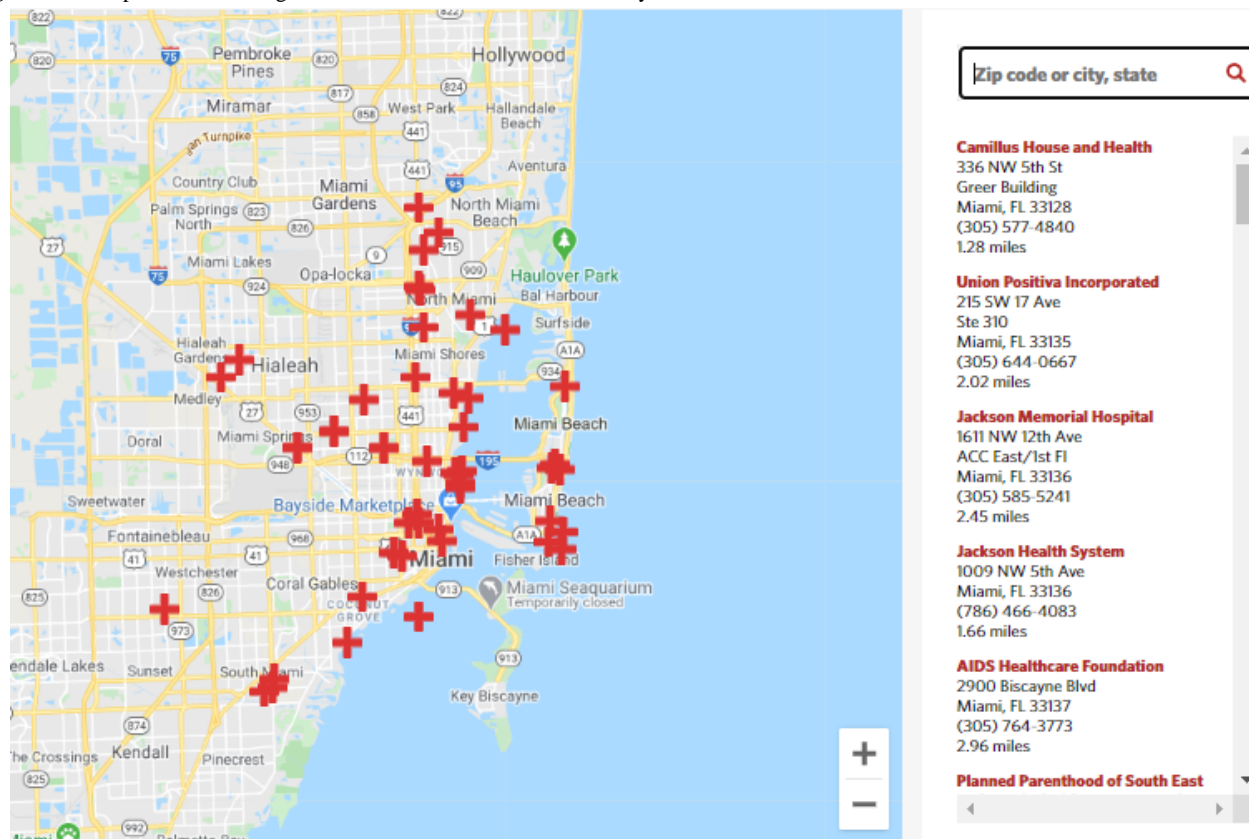


## Service and Clinic Locators

AIDSVu provides users with service locators for HIV prevention, testing, and care. Data for testing, PrEP [18], and care services are obtained from the National Prevention

Information Network through an application programming interface, so these data are as current as possible. In addition, for the nine states in the Deep South, users can locate services for stigma reduction, overdose prevention and reversal, harm reduction, and trauma-informed care (see Figure 5).

**Figure 5.** Example of HIV testing service locator for Miami Dade County.



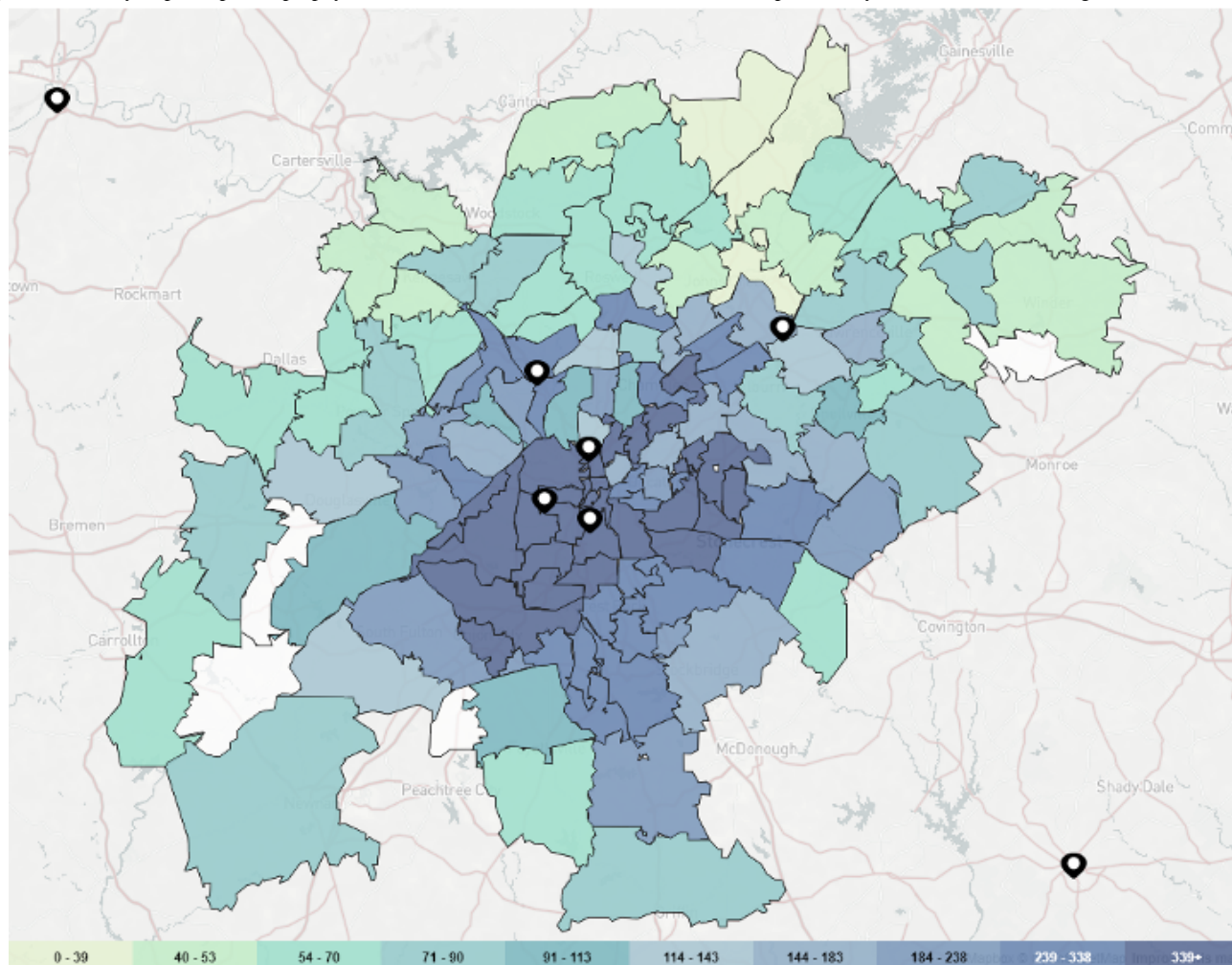
## Overlays

Users are also able to overlay service provider locations directly onto AIDSVu's maps to illustrate how services are distributed in relation to the burden of HIV. The services available for overlay include HIV testing sites, PrEP services, Ryan White HIV/AIDS Program medical care providers, the Housing Opportunities for Persons with AIDS program, National Institutes of Health (NIH)-funded HIV Prevention Trials Network sites, NIH-funded HIV Vaccine Trials Network sites, and NIH-funded HIV treatment trial sites (see Figure 6).

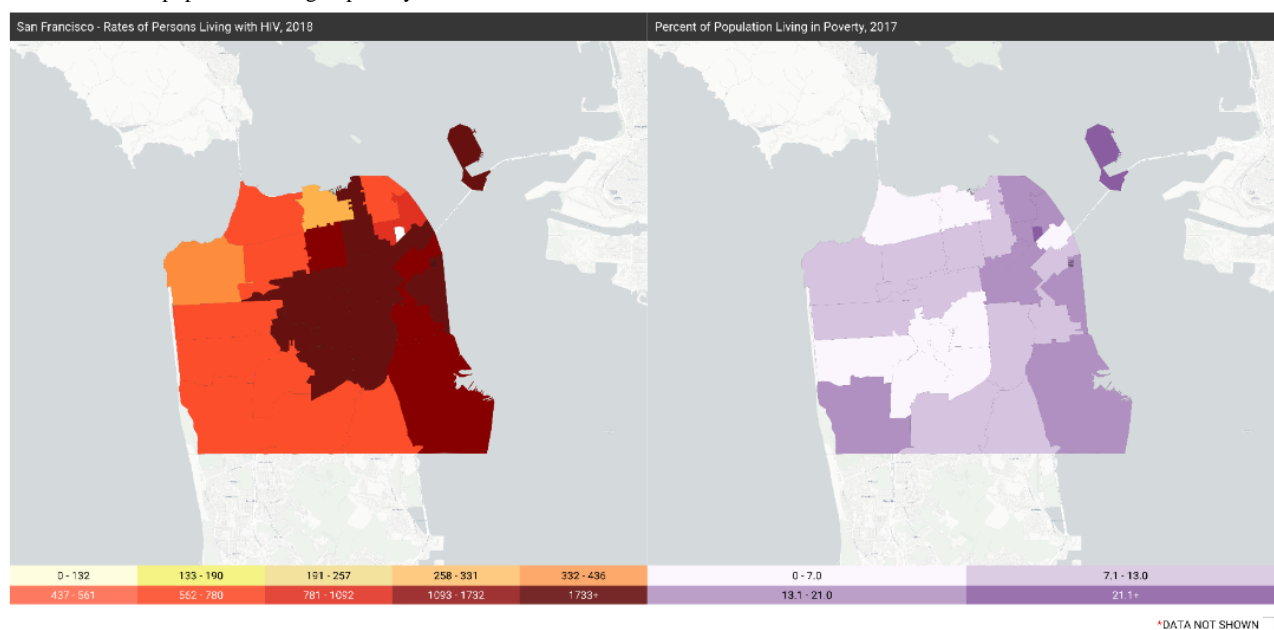
Social determinants of health can be displayed side by side on a secondary map to visualize the relationships between HIV and social determinants. The five social determinants included in AIDSVu are drawn from the US Census and are as follows: (1) poverty (percent of population living in poverty), (2) high school education (percent of population with a high school degree or equivalent), (3) median household income, (4) income inequality (measured by the Gini coefficient, a measure of income inequality where 0 reflects complete equality and 1 reflects complete inequality), and (5) people without health insurance (percent of population lacking health insurance) (see Figure 7).



**Figure 6.** Overlay of pre-exposure prophylaxis (PrEP) services on the Atlanta ZIP Code map of the 5-year risk of new HIV diagnoses, 2014-2018.



**Figure 7.** Percent of population living in poverty, San Francisco.



## Profile Pages

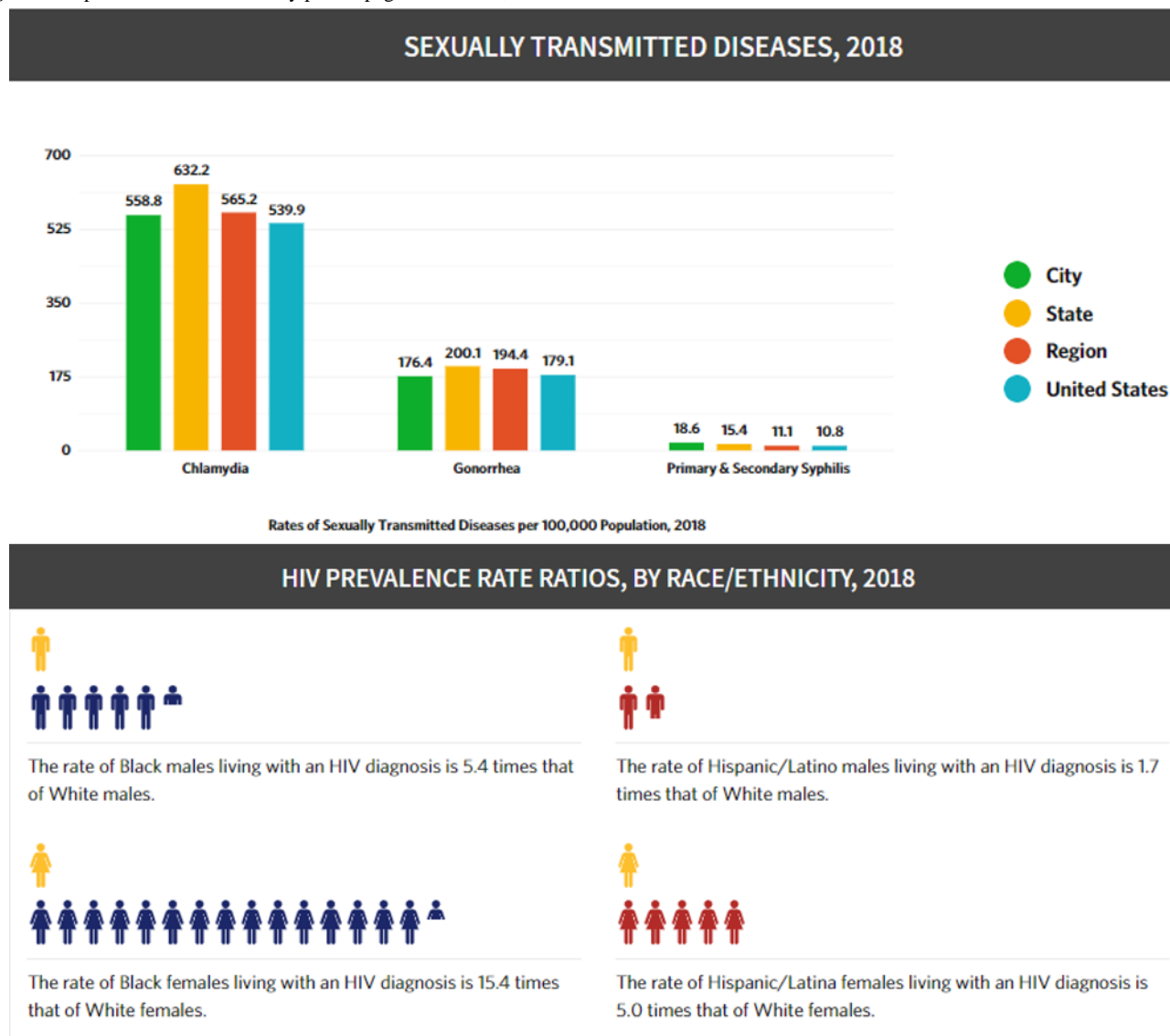
The site has unique pages with profiles for over 40 US cities and 48 counties prioritized for Phase 1 of the *Ending the HIV*

*Epidemic: A Plan for America* initiative [19]; 50 states; Washington, DC; Puerto Rico; four regions; and the nation, offering easy-to-understand, printable snapshots that summarize the impact of HIV in each of these jurisdictions. The pages

enhance the maps and display data and infographics for HIV prevalence, HIV prevalence rate ratios, new HIV diagnoses, new HIV diagnoses by transmission category, HIV mortality, AIDS diagnoses, the total population, and sexually transmitted diseases data. In some cases, AIDSvu serves as a public face

for data that are routinely reported by the CDC for all jurisdictions; for example, in 2018 some cities began reporting diagnosed HIV cases in male-to-female and female-to-male transgender people (see Figure 8).

**Figure 8.** A portion of the Atlanta city profile page in AIDSvu, 2020.



### Downloadable Data and Printable Maps

Data for the overall national level, along with those at regional, state, and county levels, are available for download. Select cities also provide their data for download. The latest national, regional, state, and county data are available for HIV new diagnoses, prevalence, and mortality; social determinants of health; PrEP utilization; and PrEP-to-need ratio. PrEP data are notable because they are a resource that was developed specifically for mapping and dissemination, to meet an urgent public health need [12,20]. Users can freely download these datasets for a number of years (2008-2018) to conduct their own analyses for use in reports, publications, grant applications, etc. Users do not need to submit a data request or obtain permission to use the available data.

Similarly, users are able to download printable maps. These maps can be customized to display a toggled area, titles, and legends, and at different resolutions and transparency levels.

### Results

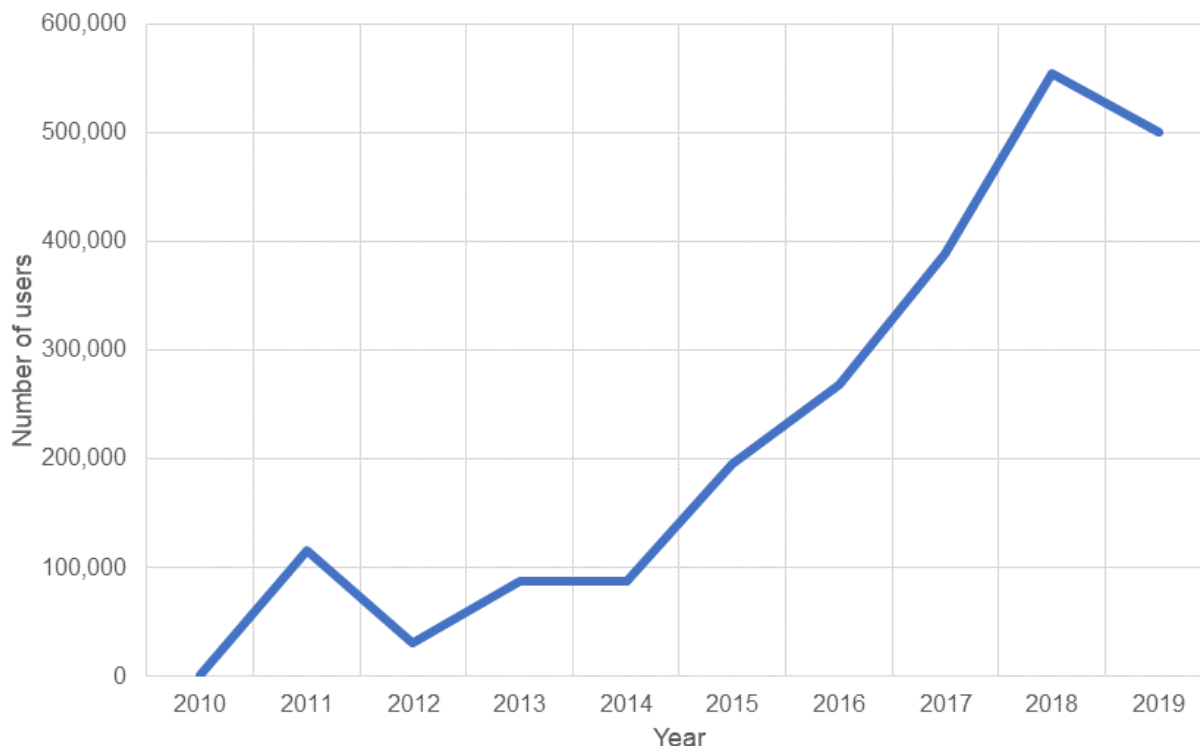
AIDSvu's scientific team collects metrics to evaluate utilization over time by key audiences, including data on number of users, social media reach, and academic publications that reference AIDSvu. All data below are as of the end of 2019 unless otherwise noted.

#### Website Sessions and Users

Since 2010 (ie, the first full year of data reporting), the number of unique website users increased from 1852 in 2010 to 501,527 in 2019 (see Figure 9). Most users (96.8%) originated in the United States, and users from five states—Georgia, Texas,

Florida, California, and New York—accounted for 43.8% of all sessions in 2019; these states represented 36% of the US population in 2019.

**Figure 9.** Total number of AIDSVu users by year, 2010-2019.

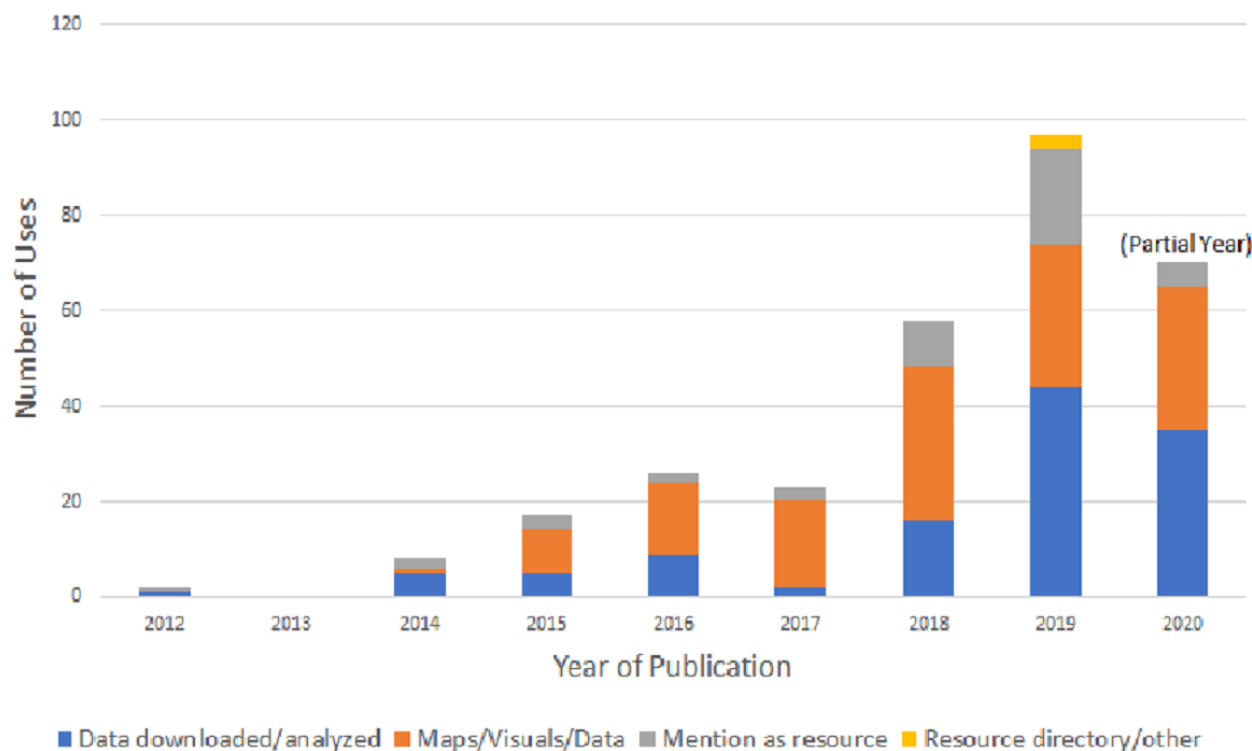


### Social Media Reach

AIDSVu's team also uses various social media platforms to reach key audiences. Key metrics include impressions, defined as the number of times content has been displayed on a screen, and engagements, defined as the number of actions that users take involving posts or tweets. In 2019, AIDSVu generated 3,955,972 impressions on social media—1,928,029 on Facebook and 2,027,943 on Twitter—and 113,444 engagements—48,614 on Facebook and 64,830 on Twitter. Moreover, AIDSVu's Facebook page was followed by 10,440 users, and AIDSVu's Twitter page was followed by 4751 users. Based on organizational types derived from IP addresses, many AIDSVu users are *high-interest* audiences (eg, health department staff, government workers, and academics). For 2019, the distribution of user types included 22,710 academic users, 5274 government users, and 6511 nonprofit users.

### Academic Publications

Given that one of AIDSVu's major objectives is to increase the availability of HIV surveillance data and related information for research purposes, the number of academic papers that cite AIDSVu is a critical metric for evaluation. As of August 22, 2020, 249 peer-reviewed journal articles, 27 academic dissertations or theses, and 6 book chapters had used AIDSVu data or cited AIDSVu's resources. Among uses of AIDSVu in these scholarly products, common uses included use of maps or visual products (136 uses), downloading and using AIDSVu data for analyses (115 uses), citation as a public health tool (46 uses), and use of locator data for research or program purposes (3 uses). Uses of AIDSVu in academic products increased steadily from 2012 through mid-2020 (see Figure 10).

**Figure 10.** Number of academic uses of AIDSvu.org by year and use, 2012-2020.

## Public Health Applications and Use Cases

Examples of how the data on AIDSvu.org are used in public health practice have been recently reviewed [11]. AIDSvu maps have been used to develop targeting for door-to-door testing campaigns, identify areas of high need for telemedicine HIV care facilities [21], and identify service gaps for HIV prevention services [11]. Funders, such as the Elton John AIDS Foundation and the Gilead COMPASS initiative, have used data on the burden of HIV diagnoses to prioritize highly impacted areas for grant funding [22]. Academic researchers have used AIDSvu data to justify the selection of study sites in grant proposals [17].

## Discussion

The US HIV surveillance system is the most comprehensive and well-funded, country-specific surveillance system for any infectious disease in the world, and the quality of the data generated by the system is unparalleled in public health surveillance activities [6]. The data have been used over the course of decades of response to the US HIV epidemic to document health disparities, to argue for resources for care, for treatment and prevention, to serve as the basis for equitable resource allocation, and to identify specific cities and facilities in need of additional resources to provide services.

Although the HIV surveillance system and quality of its data are longstanding, the development and uptake of AIDSvu is a reflection of a growing culture of access to digital data, improved computing capacity and data bandwidth to deliver dynamic content, and an emerging appetite for consuming data through visualization and interaction, in addition to, or in preference to, tabular formats. The platform builds on a

substantial and well-justified public investment in surveillance data and increases the use of those data by presenting data in a way that may be more accessible to some users. This idea—the democratization of data—drives our efforts to make data on the US HIV epidemic widely available and understandable to different types of users.

One of the major enabling factors for the development and growth of AIDSvu has been the collaboration of public health colleagues at all levels of government, as detailed in the Methods section. These advisors promote appropriate uses and interpretations of data, provide advice about the best ways to display data, identify potential weaknesses in data elements or caveats to the interpretation of data, suggest emerging needs that could be met by surveillance data, and prioritize the types of information and formats that would be most useful to the public that they serve. Where possible, AIDSvu attempts to develop resources (eg, local profiles, data overlays, and map views) that reflect questions commonly asked of surveillance programs, so that routine inquiries to these programs can be referred to the website. Surveillance programs for most cities mapped by AIDSvu have agreed to release aggregated datasets for public use, fueling academic analyses and publications by facilitating ready access to data. One of the major lessons learned is that a broad collaboration of data providers and community partners has resulted in broad trust and in having data from many local health departments that are calculated using common definitions and codes and are, therefore, directly comparable to each other.

We acknowledge that surveillance data are available through multiple channels and believe that it is important that these critical data be available in many formats and through different types of partnerships. In 2012, the CDC launched the NCHHSTP



(National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention) AtlasPlus [23], which also offers mapping capabilities at the state and county levels and custom data tabular requests. The NCHHSTP AtlasPlus also provides similar mapping and data functions for related infectious disease surveillance data, including sexually transmitted infections, tuberculosis, and viral hepatitis, which AIDSvu does not include. Conversely, AIDSvu offers some features, like ZIP Code-level mapping, HIV care continuum mapping, infographics and messaging, blog posts to facilitate data interpretation, and service provider overlays, which NCHHSTP AtlasPlus does not offer. Different platforms for presenting and facilitating the understanding of public health surveillance data are needed to serve the broad range of people who can use the data for research, program, policy, and care-planning purposes.

One of the major lessons learned in the development of AIDSvu has been the progression to displaying data at increasingly finer levels of geographic resolution. At the time of AIDSvu's launch of county-level data in 2011, national maps of HIV case counts and prevalence were only available at the state level. The first consolidated depiction of county-level data (see Figure 2) revealed the striking heterogeneity of the impact of HIV within states, and the regional patterns of highly impacted counties in the South. The understanding of the geographic concentration within, and heterogeneity between, counties brought focus to the HIV epidemic in the US South and anticipated the national *Ending the HIV Epidemic* plan's focus on highly impacted urban counties [19]. With the launch of ZIP Code-level data and maps on AIDSvu, geographic disparities within highly impacted cities were apparent. Some cities have agreed to map HIV prevalence data at the census tract level, providing even more precision to inform targeting of prevention and care needs.

AIDSvu has also moved to develop data sources when there was not an existing public data resource that met AIDSvu's strict criteria for inclusion on the site. For example, AIDSvu supported the use of commercial data to develop systematic estimates of PrEP use at the state level in 2019 and accompanied the development of the data with online mapping and public-use datasets for PrEP use at the state level [12]. In 2020, AIDSvu developed and published methods to allow for county-level estimates of PrEP use, in response to the *Ending the HIV Epidemic* initiative's programmatic focus on highly impacted counties [19,20]. This allowed for the assessment of county-level progress toward earlier goals of PrEP uptake developed under the National HIV/AIDS Strategy [24]. Further, the city-level data on AIDSvu are compiled with standardized methods and are publicly available only on AIDSvu.

The AIDSvu platform has several strengths. The strengths include the ability of the platform to provide a common platform to allow cities to display their own surveillance data. Many health departments do not have resources to build and maintain online data visualization platforms, so it makes sense to have a shared resource. Jurisdictions do not report data on ZIP Codes of reported HIV infections to the CDC, so producing city-level resources cannot be done centrally by the federal government. AIDSvu staff can develop and share code to generate local resources such as ZIP Code-level HIV continuum estimates, which is more efficient than having individual health

departments develop such code. This also means that all city-level data presentations are prepared with the same methods so that data are directly comparable across cities. At the time that AIDSvu first displayed county-level data, these data were not publicly available in a single repository, but many states published their own county-level data, sometimes using different methods and assumptions. AIDSvu advanced the field in this area by making county-level estimates public that were directly comparable across state lines. AIDSvu also provides a platform to provide tools that can help interpret, visualize, and translate data to promote public health decision making (eg, map overlays to show areas in need of additional PrEP providers). Finally, AIDSvu has been a robust platform to make population-based data other than surveillance data available to researchers for academic purposes [12,25].

There are also challenges with the AIDSvu platform. The reporting of surveillance data typically lags 12-18 months between the end of a reporting period and the availability of state- and county-level data. Processing for these data includes deduplication of reports across states. In cases where health departments provide data directly to AIDSvu for city-level data displays, the cycle time is faster and data can be made available within 6-9 months, albeit without interstate deduplication and without complete ascertainment of deaths. AIDSvu has been supported for 10 years by a public-private-academic partnership, in which monetary resources are provided by Gilead Sciences and Emory University's Center for AIDS Research; staff effort and data reporting are contributed by public health organizations. Thus, the maintenance of the website, updating of data, and development of new resources are supported with annual commitments, and are not assured in the long term.

Looking forward, AIDSvu will continue to be guided by input on the data needs and priorities of local health programs. We expect to continue moving to display more data at the most granular levels, including expansion to more US cities. Data at the city and ZIP Code level are unique data resources and will be prioritized. Similarly, we will continue to explore ways to make data on PrEP use available with more stratifications where possible, including expanding the stratification of PrEP data to include race and ethnicity. Considering other data platforms like NCHHSTP AtlasPlus, we will continue to reassess the features of AIDSvu that are value added and prioritize the development of new data resources and presentations at finer geographic levels. The AIDSvu platform, which has already been extended to map hepatitis C data [26] and SARS-CoV-2 data [27], may also be used for visualization of other diseases in the future.

Over the past decade, AIDSvu has moved to make HIV surveillance data more accessible and easy to use through innovative visualizations and tabular displays, finer levels of geography, a broad array of indicators (eg, continuum indicators and PrEP use data), and targeted dissemination for multiple stakeholders. Its impact is seen in the utilization of data resources and the growing use of AIDSvu data to support academic research and to inform policy, programmatic, and educational efforts. Surveillance data are intended to inform local responses and are most powerful when they are available for use at the city and county level.



## Acknowledgments

AIDSVu is supported by an unrestricted grant from Gilead Sciences. This work was supported by the Center for AIDS Research at Emory University (P30AI050409). We acknowledge Irene Hall, Anna Satcher Johnson, the US state and city HIV surveillance coordinators, as well as members of our TAG and PTAC: Anna Satcher Johnson, MPH; Becky Grigg, PhD; Dan O'Connell, MA, MLS; Debbie Wendell, MPH, PhD; Eve Mokotoff, MPH; Jacob Dougherty; Jeff Hitt, MEd; Jeff Lauritsen, MPH; Kathleen Brady, MD, PhD; Lorene Maddox, MPH; Margaret Vaaler, PhD; Michael McFadden; Miranda Fanning, MPH; Natalie Cramer, MS; Pascale Wortley, MD, MPH; Pat Young, BS, RN; Richard Aleshire, MSW; Stacy Cohen, MPH; Steve Saunders, MS; and Virginia Hu, MPH.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Data methods for AIDVu.org as of August 2020.

[DOCX File, 22 KB - [jmir\\_v22i10e23173\\_app1.docx](#)]

## References

1. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN, Guidelines Working Group Centers for Disease Control and Prevention (CDC). Updated guidelines for evaluating public health surveillance systems: Recommendations from the Guidelines Working Group. *MMWR Recomm Rep* 2001 Jul 27;50(RR-13):1-35; quiz CE1. [Medline: [18634202](#)]
2. Sullivan PS, Delpech V. HIV surveillance: The conscience of the epidemic. *Open AIDS J* 2012;6:65-66 [FREE Full text] [doi: [10.2174/1874613601206010065](#)] [Medline: [23049654](#)]
3. Thacker SB, Birkhead GS. Surveillance. In: Gregg M, editor. *Field Epidemiology*. 2nd edition. New York, NY: Oxford University Press; 2002:50-86.
4. Sullivan PS, McKenna MT, Waller LA, Williamson GD, Lee LM. Analyzing and interpreting public health surveillance data. In: Lee LM, Teutsch SM, Thacker SB, St. Louis ME, editors. *Principles and Practice of Public Health Surveillance*. 3rd edition. New York, NY: Oxford University Press; 2010:88-145.
5. Colorado HIV Surveillance Quarterly Report: STI/HIV/Viral Hepatitis Surveillance Program. Denver, CO: Colorado Department of Public Health & Environment; 2020 Apr. URL: <https://drive.google.com/file/d/1rc52eD3FW81qJtB5hOX1b5yV6edQ4jEb/view> [accessed 2020-06-30]
6. Nakashima AK, Fleming PL. HIV/AIDS surveillance in the United States, 1981-2001. *J Acquir Immune Defic Syndr* 2003 Feb;32 Suppl 1:S68-S85. [Medline: [12571518](#)]
7. California Department of Public Health. California HIV Surveillance Report — 2018. Sacramento, CA: California Department of Public Health, Center for Infectious Diseases, Office of AIDS; 2020 Feb 27. URL: [https://www.cdph.ca.gov/Programs/CID/DOA/CDPH%20Document%20Library/California\\_HIV\\_Surveillance\\_Report2018.pdf](https://www.cdph.ca.gov/Programs/CID/DOA/CDPH%20Document%20Library/California_HIV_Surveillance_Report2018.pdf) [accessed 2020-09-16]
8. Infectious Disease Assessment Unit, Washington State Department of Health. Washington State HIV Surveillance Report, 2019 Edition. Olympia, WA: Washington State Department of Health; 2019 Nov. URL: <https://www.doh.wa.gov/Portals/1/Documents/Pubs/150-030-WAHIVSurveillanceReport2019.pdf> [accessed 2020-06-30]
9. HIV Epidemiology Program. HIV Surveillance Annual Report, 2018. New York, NY: New York City Department of Health and Mental Hygiene; 2019 Nov. URL: <https://www1.nyc.gov/assets/doh/downloads/pdf/dires/hiv-surveillance-annualreport-2018.pdf> [accessed 2020-06-30]
10. Washington, DC: District of Columbia Department of Health; HIV/AIDS, Hepatitis, STD, & TB Administration; 2019. URL: [https://dchealth.dc.gov/sites/default/files/dc/sites/doh/publication/attachments/2019%20Annual%20Surveillance%20Report\\_09062019.pdf](https://dchealth.dc.gov/sites/default/files/dc/sites/doh/publication/attachments/2019%20Annual%20Surveillance%20Report_09062019.pdf) [accessed 2020-06-30]
11. Valdiserri RO, Sullivan PS. Data visualization promotes sound public health practice: The AIDSVu example. *AIDS Educ Prev* 2018 Feb;30(1):26-34. [doi: [10.1521/aeap.2018.30.1.26](#)] [Medline: [29481299](#)]
12. Sullivan PS, Giler RM, Mouhanna F, Pembleton ES, Guest JL, Jones J, et al. Trends in the use of oral emtricitabine/tenofovir disoproxil fumarate for pre-exposure prophylaxis against HIV infection, United States, 2012-2017. *Ann Epidemiol* 2018 Dec;28(12):833-840 [FREE Full text] [doi: [10.1016/j.annepidem.2018.06.009](#)] [Medline: [30037634](#)]
13. Siegler AJ, Mouhanna F, Giler RM, Weiss K, Pembleton E, Guest J, et al. The prevalence of pre-exposure prophylaxis use and the pre-exposure prophylaxis-to-need ratio in the fourth quarter of 2017, United States. *Ann Epidemiol* 2018 Dec;28(12):841-849 [FREE Full text] [doi: [10.1016/j.annepidem.2018.06.005](#)] [Medline: [29983236](#)]
14. National Academies of Sciences, Engineering, and Medicine. Integrating Responses at the Intersection of Opioid Use Disorder and Infectious Disease Epidemics: Proceedings of a Workshop. Washington, DC: The National Academies Press; 2018.

15. Siegler AJ, Bratcher A, Weiss KM, Mouhanna F, Ahlschlager L, Sullivan PS. Location location location: An exploration of disparities in access to publicly listed pre-exposure prophylaxis clinics in the United States. *Ann Epidemiol* 2018 Dec;28(12):858-864 [FREE Full text] [doi: [10.1016/j.annepidem.2018.05.006](https://doi.org/10.1016/j.annepidem.2018.05.006)] [Medline: [30406756](https://pubmed.ncbi.nlm.nih.gov/30406756/)]
16. Ojikutu BO, Bogart LM, Mayer KH, Stopka TJ, Sullivan PS, Ransome Y. Spatial access and willingness to use pre-exposure prophylaxis among Black/African American individuals in the United States: Cross-sectional survey. *JMIR Public Health Surveill* 2019 Feb 04;5(1):e12405 [FREE Full text] [doi: [10.2196/12405](https://doi.org/10.2196/12405)] [Medline: [30714945](https://pubmed.ncbi.nlm.nih.gov/30714945/)]
17. Bauermeister J, Sullivan PS, Gravens L, Wolfe J, Countryman K, Smith-Bankhead N, et al. Reducing HIV vulnerability through a multilevel life skills intervention for adolescent men (The iREACH Project): Protocol for a randomized controlled trial. *JMIR Res Protoc* 2018 Jul 10;7(7):e10174 [FREE Full text] [doi: [10.2196/10174](https://doi.org/10.2196/10174)] [Medline: [29991470](https://pubmed.ncbi.nlm.nih.gov/29991470/)]
18. Siegler AJ, Wirtz S, Weber S, Sullivan PS. Developing a web-based geolocated directory of HIV pre-exposure prophylaxis-providing clinics: The PrEP locator protocol and operating procedures. *JMIR Public Health Surveill* 2017 Sep 06;3(3):e58 [FREE Full text] [doi: [10.2196/publichealth.7902](https://doi.org/10.2196/publichealth.7902)] [Medline: [28877865](https://pubmed.ncbi.nlm.nih.gov/28877865/)]
19. Fauci AS, Redfield RR, Sigounas G, Weahkee MD, Giroir BP. Ending the HIV epidemic: A plan for the United States. *JAMA* 2019 Mar 05;321(9):844-845. [doi: [10.1001/jama.2019.1343](https://doi.org/10.1001/jama.2019.1343)] [Medline: [30730529](https://pubmed.ncbi.nlm.nih.gov/30730529/)]
20. Sullivan PS, Mouhanna F, Mera R, Pembleton E, Castel AD, Jaggi C, et al. Methods for county-level estimation of pre-exposure prophylaxis coverage and application to the US: Ending the HIV epidemic jurisdictions. *Ann Epidemiol* 2020 Apr;44:16-30 [FREE Full text] [doi: [10.1016/j.annepidem.2020.01.004](https://doi.org/10.1016/j.annepidem.2020.01.004)] [Medline: [32088073](https://pubmed.ncbi.nlm.nih.gov/32088073/)]
21. AIDSvu recognized by White House as public-private partnership making a difference. AIDSvu. 2013. URL: <https://aidsvu.org/aidsvu-recognized-by-white-house-as-public-private-partnership-making-a-difference/> [accessed 2020-09-21]
22. AIDSvu. 2018. URL: <https://aidsvu.org/ejaf/> [accessed 2020-07-01]
23. NCHHSTP AtlasPlus. Centers for Disease Control and Prevention. 2012. URL: <https://www.cdc.gov/nchhstp/atlas/index.htm> [accessed 2020-09-15]
24. Office of National AIDS Policy. National HIV/AIDS Strategy. Improving Outcomes: Accelerating Progress Along The HIV Care Continuum. Washington, DC: The White House Office of National AIDS Policy; 2013 Dec. URL: [https://obamawhitehouse.archives.gov/sites/default/files/nap\\_nhas\\_improving\\_outcomes\\_dec\\_2013.pdf](https://obamawhitehouse.archives.gov/sites/default/files/nap_nhas_improving_outcomes_dec_2013.pdf) [accessed 2020-09-21]
25. Grey JA, Bernstein KT, Sullivan PS, Purcell DW, Chesson HW, Gift TL, et al. Estimating the population sizes of men who have sex with men in US states and counties using data from the American Community Survey. *JMIR Public Health Surveill* 2016;2(1):e14 [FREE Full text] [doi: [10.2196/publichealth.5365](https://doi.org/10.2196/publichealth.5365)] [Medline: [27227149](https://pubmed.ncbi.nlm.nih.gov/27227149/)]
26. Rosenberg ES, Rosenthal EM, Hall EW, Barker L, Hofmeister MG, Sullivan PS, et al. Prevalence of hepatitis C virus infection in US states and the District of Columbia, 2013 to 2016. *JAMA Netw Open* 2018 Dec 21;1(8):e186371 [FREE Full text] [doi: [10.1001/jamanetworkopen.2018.6371](https://doi.org/10.1001/jamanetworkopen.2018.6371)] [Medline: [30646319](https://pubmed.ncbi.nlm.nih.gov/30646319/)]
27. Siegler AJ, Sullivan PS, Sanchez T, Lopman B, Fahimi M, Sailey C, et al. Protocol for a national probability survey using home specimen collection methods to assess prevalence and incidence of SARS-CoV-2 infection and antibody response. *Ann Epidemiol* 2020 Aug 10;49:50-60 [FREE Full text] [doi: [10.1016/j.annepidem.2020.07.015](https://doi.org/10.1016/j.annepidem.2020.07.015)] [Medline: [32791199](https://pubmed.ncbi.nlm.nih.gov/32791199/)]

## Abbreviations

**CDC:** Centers for Disease Control and Prevention

**NCHHSTP:** National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention

**NIH:** National Institutes of Health

**PrEP:** pre-exposure prophylaxis

**PTAC:** Prevention and Treatment Advisory Committee

**TAG:** Technical Advisory Group

*Edited by G Eysenbach; submitted 03.08.20; peer-reviewed by A Greenberg, C Ciesielski, A Hamp, A Mavragani; comments to author 20.08.20; revised version received 25.08.20; accepted 13.09.20; published 23.10.20.*

### *Please cite as:*

Sullivan PS, Woodyatt C, Koski C, Pembleton E, McGuinness P, Taussig J, Ricca A, Luisi N, Mokotoff E, Benbow N, Castel AD, Do AN, Valdiserri RO, Bradley H, Jaggi C, O'Farrell D, Filipowicz R, Siegler AJ, Curran J, Sanchez TH  
*A Data Visualization and Dissemination Resource to Support HIV Prevention and Care at the Local Level: Analysis and Uses of the AIDSvu Public Data Resource*  
*J Med Internet Res* 2020;22(10):e23173  
 URL: <https://www.jmir.org/2020/10/e23173>  
 doi:[10.2196/23173](https://doi.org/10.2196/23173)  
 PMID:[33095177](https://pubmed.ncbi.nlm.nih.gov/33095177/)

©Patrick Sean Sullivan, Cory Woodyatt, Chelsea Koski, Elizabeth Pembleton, Pema McGuinness, Jennifer Taussig, Alexandra Ricca, Nicole Luisi, Eve Mokotoff, Nanette Benbow, Amanda D Castel, Ann N Do, Ronald O Valdiserri, Heather Bradley, Chandni Jaggi, Daniel O'Farrell, Rebecca Filipowicz, Aaron J Siegler, James Curran, Travis H Sanchez. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Intersection of the Web-Based Vaping Narrative With COVID-19: Topic Modeling Study

Kamila Janmohamed<sup>1</sup>; Abdul-Nasah Soale<sup>2</sup>, MSc; Laura Forastiere<sup>1,3</sup>, PhD; Weiming Tang<sup>4,5,6</sup>, PhD; Yongjie Sha<sup>4</sup>, BA; Jakob Demant<sup>7</sup>, PhD; Edoardo Airolidi<sup>2</sup>, PhD; Navin Kumar<sup>1</sup>, MPhil

<sup>1</sup>Human Nature Lab, Department of Sociology, Yale University, New Haven, CT, United States

<sup>2</sup>Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA, United States

<sup>3</sup>Department of Biostatistics, Yale School of Public Health, New Haven, CT, United States

<sup>4</sup>UNC Project-China, University of North Carolina at Chapel Hill, Guangzhou, China

<sup>5</sup>Southern Medical University Dermatology Hospital, Guangzhou, China

<sup>6</sup>School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

<sup>7</sup>University of Copenhagen, Copenhagen, Denmark

**Corresponding Author:**

Navin Kumar, MPhil

Human Nature Lab

Department of Sociology

Yale University

17 Hillhouse Ave

New Haven, CT, 06511

United States

Phone: 1 4754148375

Email: [navin.kumar@yale.edu](mailto:navin.kumar@yale.edu)

## Abstract

**Background:** The COVID-19 outbreak was designated a global pandemic on March 11, 2020. The relationship between vaping and contracting COVID-19 is unclear, and information on the internet is conflicting. There is some scientific evidence that vaping cannabidiol (CBD), an active ingredient in cannabis that is obtained from the hemp plant, or other substances is associated with more severe manifestations of COVID-19. However, there is also inaccurate information that vaping can aid COVID-19 treatment, as well as expert opinion that CBD, possibly administered through vaping, can mitigate COVID-19 symptoms. Thus, it is necessary to study the spread of inaccurate information to better understand how to promote scientific knowledge and curb inaccurate information, which is critical to the health of vapers. Inaccurate information about vaping and COVID-19 may affect COVID-19 treatment outcomes.

**Objective:** Using structural topic modeling, we aimed to map temporal trends in the web-based vaping narrative (a large data set comprising web-based vaping chatter from several sources) to indicate how the narrative changed from before to during the COVID-19 pandemic.

**Methods:** We obtained data using a textual query that scanned a data pool of approximately 200,000 different domains (4,027,172 documents and 361,100,284 words) such as public internet forums, blogs, and social media, from August 1, 2019, to April 21, 2020. We then used structural topic modeling to understand changes in word prevalence and semantic structures within topics around vaping before and after December 31, 2019, when COVID-19 was reported to the World Health Organization.

**Results:** Broadly, the web-based vaping narrative can be organized into the following groups or archetypes: harms from vaping; Vaping Regulation; Vaping as Harm Reduction or Treatment; and Vaping Lifestyle. Three archetypes were observed prior to the emergence of COVID-19; however, four archetypes were identified post-COVID-19 (Vaping as Harm Reduction or Treatment was the additional archetype). A topic related to CBD product preference emerged after COVID-19 was first reported, which may be related to the use of CBD by vapers as a COVID-19 treatment.

**Conclusions:** Our main finding is the emergence of a vape-administered CBD treatment narrative around COVID-19 when comparing the web-based vaping narratives before and during the COVID-19 pandemic. These results are key to understanding how vapers respond to inaccurate information about COVID-19, optimizing treatment of vapers who contract COVID-19, and possibly minimizing instances of inaccurate information. The findings have implications for the management of COVID-19

among vapers and the monitoring of web-based content pertinent to tobacco to develop targeted interventions to manage COVID-19 among vapers.

(*J Med Internet Res* 2020;22(10):e21743) doi:[10.2196/21743](https://doi.org/10.2196/21743)

## KEYWORDS

vaping; COVID-19; topic modeling; web-based narrative; modeling; trend; narrative; social media; internet; web-based health information

## Introduction

COVID-19 is spreading rapidly and has been declared a global pandemic [1]. COVID-19 was first reported in Wuhan, China, in December 2019 [2], and it was declared a pandemic by the World Health Organization (WHO) on March 11, 2020 [3]. With the pandemic currently in progress, research on the determinants of disease progression and communities that may be more vulnerable to COVID-19 is of key importance [4].

There is some scientific knowledge that vaping cannabidiol (CBD) or other substances may be associated with more severe manifestations of COVID-19 [5,6]. Use of electronic cigarettes (e-cigarettes) has been associated with a reduction in the ability of the lungs to respond to infection [5,6]; thus, people who use e-cigarettes may be at increased risk of contracting COVID-19 [4,7]. Several studies have indicated that smokers, including vapers, are more vulnerable to COVID-19 infections or more likely to develop serious complications after contracting SARS-CoV-2, the virus that causes COVID-19 [8-10]. Vaping devices may also be possible sites of COVID-19 transmission [9]. However, inaccurate information that vaping can aid COVID-19 treatment is also circulating [9], and experts have expressed the opinion that CBD, which can be administered through vaping, can mitigate COVID-19 symptoms [11,12].

For example, some Twitter posts around vaping indicated that e-cigarette devices may increase lung humidity and prevent COVID-19, and other posts stated that these devices can be used to administer COVID-19 medication to the lungs and possibly destroy the virus [9]. It has also been suggested that CBD products, often delivered through vaping, can be used to treat COVID-19, perhaps by augmenting the immune system [11,12]. Much of the information around vaping and other tobacco products is disseminated through the internet [13,14] and can affect health outcomes [15,16]. For example, vapers who develop COVID-19 may mistakenly believe that vaping CBD or other substances can alleviate COVID-19 symptoms; however, it may instead create additional disease complications. Thus, it is necessary to study the spread of inaccurate information to better understand how to promote scientific knowledge and curb misinformation, which may be critical to vapers' health [17].

A previous study analyzed Twitter content around COVID-19 and vaping [9]. The indicated study surveilled vaping tweets and detailed conversations around COVID-19 and vaping. The web-based conversations centered on the possibly heightened risk of COVID-19 for vapers and how vaping could potentially protect against COVID-19 [9]. However, past work did not use web-based vaping-related data from a range of sources or

detailed data from before the COVID-19 pandemic and through its progression. Detailing a large scope of sources is necessary to document the broad range of web-based vaping conversations, and collecting data from both before and during the COVID-19 pandemic is key to understanding the narrative prior to the emergence of COVID-19 and how it is changing as the pandemic progresses. Our study builds on previous work by using a large data set to represent the web-based vaping narrative (August 1, 2019, to April 21, 2020) that combines analysis of a multitude of sources, such as blogs, forums, and social media posts; also, we used novel computational techniques to examine how the vaping narrative has changed from before to during the COVID-19 pandemic.

Using the novel computational technique of structural topic modeling (STM), we mapped temporal trends in the web-based vaping narrative (a large data set comprising web-based vaping chatter from several sources) to show how discourse differed before versus during the COVID-19 pandemic. Topic modeling is a computer-aided content analysis technique where texts are organized into themes known as topics. These topics are not provided to the machine prior to modeling but emerge inductively as the algorithm learns patterns within the texts. The model assumes a relational theory of meaning by identifying structures of co-occurrence of words in individual texts and across all the texts. The model thus provides content analysis of text data sets that are too large to code by hand. Topic models use machine learning to uncover patterns and relationships that may be omitted by hand coding or traditional content analysis. Unsupervised machine learning methods have performed similarly to human coders on identical documents [18]. Unsupervised machine learning is a variant of machine learning that looks for new patterns in a data set without pre-existing labels and with limited human supervision. We used an approach to topic modeling known as STM [19,20]. STM enables discovery of topics and their prevalence based on document metadata, such as dates, or other important attributes, such as the number of new COVID-19 cases worldwide per day. Adding this metadata is useful, as the data are obtained over several months (August 2019 to April 2020), and the web-based vaping narrative may be susceptible to thematic change based on the progression of the COVID-19 pandemic. STM has been used to address several social scientific research questions around areas such as climate change [21,22] and web-based drug marketplaces [23]. As vapers may be at greater risk for contracting SARS-CoV-2 and COVID-19 disease progression [8-10], and CBD, which is often administered through vaping, may have interactions with COVID-19 treatment outcomes, we hope to provide insight on how vapers are responding to the pandemic. This may help improve the treatment outcomes of vapers who develop COVID-19.



## Methods

### Ethics Statement

Approval and informed consent were not needed because the data were collected using publicly available textual query techniques. All data are publicly available and can be accessed by anyone. The data were provided to the research team with all identifiers removed.

### Data Acquisition and Processing

Data were obtained using a textual query that scanned a data pool of approximately 200,000 different domains, such as public forum posts, blogs, news articles, message boards, health care provider forums, and social media (see [Multimedia Appendix 1](#) for the full list of sources). Textual queries were used to automatically search the indicated sources for text fragments related to keywords such as *vape*, *vaping*, and *e-cigarette*. The data that comprised vaping-related text fragments were collected from August 1, 2019, to April 21, 2020. As the data set represents a multitude of sources for web-based chatter related to vaping, our data set is likely representative of the web-based vaping narrative during the indicated period. The start date for the COVID-19 pandemic was denoted as December 31, 2019, when the Chinese government disclosed the existence of COVID-19 to the World Health Organization (WHO) [3]. Although the date of the first COVID-19 case is prior to December 31, 2019 [2], COVID-19 is unlikely to have influenced vaping-related discourse in the United States prior to December 31, 2019, due to low global awareness. The time period of August 1, 2019, to April 21, 2020, was chosen to provide sufficient data to detail the vaping narrative prior to the COVID-19 pandemic. Given that the date demarcating pre- and post-COVID-19 is December 31, 2019, our time period allowed for approximately four months before and approximately the same period after the first report of COVID-19 to the WHO.

### Word Prevalence and Topic Modeling

To prepare the data for word prevalence and topic modeling analysis, English stop words such as “the,” “a,” and “an” were removed, along with abbreviations, and terms were stemmed using Porter’s stemming algorithm [24]. Stemming converts words with the same “stem” or root (eg, “innovative” and “innovator”) to a single word type (eg, “innovate”). As our study centered on the intersection of vaping and COVID-19, it was expected that words such as “cigarette,” “vape,” and “coronavirus” would dominate our findings. However, these terms may crowd out other words, perhaps causing us to miss key topics occurring in the text. For example, if we were interested in understanding different cooking techniques such as roasting and frying, and we sourced data from web-based forums frequented by amateur chefs, the most frequent words in the data might be “cook” and “recipe.” However, these words might obscure information around the cooking techniques we were interested in. Thus, in some cases, such as our study, it may be necessary to remove frequently occurring words to detail underlying themes in the data. All data were first processed to remove mentions of COVID-19, tobacco, and vaping. These data were used to generate word clouds by word prevalence. As we will later detail, the word clouds generated by word

prevalence contained significant mentions of CBD after the emergence of COVID-19. When conducting topic modeling, mentions of CBD may crowd out other words and reduce our ability to identify salient topics. As such, we further processed the data set for topic modeling by removing mentions of cannabis, inclusive of CBD.

We first generated word clouds based on the top 200 terms ranked by prevalence before and after COVID-19 was reported to the WHO. In a word cloud, a larger font size indicates a greater prevalence of a single word. Word clouds thus provide a relative yardstick of the importance of a word in a particular time period. This visualization enabled us to qualitatively assess words by importance. Documents were processed (words removed) for mentions of cannabis, COVID-19, tobacco, and vaping. References to cannabis were determined using these search terms: [*bud OR cannabis OR cannabidiol OR cbd OR ganja OR hash OR hashish OR hemp OR indica OR joint OR marijuana OR mary jane OR ruderalis OR pot OR sativa OR weed OR THC*]. References to COVID-19 were determined using these search terms: [*COVID-19 OR covid 19 OR novel coronavirus OR coronavirus OR sars cov-2 OR sars cov 2 OR sars-cov-2 OR n-cov OR cov OR covid*]. References to tobacco were determined using these search terms: [*baccy OR bidi OR cig OR cigar OR cigarillo OR cigarette OR ciggy OR fag OR hookah OR pipe OR shag OR sheesha OR shisha OR snuff OR snus OR tobacco*]. References to vaping were determined using these search terms: [*e-cig OR electronic cigarette OR vape OR vaper OR vaping OR vapelife OR vapist OR vapiin OR vapylyfe*].

We then used topic modeling to understand changes in word prevalence within topics around vaping and COVID-19. Topic modeling is a computer-aided content analysis technique in which texts are organized into themes known as topics [25,26]. In topic modeling, a topic is a distribution over a vocabulary of words that represent semantically interpretable themes [19]. For example, in a topic denoted “vape,” the terms “smoke” and “device” are more likely to occur than the words “peanut” and “tomato.” “Smoke” may appear in both “vape” and “cooking” topics with different contextual meanings. Given that the topic is a distribution, “smoke” may appear with other high-probability terms such as “roast” and “fry” in the “cooking” topic but may appear with terms such as “nicotine” and “device” in the “vape” topic. Thus, topics can be understood by considering that a person who was talking about the topic of “cooking” would tend to use some words more frequently than others compared to if they were talking about the topic of “vape.” Topic models are suitable for analyzing large quantities of textual data via an automated technique for providing context.

We used an approach to topic modeling known as STM [18,20]. STM [18,20] enables the generation of topics regarding document metadata such as date and source as well as other covariates relevant to the research question, such as new COVID-19 cases. This is vital to understanding how the narrative and topic proportions change over time. This enabled a robust quantitative analysis of how the COVID-19 pandemic has shaped the web-based narrative on vaping [19]. The key innovation of STM is that it can incorporate metadata or information about each document. This allows metadata covariates, such as new COVID-19 cases per day, to influence

topic discovery. Metadata can affect both the prevalence and content of a topic. Metadata covariates for topical prevalence allow the metadata to affect topic frequency. Similarly, covariates in topical content allow the metadata to affect the word rate within a topic or how a topic is discussed [20]. The STM process will output documents and vocabulary for analysis [20]. This output can be investigated in a range of ways, such as detailing words associated with topics or the relationships between metadata and topics. Model output can be used to conduct hypothesis testing around these relationships. STM [18,20] was applied to the whole data set (August 1, 2019, to April 21, 2020); the data prior to the reporting of COVID-19 to the WHO only (August 1 to December 31, 2019); and the data after COVID-19 was reported to the WHO only (January 1 to April 21, 2020). We subsetting the data to see if the vaping narratives were different before and after COVID-19 was reported to the WHO. We used the following metadata covariates for the STM models. For the full data set, the covariates were the binary variable for before and after COVID-19 was reported to the WHO, COVID-19 content covariate (variable to control for COVID-19-related content), date (the first day was denoted as 1, and the days were numbered sequentially after), source (0=social media, 1=news), new COVID-19 cases per day worldwide, and new COVID-19 deaths per day worldwide. For the pre-COVID-19 data set, the covariates were the date and source. For the post-COVID-19 data set, the covariates were the date, source, new COVID-19 cases per day worldwide, and new COVID-19 deaths per day worldwide.

Because STM is an unsupervised approach, the number of topics to estimate ( $k$ ) is key to the analysis. We first estimated several models ranging from 5 to 200 topics. These models were then evaluated qualitatively by their ability to produce coherent topics and capture topics regarding vaping and COVID-19 [27]. The number of topics was based on our understanding of the data set and how other researchers interpreted STM results [27,28]. The choice of the number of topics was also influenced by postestimation validation outcomes and past work [27]. As per standard content analysis [29], topic model validation also requires qualitative review, where researchers assess the interpretability and relative efficacy of models based on their subject matter expertise and data context. Our final models ( $k=15$  for the whole data set;  $k=20$  for the pre-COVID-19 report data set;  $k=20$  for the post-COVID-19 report data set) provided the greatest external validity and the most semantically coherent output of distinctive topics. When the number of topics was greater than indicated above, there were diminishing returns for solutions, as the substantive meaning and coherence of categories started to break down [21]. When the number of topics was lower, variation decreased and specific topics were placed into more generic categories. Validating a topic model is not the same as evaluating a statistical model regarding a population sample [30]. The goal is to identify the framework that best describes the data, not to estimate population parameters [30].

We conducted qualitative analysis to determine the number of topics based on past social science studies in which topic modeling was used to extract meaning from large text samples.

These studies [21,31] determined topics by qualitative coding based on word prevalence and researchers' topic expertise. We applied similar techniques in this study. Methods such as interrater reliability ratings may guard against subjective bias based on subject matter expertise and data context [32]. Adding interrater reliability to the qualitative component of topic modeling may improve data quality. However, we sought to use topic modeling to answer a specific research question, not to improve on methodological techniques. Thus, we used best practices implemented in previous studies regarding topic interpretation but did not advance these methods.

Topic interpretation was influenced by prior knowledge about why texts were written and what they sought to accomplish. Most of the text was produced and consumed by people who engaged in vaping and other forms of tobacco use, and this lens was used to interpret the presence or absence of topics and words. Most of the topic labels were straightforward and did not require much interpretation. To characterize topics in the vaping narrative, we qualitatively coded each topic by investigating word clouds based on each topic and reviewing exemplar documents which detailed high proportions of each topic [19]. The topic we classified as "tobacco company merger called off" had the following most frequently occurring words: "sue," "analyst," "halt," "1st," "imperial," "judge," "backlash," "advisor," "merger," "stake," "acquire," "outbreak," "carolina," "confirm," and "mint." Exemplar documents that exhibited high proportions of this topic indicated a preoccupation with these words. This detailed a topical preoccupation with a tobacco firm merger being called off. Thus, the interpretation of the topic was clear, given the genre of the narrative and the reliance on research regarding prominent topics around vaping. Two authors assigned the topics, and a third author resolved disagreements when they arose.

Topic validation is key to assessing whether the substantive meaning of the topic and its words are parallel with the qualitative meaning of the text, and we used methodological guidance from past research for this purpose [19,26]. Past work advocated the use of sample documents to validate the substantive meaning of each topic. Determining the number of sample documents to use is based on the amount of resolution needed by a social scientist to answer the research question using topic modeling methods [33]. Thus, determining the number of sample documents is a largely qualitative process that is dependent on the research question at hand. To determine the appropriate number of documents to sample, we searched the social science literature for studies that used topic modeling based on whether they asked similar research questions to our study, addressed similar topic areas to our study, and drew their study data from similar sources to those used in our study. We searched databases such as Web of Science Core Collection, Embase, PsycINFO, MEDLINE, and Sociological Abstracts. We used keywords such as *contrarian*, *polarized*, and *topic modeling*. The 2016 paper by Farrell [21] was determined to be the most similar to our study based on the assessed characteristics. Farrell explored ideological polarization in climate change and used a broad range of sources, such as press releases, published papers, and website articles. Based on the nature of the research question and large range of sources,

Farrell determined that a sample of 50 documents was sufficient to validate the substantive meaning of the topic output. Given the similarities between Farrell's 2016 study and ours in a range of characteristics, we similarly determined that a sample of 50 documents was adequate to validate the topics. We used `findThoughts` and `plotQuote` within the STM package to examine the top 50 associated documents for each topic to validate a topic's substantive meaning. Determination of the top 50 documents was based on ranking topics by the maximum a posteriori estimate of the topic's theta value, which represents the modal estimate of the proportion of word tokens assigned to the topic with the model. These top 50 documents were read by two of the authors to determine validity. A third author resolved disagreements where necessary. As indicated above, interrater reliability was not determined; however, we believe our methods were sufficiently robust.

Finally, word clouds were again generated, this time to visualize topics with the top 100 words ranked by STM-generated weights per topic, for the models representing the full data set and the data sets from before and after COVID-19 was reported to the WHO. In these word clouds, a larger font size indicated a greater weight, with word clouds indicating the importance of a word within a topic. We then grouped topics and their associated word clouds into larger categories or archetypes based on shared concepts across topics [34]. All analysis was conducted using R with the following packages: `dplyr`, `quanteda`, `textclean`, `tm`, and `stm` [20,24,35-37].

## Results

### Data

We collected 4,027,172 documents (361,100,284 words) comprised of text from blogs (86.01%, N=310,546,244), news articles (11.02%, N=39,721,031), forums (3.01%, N=10,833,008), comments (<1%), professional reviews and Facebook posts (both <1%).

### Word Prevalence

The most frequently observed words in the data set were "ban" (216,735/361,100,284 words, 0.06%, rank 1), "product" (135,607/361,100,284 words, 0.04%, rank 2), and "make" (115,413/361,100,284 words, 0.03%, rank 3).

Figure 1 shows a word cloud displaying the 200 most frequently featured words in the data sets divided by time period. The words are colored, sized, and positioned radially based on frequency of appearance, with larger, more central words appearing most frequently. In the period before COVID-19 was reported to the WHO (prior to December 31, 2019), commonly featured words included "ban," "lung," and "quit." Over the next four months, mentions of "CBD" and "oil" increased, along with positive words such as "good" and "best." There was a clear shift in word prevalence before COVID-19 was reported to the WHO compared to after the COVID-19 report. Word prevalence shifted from words related to the vaping ban to positive words and words associated with CBD. The vaping ban was a move by the US government on September 11, 2019, to remove all flavored vaping products from the market [38].



**Month 2 (February 1 2020 – February 28 2020)**

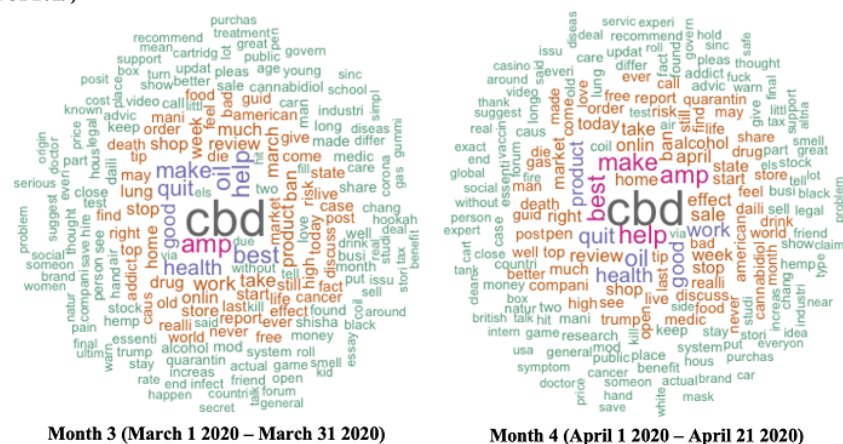


Figure 2 (all observations), Figure 3 (before COVID-19 was reported to the WHO), and Figure 4 (after COVID-19 was reported to the WHO) detailed results of the topic modeling analysis. Topics not directly relevant to our analysis were not indicated in the figures, such as the Australian bushfires and

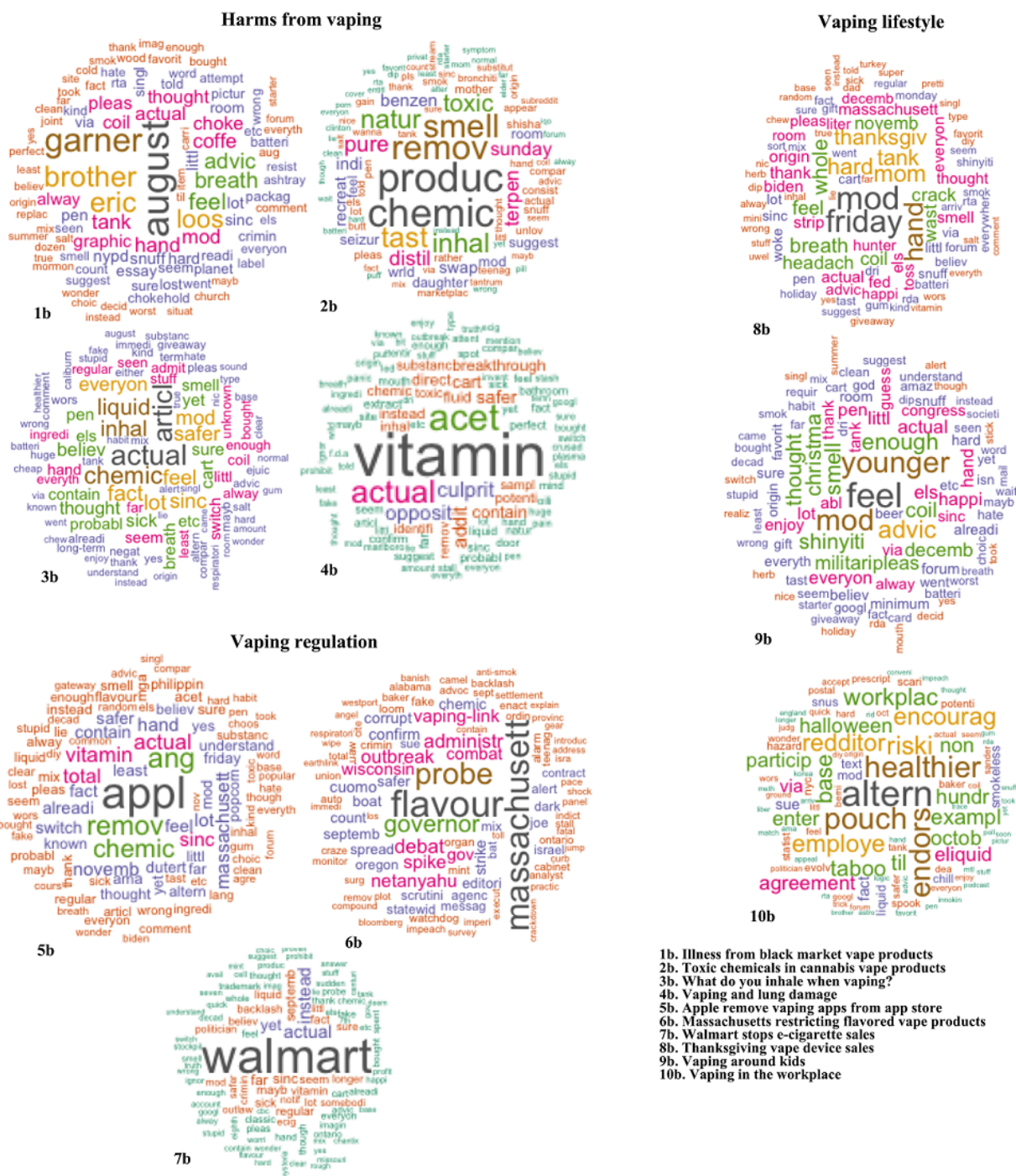
J Med Internet Res 2020 | vol. 22 | iss. 10 | e21743 | p.991  
(page number not for citation purposes)

**Figure 2.** Major archetypes of vaping-related topics with example topics per archetype obtained by structural topic modeling for all observations (August 1, 2019, to April 21, 2020). The word clouds are generated from the weights of the top 100 terms within a topic. Terms with larger weights are depicted in larger font sizes. Terms with approximately the same weight are depicted in the same color.





**Figure 3.** Major archetypes of vaping-related topics with example topics per archetype obtained by structural topic modeling before COVID-19 was reported to the WHO (August 1 to December 31, 2019). The word clouds are generated from the weights of the top 100 terms within a topic. Terms with larger weights are depicted in larger font sizes. Terms with approximately the same weight are depicted in the same color.



**Figure 4.** Major archetypes of vaping-related topics with example topics per archetype obtained by structural topic modeling after COVID-19 was reported to the WHO (January 1 to April 21, 2020). The word clouds are generated from the weights of the top 100 terms within a topic. Terms with larger weights are depicted in larger font sizes. Terms with approximately the same weight are depicted in the same color.



Topic modeling captured significant events in the vaping environment, such as the vaping health crisis and Walmart stopping e-cigarette sales. The vaping health crisis referred to the 2019 emergence of vaping-associated pulmonary injury (VAPI) in the United States [39]. Walmart stopping e-cigarette sales denoted the September 2019 termination of vape sales at Walmart after the advent of VAPI [40]. These were likely significant events in the vaping narrative, as they emphasized the possible harms of vaping and were highly prominent in US media. Accordingly, these events were identified as individual topics. We organized the topics into the following groups or archetypes: Harms from Vaping; Vaping Regulation; Vaping as Harm Reduction or Treatment; and Vaping Lifestyle. We

generated word clouds from the weights of the top 100 terms within each topic, and Figures 2–4 show sample word clouds for each of the three archetypes. Broadly, across the whole time period (Figure 2), the web-based vaping narrative was centered around harms from vaping and vaping regulation. Archetypes pre-COVID-19 (Figure 3) versus during COVID-19 (Figure 4) were largely similar, except that three archetypes were identified pre-COVID-19 (Harms From Vaping; Vaping Regulation; Vaping Lifestyle) but four archetypes were identified post-COVID-19 (Harms From Vaping; Vaping Regulation; Vaping as Harm Reduction or Treatment; Vaping Lifestyle). This suggests that the emergence of COVID-19 is related to the appearance of topics around vaping as a form of harm reduction



or treatment. There was also variation in the topics within an archetype. For example, pre-COVID-19, the Harms From Vaping archetype included topics such as “illness from black market vape products” and “toxic chemicals in cannabis vape products.” After COVID-19 was reported to the WHO, example topics in the same archetype were “vaping and dental health” and “vaping and lung damage.” This suggests that while underlying themes in the vaping narrative are largely stagnant, specific topics in an archetype may vary over time. In line with the difference in archetypes before and during the COVID-19 pandemic, there were also different event-based topics before and during COVID-19. Before COVID-19, several topics represented events significant to vapers in that time period, such as “Walmart stops e-cigarette sales” and “Thanksgiving vape device sales.” These event-related topics were not present after COVID-19 was reported to the WHO. Instead, we noticed new event-based topics, such as “Rush Limbaugh supports vaping” and “disposable vape devices not covered by ban.” After COVID-19 was reported to the WHO, we noticed the emergence of topics specific to COVID-19 that were not present pre-COVID-19. These topics were “vape stores as essential service” (vape stores not being designated as essential services when COVID-19 lockdowns occurred in the United States) and “sharing vapes in COVID-19” (vape devices being possible sites of SARS-CoV-2 transmission). Similarly, a “CBD product preference” topic emerged after the COVID-19 report (Figure 4, word cloud 8c) under the Vaping as Harm Reduction or Treatment archetype that may be related to the advent of the pandemic. As indicated in the Methods section, we removed words around cannabis to provide for more nuanced analysis; accordingly, CBD and other related terms did not appear in the topic-based word clouds. This CBD topic was not present pre-COVID-19. The appearance of the “CBD product preference” topic may be related to vapers using CBD as a treatment for COVID-19. As a validity check, we examined the top 50 associated documents for the “CBD product preference” topic to validate the substantive meaning of the topic. Convenience sampling was not used to sample the top 50 topics; instead, we used the theta values of the topics, as detailed earlier. The number of top-ranked documents to be sampled was based on the methodology outlined earlier. There were 114,622 documents in total for the “CBD product preference” topic. These documents were read by two of the authors to determine validity. A third author resolved disagreements where necessary. As indicated above, interrater reliability was not used. We found that a majority of these top 50 documents (31, 62%) detailed CBD administered through vaping as a possible COVID-19 cure or protective agent. Example text fragments regarding how vaping CBD can prevent or treat COVID-19:

*Pot smoke is the best expectorant I've ever used and fresh cbd oil or weed brownies are verrrrrrrry healing, promotes good sleep and good healing. If you are a non-smoker, tobacco smoke should help you clear out your lungs if you have nothing else.* [March 10, 2020]

*COVID-19 deaths invariably involve a "cytokine storm," an excessive, un-checked immune system response. Cannabinoids from cannabis, cbd in*

*particular, can lower cytokine production naturally. research needed asap!* [March 25, 2020]

Example text fragments regarding websites marketing CBD for COVID-19 prevention or treatment:

*Why hemp cbd flowers and a vaporizer are the best COVID-19 coronavirus prepping tools.* [February 18, 2020]

*REDACTED COMPANY NAME applauds the use of cbd during the coronavirus outbreak* [March 30, 2020]

*Toronto-based cannabis seller testing cbd's effectiveness on reducing symptoms of coronavirus.* [April 1, 2020]

## Discussion

### Principal Findings

Our main finding was the emergence of discourse around vape-administered CBD treatment for COVID-19 when comparing web-based vaping narratives before and after the outbreak of COVID-19. Recent work suggested that CBD use may increase COVID-19 risks [41]. Other studies indicated that CBD may aid COVID-19 treatment outcomes [12,42]. Vaping CBD products as a treatment for COVID-19 is still largely unsubstantiated. Beliefs around CBD as a COVID-19 treatment, bolstered by marketing campaigns and early-stage research [11,12,42], may be responsible for the emergence of discussion around CBD.

There is limited empirical research on the intersection of the vaping narrative and COVID-19, especially around the emergence of CBD-related COVID-19 treatments and comparing the narratives before and during the COVID-19 pandemic. The role of web-based narratives in tobacco control, especially within social media, is a growing field of study [43]. We expand on past work that used novel computational approaches to examine trends in digital media to understand how these web-based behaviors may influence health behaviors [44,45]; our results indicate that the web-based environment is key to comprehending vaping and related health outcomes, especially in response to public health events. Previous research suggested the need to monitor social media content around tobacco to protect youth and mitigate tobacco use [43]. We expand on these studies, bolstering the need to surveil web-based tobacco content given our findings around increased discussion of inaccurate COVID-19 vape-administered treatments that are not evidence-based and may worsen health outcomes. A recent review detailed the role of misinformation in public health outcomes [17], and we expand on past work by providing evidence of how large-scale events may create misinformation in the health sphere. The strength of this study is our use of innovative computational methods to explore the content of the vaping narrative and how it is affected by COVID-19, comparing narrative content before the COVID-19 outbreak and after the pandemic took shape. This outcome measurement is key to understanding how vapers respond to COVID-19, enabling optimized treatment of vapers who develop COVID-19, and possibly minimizing instances of inaccurate information.

Our findings around the use of CBD as a non-evidence-based and possibly injurious COVID-19 treatment, likely administered through vaping, are indicative of the earlier discussed point. Inaccurate information on the internet may create complications in COVID-19 treatment. Vapers who develop COVID-19 may use vape-administered CBD treatments; meanwhile, CBD is associated with reduced immune system functioning [41] and may heighten the disease progression of COVID-19. It is possible that upon contracting COVID-19, people may use vape devices to administer medication to themselves. Given the possibility of device malfunction [46,47], some individuals may further harm their health if they develop COVID-19 symptoms. As levels of misinformation around devices as a means to administer purported COVID-19 treatments increases, more people may share these modified devices; this creates possible sites of transmission [9,48] and may further increase SARS-CoV-2 transmission rates.

If health care professionals are aware that vapers with COVID-19 may use CBD as a treatment based on inaccurate information, these professionals may be better able to respond to vapers with COVID-19 who demonstrate CBD-related complications. Professionals can provide accurate information regarding COVID-19 to vapers who seek health care. Misinformation can also be combated with trusted information. Public health authorities can include COVID-19-specific information in targeted vaping-related messaging, perhaps mitigating consumption of inaccurate web-based information. Information-based campaigns can target inaccurate information in line with the topics identified by our results, such as vaping CBD as a COVID-19 treatment. There are several experimental interventions around reducing levels of inaccurate information [49,50], with some centering specifically on COVID-19 [51] and other health issues [17]. Interventions that harness similar techniques, such as asking respondents to judge information accuracy, may nudge individuals toward obtaining accurate information around COVID-19 and vaping. Interventions can also center on vaping-related health literacy in various media outlets, which may reduce misinformation about the topic [52]. Thus, our results may improve COVID-19 treatment for individuals who may have received inaccurate information around COVID-19 and vaping; they may also provide insight on reducing levels of misinformation among vapers during the pandemic.

Our findings have several implications. From a policy standpoint, we suggest that vaping forums be mandated to provide accurate data around the interactions between vaping and COVID-19. These efforts may reduce levels of inaccurate information around COVID-19 and vaping and may minimize any COVID-19 complications associated with vaping. Future research can identify changes in the vaping narrative as the pandemic progresses further, allowing public health authorities to adjust treatment provision for vapers at risk of contracting COVID-19. Future work can also address how inaccurate information on the internet can be mitigated, especially as the pandemic progresses.

### Limitations

Our results depended on the validity of the data collected through our textual query. We searched a wide range of web-based media and identified key themes that validated our results (eg, Walmart stops e-cigarette sales, Vitamin E acetate and vaping illness), and we are thus confident in the comprehensiveness of our data. We may have overlooked some slang terms for vaping and thus underestimated the web-based narrative. We did not obtain location data for individual text fragments; thus, we were not able to determine how COVID-19 cases in certain areas affected the narrative. Our data were drawn from August 1, 2019, to April 21, 2020, and we were not able to determine changes in the narrative after April. We were not able to collect all web-based vaping chatter and may have missed some themes. In future research, we will collect qualitative and survey data from vapers to enhance the current findings. We did not use interrater reliability for our qualitative analysis, and we will use such methods in future research.

### Conclusions

We demonstrated the advent of discourse around vape-administered CBD treatment for COVID-19 by comparing the web-based vaping narratives before and during the COVID-19 pandemic. The increase in CBD-related discussion within the vaping narrative may be due to the marketing of CBD products consumed through vaping as a COVID-19 treatment [11]. Our findings have implications for the management of COVID-19 among vapers and for monitoring of web-based content pertinent to tobacco.

### Acknowledgments

This study was funded by The Foundation for a Smoke-Free World. The funding body had no role in the design, analysis, or interpretation of the data in the study. This study was preregistered on the Open Science Framework (DOI 10.17605/OSF.IO/XBVF2).

### Authors' Contributions

KJ, NK, AN, and EA contributed to the study design, hypothesis generation, data collection, data analysis, data interpretation, and manuscript writing and review. LF, WT, YS, DM, and JD contributed to the manuscript writing and review.

### Conflicts of Interest

None declared.

## Multimedia Appendix 1

List of sources from which data were obtained.

[XLS File (Microsoft Excel File), 1228 KB - [jmir\\_v22i10e21743\\_app1.xls](#)]

## References

1. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet* 2020 Apr;395(10231):1225-1228. [doi: [10.1016/s0140-6736\(20\)30627-9](#)]
2. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020 Apr 07;323(13):1239-1242. [doi: [10.1001/jama.2020.2648](#)] [Medline: [32091533](#)]
3. Timeline of WHO's response to COVID-19. World Health Organization. URL: <https://www.who.int/news-room/detail/29-06-2020-covidtimeline> [accessed 2020-05-29]
4. Volkow ND. Collision of the COVID-19 and Addiction Epidemics. *Ann Intern Med* 2020 Jul 07;173(1):61-62. [doi: [10.7326/m20-1212](#)]
5. Madison MC, Landers CT, Gu BH, Chang CY, Tung HY, You R, et al. Electronic cigarettes disrupt lung lipid homeostasis and innate immunity independent of nicotine. *J Clin Invest* 2019 Oct 01;129(10):4290-4304 [FREE Full text] [doi: [10.1172/JCI128531](#)] [Medline: [31483291](#)]
6. Gilpin DF, McGown K, Gallagher K, Bengoechea J, Dumigan A, Einarsson G, et al. Electronic cigarette vapour increases virulence and inflammatory potential of respiratory pathogens. *Respir Res* 2019 Dec 18;20(1):267 [FREE Full text] [doi: [10.1186/s12931-019-1206-8](#)] [Medline: [31847850](#)]
7. McAlinden KD, Eapen MS, Lu W, Chia C, Haug G, Sohal SS. COVID-19 and vaping: risk for increased susceptibility to SARS-CoV-2 infection? *Eur Respir J* 2020 Jul 19;56(1):2001645. [doi: [10.1183/13993003.01645-2020](#)] [Medline: [32430427](#)]
8. Glantz SA. Reduce your risk of serious lung disease caused by corona virus by quitting smoking and vaping. Center for Tobacco Control Research and Education. 2020. URL: <https://tobacco.ucsf.edu/reduce-your-risk-serious-lung-disease-caused-corona-virus-quitting-smoking-and-vaping> [accessed 2020-04-08]
9. Majmundar A, Allem J, Cruz TB, Unger JB. Public Health Concerns and Unsubstantiated Claims at the Intersection of Vaping and COVID-19. *Nicotine Tob Res* 2020 Aug 24;22(9):1667-1668 [FREE Full text] [doi: [10.1093/ntr/ntaa064](#)] [Medline: [32285129](#)]
10. Lewis T. Smoking or Vaping May Increase the Risk of Severe Coronavirus Infection. *Scientific American*. 2020 Mar 17. URL: <https://www.scientificamerican.com/article/smoking-or-vaping-may-increase-the-risk-of-a-severe-coronavirus-infection1/> [accessed 2020-10-02]
11. Fraudulent Coronavirus Disease 2019 (COVID-19) Products. US Food and Drug Administration. URL: <https://www.fda.gov/consumers/health-fraud-scams/fraudulent-coronavirus-disease-2019-covid-19-products> [accessed 2020-05-29]
12. Esposito G, Pesce M, Seguela L, Sanseverino W, Lu J, Corpetti C, et al. The potential of cannabidiol in the COVID-19 pandemic. *Br J Pharmacol* 2020 Jun 10;2020 [FREE Full text] [doi: [10.1111/bph.15157](#)] [Medline: [32519753](#)]
13. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *Am J Prev Med* 2011 Apr;40(4):448-453. [doi: [10.1016/j.amepre.2010.12.007](#)] [Medline: [21406279](#)]
14. Yamin CK, Bitton A, Bates DW. E-cigarettes: a rapidly growing Internet phenomenon. *Ann Intern Med* 2010 Nov 02;153(9):607-609. [doi: [10.7326/0003-4819-153-9-201011020-00011](#)] [Medline: [21041581](#)]
15. Freeman B. New media and tobacco control. *Tob Control* 2012 Mar 16;21(2):139-144. [doi: [10.1136/tobaccocontrol-2011-050193](#)] [Medline: [22345236](#)]
16. Forsyth SR, Malone RE. "I'll be your cigarette--light me up and get on with it": examining smoking imagery on YouTube. *Nicotine Tob Res* 2010 Aug 15;12(8):810-816 [FREE Full text] [doi: [10.1093/ntr/ntq101](#)] [Medline: [20634267](#)]
17. Swire-Thompson B, Lazer D. Public Health and Online Misinformation: Challenges and Recommendations. *Annu Rev Public Health* 2020 Apr 02;41(1):433-451. [doi: [10.1146/annurev-publhealth-040119-094127](#)] [Medline: [31874069](#)]
18. Roberts ME, Stewart BM, Airoidi EM. A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association* 2016 Oct 18;111(515):988-1003. [doi: [10.1080/01621459.2016.1141684](#)]
19. Roberts ME, Stewart BM, Tingley D, Lucas C, Leder-Luis J, Gadarian SK, et al. Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science* 2014 Mar 06;58(4):1064-1082. [doi: [10.1111/ajps.12103](#)]
20. Roberts ME, Stewart BM, Tingley D. stm: An R Package for Structural Topic Models. *J Stat Soft* 2019;91(2):1-40. [doi: [10.18637/jss.v091.i02](#)]
21. Farrell J. Corporate funding and ideological polarization about climate change. *Proc Natl Acad Sci USA* 2016 Jan 05;113(1):92-97. [doi: [10.1073/pnas.1509433112](#)] [Medline: [26598653](#)]
22. Tvinnereim E, Fløttum K. Explaining topic prevalence in answers to open-ended survey questions about climate change. *Nature Clim Change* 2015 Jun 1;5(8):744-747. [doi: [10.1038/nclimate2663](#)]
23. Guarino S, Santoro M. Multi-word structural topic modelling of ToR drug marketplaces. In: 2018 IEEE 12th International Conference on Semantic Computing (ICSC). 2018 Presented at: IEEE 12th International Conference on Semantic Computing (ICSC); Jan 31-Feb 2, 2018; Laguna Hills, CA p. 269-273. [doi: [10.1109/icsc.2018.00048](#)]



24. Feinerer I, Hornik K, Meyer D. Text Mining Infrastructure in R. *J Stat Soft* 2008;25(5):54. [doi: [10.18637/jss.v025.i05](https://doi.org/10.18637/jss.v025.i05)]
25. Mohr JW, Bogdanov P. Introduction—Topic models: What they are and why they matter. *Poetics* 2013 Dec;41(6):545-569. [doi: [10.1016/j.poetic.2013.10.001](https://doi.org/10.1016/j.poetic.2013.10.001)]
26. Blei DM. Probabilistic topic models. *Commun ACM* 2012 Apr;55(4):77-84. [doi: [10.1145/2133806.2133826](https://doi.org/10.1145/2133806.2133826)]
27. Grimmer J, Stewart BM. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Polit Anal* 2017 Jan 04;21(3):267-297. [doi: [10.1093/pan/mps028](https://doi.org/10.1093/pan/mps028)]
28. Wallach HM, Murray I, Salakhutdinov S, Mimno D. Evaluation methods for topic models. In: *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*. 2009 Jun Presented at: 26th Annual International Conference on Machine Learning; Jun 14-18, 2009; Montréal, QC p. 1105-1112. [doi: [10.1145/1553374.1553515](https://doi.org/10.1145/1553374.1553515)]
29. Krippendorff K. *Content Analysis: An Introduction to its Methodology*. 4th Edition. Los Angeles, CA: Sage Publications; 2018.
30. DiMaggio P, Nag M, Blei D. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics* 2013 Dec;41(6):570-606. [doi: [10.1016/j.poetic.2013.08.004](https://doi.org/10.1016/j.poetic.2013.08.004)]
31. Munksgaard R, Demant J. Mixing politics and crime - The prevalence and decline of political discourse on the cryptomarket. *Int J Drug Policy* 2016 Sep;35:77-83. [doi: [10.1016/j.drugpo.2016.04.021](https://doi.org/10.1016/j.drugpo.2016.04.021)] [Medline: [27424175](https://pubmed.ncbi.nlm.nih.gov/27424175/)]
32. Saal FE, Downey RG, Lahey MA. Rating the ratings: Assessing the psychometric quality of rating data. *Psychol Bull* 1980;88(2):413-428. [doi: [10.1037/0033-2909.88.2.413](https://doi.org/10.1037/0033-2909.88.2.413)]
33. Nikolenko SI, Koltcov S, Koltsova O. Topic modelling for qualitative studies. *J Inf Sci* 2016 Jul 10;43(1):88-102. [doi: [10.1177/0165551515617393](https://doi.org/10.1177/0165551515617393)]
34. Evans MS. A computational approach to qualitative analysis in large textual datasets. *PLoS One* 2014 Feb 3;9(2):e87908 [FREE Full text] [doi: [10.1371/journal.pone.0087908](https://doi.org/10.1371/journal.pone.0087908)] [Medline: [24498398](https://pubmed.ncbi.nlm.nih.gov/24498398/)]
35. Wickham H, François R, Henry L, Müller K. *dplyr: A Grammar of Data Manipulation*. URL: <https://dplyr.tidyverse.org/reference/dplyr-package.html> [accessed 2020-10-02]
36. Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, et al. quanteda: An R package for the quantitative analysis of textual data. *JOSS* 2018 Oct;3(30):774. [doi: [10.21105/joss.00774](https://doi.org/10.21105/joss.00774)]
37. Rinker TW. *textclean: Text Cleaning Tools*. GitHub. URL: <https://github.com/trinker/textclean> [accessed 2020-10-02]
38. Furlow B. US Government contemplates a nationwide ban on flavoured e-cigarette products. *Lancet Respir Med* 2019 Nov;7(11):932. [doi: [10.1016/s2213-2600\(19\)30329-7](https://doi.org/10.1016/s2213-2600(19)30329-7)]
39. Hooper RW, Garfield JL. An Emerging Crisis: Vaping-Associated Pulmonary Injury. *Ann Intern Med* 2019 Oct 08;172(1):57. [doi: [10.7326/m19-2908](https://doi.org/10.7326/m19-2908)]
40. Lawrence K. *e-Cigarette Industry in Crisis: How Can Juul Survive?* SAGE Publications: SAGE Business Cases Originals 2020 Jan 15:online. [doi: [10.4135/9781529728613](https://doi.org/10.4135/9781529728613)]
41. Brown JD. Cannabidiol as prophylaxis for SARS-CoV-2 and COVID-19? Unfounded claims versus potential risks of medications during the pandemic. *Res Social Adm Pharm* 2020 Mar 31:online ahead of print [FREE Full text] [doi: [10.1016/j.sapharm.2020.03.020](https://doi.org/10.1016/j.sapharm.2020.03.020)] [Medline: [32273254](https://pubmed.ncbi.nlm.nih.gov/32273254/)]
42. Wang B, Kovalchuk A, Li D, Ilynskyy Y, Kovalchuk I, Kovalchuk O. In Search of Preventative Strategies: Novel Anti-Inflammatory High-CBD Cannabis Sativa Extracts Modulate ACE2 Expression in COVID-19 Gateway Tissues. Preprints. Preprint posted online on April 19, 2020 .
43. Kostygina G, Tran H, Shi Y, Kim Y, Emery S. 'Sweeter Than a Swisher': amount and themes of little cigar and cigarillo content on Twitter. *Tob Control* 2016 Oct 03;25(Suppl 1):i75-i82. [doi: [10.1136/tobaccocontrol-2016-053094](https://doi.org/10.1136/tobaccocontrol-2016-053094)] [Medline: [27697951](https://pubmed.ncbi.nlm.nih.gov/27697951/)]
44. Huang J, Kornfield R, Emery SL. 100 Million Views of Electronic Cigarette YouTube Videos and Counting: Quantification, Content Evaluation, and Engagement Levels of Videos. *J Med Internet Res* 2016 Mar 18;18(3):e67 [FREE Full text] [doi: [10.2196/jmir.4265](https://doi.org/10.2196/jmir.4265)] [Medline: [26993213](https://pubmed.ncbi.nlm.nih.gov/26993213/)]
45. Huang J, Kornfield R, Szczycka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. *Tob Control* 2014 Jul;23 Suppl 3:iii26-iii30. [doi: [10.1136/tobaccocontrol-2014-051551](https://doi.org/10.1136/tobaccocontrol-2014-051551)] [Medline: [24935894](https://pubmed.ncbi.nlm.nih.gov/24935894/)]
46. Chadi N, Belanger RE. Teen vaping: There is no vapour without fire. *Paediatr Child Health* 2020 Oct;25(6):337-339. [doi: [10.1093/pch/pxz137](https://doi.org/10.1093/pch/pxz137)] [Medline: [32968465](https://pubmed.ncbi.nlm.nih.gov/32968465/)]
47. Reasoner JJ, Regier BA, Beckendorf R, McAllister RK. Update on the Risks of Electronic Cigarettes-Vaping. *Ochsner J* 2020 Apr 03;20(1):2-4 [FREE Full text] [doi: [10.31486/toj.20.0012](https://doi.org/10.31486/toj.20.0012)] [Medline: [32284674](https://pubmed.ncbi.nlm.nih.gov/32284674/)]
48. Winickoff J, Hartman L, Reynolds M. Smoking and E-cigarette Use Put Individuals at Increased Risk of Coronavirus (COVID-19). Massachusetts Health Officers Association. URL: <https://mhoa.com/wp-content/uploads/2020/07/Smoking-and-E-cigarette-Use-Put-Individuals-at-Increased-Risk-of-COVID-19-1.pdf> [accessed 2020-08-15]
49. Pennycook G, Rand DG. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc Natl Acad Sci USA* 2019 Feb 12;116(7):2521-2526. [doi: [10.1073/pnas.1806781116](https://doi.org/10.1073/pnas.1806781116)] [Medline: [30692252](https://pubmed.ncbi.nlm.nih.gov/30692252/)]
50. Barnett PA, Hoskins CE, Alhoti JA, Carpenter LJ. Reducing Public Misinformation About Organ Donation: An Experimental Intervention. *Journal of Social Distress and the Homeless* 2013 Jul 19;18(1-2):57-73. [doi: [10.1179/105307809805365235](https://doi.org/10.1179/105307809805365235)]

51. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychol Sci* 2020 Jul;31(7):770-780. [doi: [10.1177/0956797620939054](https://doi.org/10.1177/0956797620939054)] [Medline: [32603243](https://pubmed.ncbi.nlm.nih.gov/32603243/)]
52. Nordheim LV, Gundersen MW, Espehaug B, Guttersrud Ø, Flottorp S. Effects of School-Based Educational Interventions for Enhancing Adolescents Abilities in Critical Appraisal of Health Claims: A Systematic Review. *PLoS One* 2016 Aug 24;11(8):e0161485 [FREE Full text] [doi: [10.1371/journal.pone.0161485](https://doi.org/10.1371/journal.pone.0161485)] [Medline: [27557129](https://pubmed.ncbi.nlm.nih.gov/27557129/)]

## Abbreviations

**CBD:** cannabidiol

**e-cigarette:** electronic cigarette

**STM:** structural topic modeling

**VAPI:** vaping-assisted pulmonary injury

**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 29.06.20; peer-reviewed by C Harshbarger, M El Tantawi; comments to author 15.07.20; revised version received 16.07.20; accepted 16.09.20; published 30.10.20.*

*Please cite as:*

Janmohamed K, Soale AN, Forastiere L, Tang W, Sha Y, Demant J, Airoidi E, Kumar N  
*Intersection of the Web-Based Vaping Narrative With COVID-19: Topic Modeling Study*  
*J Med Internet Res* 2020;22(10):e21743

URL: <http://www.jmir.org/2020/10/e21743/>

doi: [10.2196/21743](https://doi.org/10.2196/21743)

PMID: [33001829](https://pubmed.ncbi.nlm.nih.gov/33001829/)

©Kamila Janmohamed, Abdul-Nasah Soale, Laura Forastiere, Weiming Tang, Yongjie Sha, Jakob Demant, Edoardo Airoidi, Navin Kumar. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Building a Digital Tool for the Adoption of the World Health Organization's Antenatal Care Recommendations: Methodological Intersection of Evidence, Clinical Logic, and Digital Technology

Samira M Haddad<sup>1,2</sup>, MSc, MD, PhD; Renato T Souza<sup>1,2</sup>, MSc, MD, PhD; Jose Guilherme Cecatti<sup>1,2</sup>, MSc, MD, PhD; Maria Barreix<sup>3</sup>, MHS; Tigest Tamrat<sup>3</sup>, MPH; Carolyn Footitt<sup>4</sup>, MSPH; Garrett L Mehl<sup>3</sup>, PhD; Inraini F Syah<sup>5</sup>, MPH; Anuraj H Shankar<sup>5,6,7</sup>, PhD; Özge Tunçalp<sup>3</sup>, MD, PhD

<sup>1</sup>Department of Obstetrics and Gynecology, School of Medical Sciences, University of Campinas, Campinas, Brazil

<sup>2</sup>Center for Research in Reproductive Health of Campinas (CEMICAMP), Campinas, Brazil

<sup>3</sup>UNDP–UNFPA–UNICEF–WHO–World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Reproductive Health and Research, World Health Organization, Geneva, Switzerland

<sup>4</sup>Ona Systems Inc, Nairobi, Kenya

<sup>5</sup>Summit Institute of Development, Mataram, Indonesia

<sup>6</sup>Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

<sup>7</sup>Eijkman-Oxford Clinical Research Unit, Eijkman Institute for Molecular Biology, Jakarta, Indonesia

**Corresponding Author:**

Jose Guilherme Cecatti, MSc, MD, PhD  
Department of Obstetrics and Gynecology  
School of Medical Sciences  
University of Campinas  
101 Alexander Fleming Street  
University of Campinas  
Campinas, 13083-891  
Brazil  
Phone: 55 1935219482  
Email: [cecatti@unicamp.br](mailto:cecatti@unicamp.br)

**Related Article:**

This is a corrected version. See correction statement: <https://www.jmir.org/2020/10/e24891/>

## Abstract

**Background:** One of the key mandates of the World Health Organization (WHO) is to develop guidelines, defined as “a document containing recommendations for clinical practice or public health policy.” Guidelines represent the global standard for information sources shaping clinical practice and public health policies. Despite the rigorous development process and the value of guidelines for setting standards, implementing such standards within local contexts and at the point of care is a well-documented challenge. Digital technologies enable agile information management and may facilitate the adaptation of guidelines to diverse settings of health services delivery.

**Objective:** The objective of this paper is to detail the systematic and iterative process involved in transforming the WHO Antenatal Care (ANC) guidelines into a digital decision-support and patient-record application for routine use in primary health care settings, known as the WHO digital ANC module.

**Methods:** The WHO convened a team of clinical and digital health experts to develop the WHO digital ANC module as a tool to assist health care professionals in the implementation of WHO evidence-based recommendations for pregnant women. The WHO digital ANC module's creation included the following steps: defining a minimum viable product (MVP), developing clinical workflows and algorithms, algorithm testing, developing a data dictionary, and the creation of a user interface or application development. The overall process of development took approximately 1 year to reach a stable prototype and to finalize the underlying content requirements of the data dictionary and decision support algorithms.

**Results:** The first output is a reference software reflecting the generic WHO ANC guideline content, known as the WHO digital ANC module. Within it, all actionable ANC recommendations have related data fields and algorithms to confirm whether the associated task was performed. WHO recommendations that are not carried out by the health care worker are saved as pending tasks on a woman's health record, and those that are adequately fulfilled trigger messages with positive reinforcement. The second output consists of the structured documentation of the different components which contributed to the development of the WHO digital ANC module, such as the data dictionary and clinical decision support workflows.

**Conclusions:** This is a novel approach to facilitate the adoption and adaptation of recommendations through digital systems at the health service delivery level. It is expected that the WHO digital ANC module will support the implementation of evidence-based practices and provide information for monitoring and surveillance; however, further evidence is needed to understand how the WHO digital ANC module impacts the implementation of WHO recommendations. Further, the module's implementation will inform the WHO's ongoing efforts to create a pathway to adaptive and integrated (Smart) Guidelines in Digital Systems to improve health system quality, coverage, and accountability.

(*J Med Internet Res* 2020;22(10):e16355) doi:[10.2196/16355](https://doi.org/10.2196/16355)

## KEYWORDS

clinical decision support; antenatal care; digital health; requirements gathering; implementation; WHO guidelines; evidence-based

## Introduction

One of the key mandates of the World Health Organization (WHO) is to develop guidelines, defined as “a document containing recommendations for clinical practice or public health policy” [1]. These recommendations aim to assist policymakers and clinicians in making an informed decision on whether to undertake specific interventions based on an assessment of the anticipated impact, values and preferences, and use of resources [1]. For WHO and other institutions, such as the Institute of Medicine (IOM) and the National Institute for Health and Care Excellence (NICE), the development of guidelines entails a structured process and an underlying evidence base to support the formulation of recommendations [1,2]. As such, guidelines often represent a global standard and credible information sources for shaping clinical practice and public health policies.

Despite the rigorous development process and the value of guidelines for setting standards, applying global recommendations within local contexts and at the point of care is a well-documented challenge [2-9]. These challenges arise from both systemic issues (such as a lack of supplies, poor staff training, excessive workload, high staff turnover, lack of quality data collection, and difficulty monitoring clinical practice [7]) and barriers in accessing and interpreting recommendations during routine clinical and public health practice. Furthermore, a guideline alone is not sufficient for changing entrenched practices and often requires a combination of implementation strategies, including educational and training materials, convening conferences and meetings, auditing and providing feedback, encouraging the participation of opinion leaders, and other approaches [10]. Lastly, the variability and complexity of local conditions require significant adaptations of global guidelines, thus driving a push towards more agile, adaptive, and efficient approaches enabled by digital technologies to implement guidelines at local levels.

A complementary approach to overcoming some of the constraints associated with diffusing and incorporating static guideline recommendations into the delivery of health services is to utilize digital technologies. In particular, digital

decision-support systems have emerged as an innovative mechanism to facilitate the uptake of guideline recommendations and reinforce adherence to clinical practices at the point of care [3,11-13]. Digital decision support consists of digitized job aids that can combine an individual's health information with the health worker's knowledge and clinical protocols to assist health workers in making diagnosis and treatment decisions [11,14,15]. Different forms of digital decision support include prompts and alerts based on a clinical algorithm, checklists according to a clinical protocol, and screening tools to identify risk and prioritize patients. Increasingly, digital decision support tools are being integrated with patient record systems to promote continuity of care and provide a more comprehensive history of the patient to bolster clinical decision-making. Despite the increasing availability of decision support systems, the accuracy of the underlying health content and adherence to the WHO (or other national bodies) guideline recommendations cannot always be verified. As such, WHO's role in monitoring the adaptation of guidelines in digital form is a necessity.

In November 2016, the WHO published the first evidence-based guideline on antenatal care (ANC), entitled “WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience” [16]. This guideline is a comprehensive document with 49 recommendations, encompassing nutritional interventions, maternal and fetal assessment, preventative interventions, common physiological symptoms, and health system interventions [17]. The recommendations ensure quality care throughout pregnancy and are based on a holistic consideration of comparative effects, values, resources, equity, acceptability, and feasibility [17].

While this guideline serves as the global standard for managing routine ANC, the guideline development team recognized that its adoption and uptake would be limited without effective implementation strategies. Considering the increasing ubiquity of mobile devices [18] and emerging uses of decision support tools, the guideline development team embarked on developing a digital decision support and health record module, which would contain the recommendations for routine use at the point of care. However, the structure of the ANC guideline, as with many other guidelines, did not readily lend itself for directly



encoding into a digital system and therefore required an additional year-long process to extract the recommendations for use in a point-of-care digital system [3,6].

This paper details the systematic and iterative process involved in translating the WHO ANC guidelines into a digital decision-support and patient-record software intended for routine use in primary health care settings, hereafter referred to as the WHO digital ANC module. By describing the methods and outputs, this paper aims to offer a replicable model for distilling and optimizing guideline content for digital systems.

## Methods

### The Process

The distillation of the ANC guideline content for a digital decision-support and client-record tool required a multidisciplinary team and a series of sequential but iterative steps. These steps, detailed below, include defining a minimum viable product (MVP), developing clinical workflows and algorithms, algorithm testing, developing a data dictionary, and the creation of a user interface or application development. The overall process of development took approximately 11 months to reach a stable prototype and finalize the underlying content requirements of the data dictionary and decision support algorithms.

### Establishing Fundamental Principles

The WHO convened a team of clinical and digital health experts (henceforth called “the team”) to lead the process of developing the WHO digital ANC module. In addition to WHO staff representing maternal health and digital health expertise, the team comprised obstetrician/gynecologists from the Brazilian Center for Research in Reproductive Health of Campinas (CEMICAMP, the University of Campinas’ research arm), digital and public health-implementation experts from the SUMMIT Institute for Development, and a technology social enterprise, Ona, which is the lead technology partner developing Open Smart Register Platform (OpenSRP) open-source software. The team worked mostly virtually and held 2 in-person meetings (in November 2017 and February 2018) throughout the development process.

The team’s aim was to develop a tool that assists health care professionals in implementing WHO evidence-based recommendations when providing quality medical care to pregnant women. Initially, the team established fundamental requirements for the digital tool, which would provide decision support and a longitudinal patient record of ANC based on WHO’s evidence-based recommendations. These overarching requirements included being user-friendly and applying human-centered design principles, adaptability to a wide range of contexts (ie, based on the availability of supplies, equipment, and human resources), and the ability to monitor pregnancy progression. Further, the team sought to create a comprehensive digital tool that would be able to link with existing health and logistical management information systems (Health Management Information Systems and Logistics Management Information Systems) and local electronic medical records (EMRs). Additionally, the tool would need to generate information and

monitor the health care process during the pregnancy of all women.

### Defining the Minimum Viable Product (MVP)

To set the parameters for the software’s MVP, the first iteration is designed for the primary health care (PHC) level to generate individual patient information (simple EMR) and to monitor the health care process as services are provided to all women during pregnancy. In line with the WHO recommendations for optimizing health worker roles for maternal and newborn health through task shifting, the ANC module was designed for the health worker cadre at the PHC level, namely, auxiliary nurse midwives, nurses, and midwives [19]. Future iterations, however, will be comprised of interlinked modules for community-level service provision and a patient version for women, allowing for management, feedback, and information exchange.

The first MVP provides decision support through prompts and alerts, checklists, and risk screening by health status based on WHO guidelines and clinical protocols. The WHO digital ANC module is a generic digital tool broadly applicable across all 49 ANC WHO recommendations, focusing on routine ANC (including screening and referral but not treatment for complications). Furthermore, the software documents whether required measures were implemented, and if they were not, then why not, which is important for decision-making, real-time feedback, and quality control. It is, however, customizable to diverse country contexts.

In addition to providing decision support (2.3 of WHO’s Classification of Digital Health Interventions), the module will help PHC workers identify, register, and longitudinally track pregnant women’s health status and services (2.1 & 2.2), as well as support their activity planning, scheduling (2.7), and referral coordination (2.6) [15]. Additionally, in planning for the future, the tool would be established in a way that could eventually employ machine learning algorithms, in anticipation of subsequent incorporation of artificial intelligence and predictive analytics.

### Developing Clinical Workflows and Algorithms

To better understand the standard patient and health worker’s actions during routine ANC contacts, the team developed relevant business processes and workflows (Figure 1) conducted by the PHC worker during ANC service delivery. The WHO’s *Pregnancy, Childbirth, Postpartum, and Newborn Care Guideline: A Guide for Essential Practice* served as a basis for this process, as it describes which procedures should be performed in clinical sequence [20]. While it is expected that each countries’ ANC workflows will vary from this generic idea, the team applied the Pareto principle, in which 80% of the content applies to all settings and 20% must be modified to local needs. Additionally, the workflows and the contact schedule (Figure 2) served as a framework for the module’s content and design.

Initially, for algorithm development, the WHO ANC recommendations were analyzed by 2 independent reviewers (SMH and RTS). Recommendations were identified, extracted, and split into spreadsheets according to their respective areas:

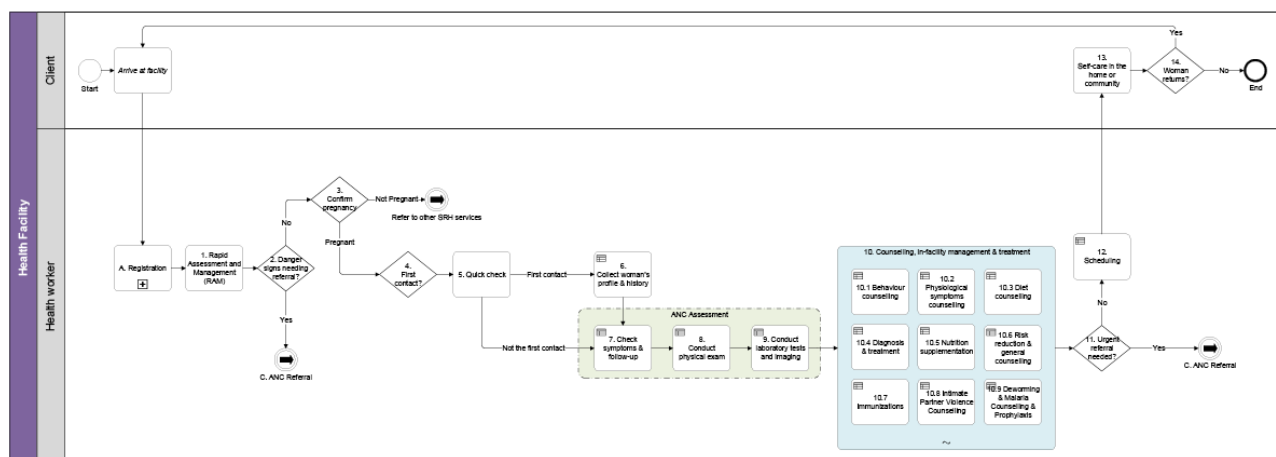


nutritional interventions, maternal and fetal assessment, preventive measures, common physiological symptoms, and health care system interventions. Recommendations were classified in line with the ANC guideline's categorizations: "universal" when they could be implemented in any scenario; "context-specific" when characteristics of the population were necessary to trigger the actions; "context-specific research"

when further studies were needed to formalize the recommendation; and "not recommended" when the action should be avoided due to a lack of a sufficient scientific basis to support its performance.

Upon further review, the team removed the "context-specific research" recommendations from the MVP.

**Figure 1.** The antenatal care (ANC) contact business process. ANC: Antenatal care, SRH: sexual and reproductive health.



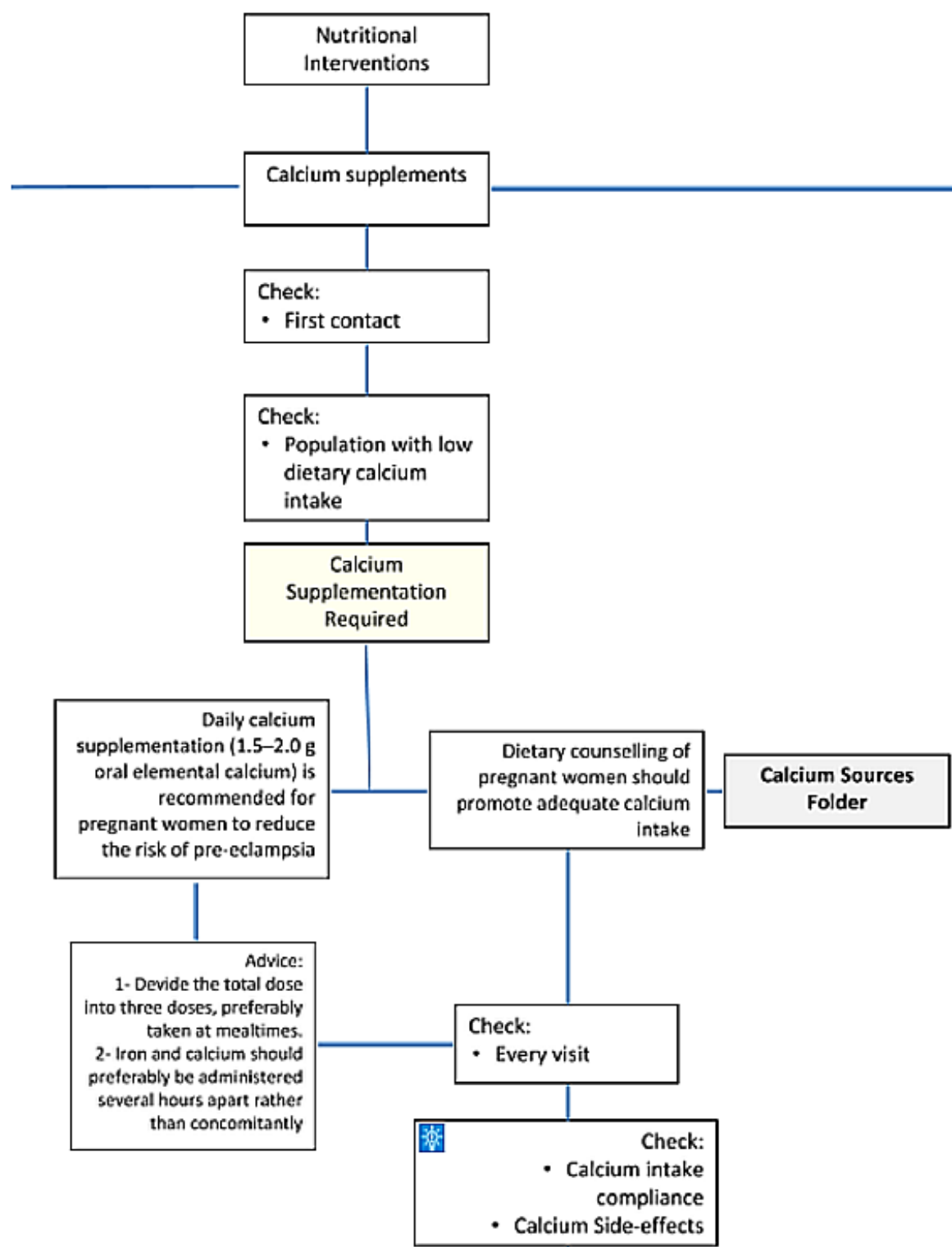
Other official WHO guidelines were consulted for clarification or for additional recommendations considered important for routine ANC (ie, influenza vaccination, pre-eclampsia screening, and hepatitis B and C testing [21-24]). When necessary, some adaptations were made to align with the clinical logic of the ANC service delivery process. These suggestions were assessed and agreed upon by the team. Decision trees were then created for each of the ANC and additional recommendations, considering parameters that trigger action during an ANC contact (for example, calcium supplementation after 20 weeks

gestation in women at high-risk for pre-eclampsia). Decision trees were also shaped by population-level characteristics (eg, folic acid and elemental iron supplementation according to the prevalence of anemia in the population), the local availability of health care equipment (eg, type of urine or syphilis tests depending on the laboratory facilities), and individual-level variables (eg, hepatitis C testing, if injectable drug user). Thus, a single algorithm structure or decision tree was developed to take all of these elements into consideration (Figure 3).

**Figure 2.** Antenatal care (ANC) 8-contact delivery schedule.

Interval Rules & Deviations										
Rule 1	< 20 weeks, contacts should be scheduled every 8 weeks									
Rule 2	≥ 20 weeks, contacts should be scheduled every 6 weeks									
Rule 3	≥ 26 weeks, contacts should be scheduled every 4 weeks									
Rule 4	≥ 34 weeks, contacts should be scheduled every 2 weeks									
Rule 5	if the logic skips the 40 weeks contact, she must come in, so add									
Rule 6	the woman MUST come in at 41 weeks if she hasn't given birth already (however this does not count as an official contact)									
Rule 7	if the woman comes in for the first time at 22 weeks or later, we have changed the logic to increase the number of visits, especially those in the 3rd trimester, as follows: - for women who come in at 22 and 23 weeks (they have one visit added) by increasing visits to every 2 weeks after 30 weeks (rather than 34 weeks) - for women who come in at 28 and 29 weeks (they have one visit added) by increasing visits to every 2 weeks after 32 weeks (rather than 34 weeks) - for women who come in after 30 weeks all her visits are every 2 weeks									
Key										
	cells that deviate from the norm because of Rule 7									
	doesn't count as a contact, delivery only									

	Contact 1	Contact 2	Contact 3	Contact 4	Contact 5	Contact 6	Contact 7	Contact 8	Contact 9	Delivery	
End of 1st trimester	4	12	20	26	30	34	36	38	40	41	
	5	13	21	27	31	35	37	39	40	41	
	6	14	22	28	32	36	38	40	41		
	7	15	23	29	33	37	39	40	41		
	8	16	24	30	34	36	38	40	41		
	9	17	25	31	35	37	39	40	41		
	10	18	26	30	34	36	38	40	41		
	11	19	27	31	35	37	39	40	41		
	12	20	26	30	34	36	38	40	41		
	13	21	27	31	35	37	39	40	41		
	Middle of pregnancy	14	22	28	32	36	38	40	41		
		15	23	29	33	37	39	40	41		
16		24	30	34	36	38	40	41			
17		25	31	35	37	39	40	41			
18		26	30	34	36	38	40	41			
19		27	31	35	37	39	40	41			
20		26	30	34	36	38	40	41			
21		27	31	35	37	39	40	41			
22		28	32	34	36	38	40	41			
23		29	33	35	37	39	40	41			
24		30	34	36	38	40	41				
25		31	35	37	39	40	41				
26		30	34	36	38	40	41				
27		31	35	37	39	40	41				
28		32	34	36	38	40	41				
29		33	35	37	39	40	41				
30		32	34	36	38	40	41				
31		33	35	37	39	40	41				
32		34	36	38	40	41					
33		35	37	39	40	41					
End of 2nd trimester		34	36	38	40	41					
	35	37	39	40	41						
	36	38	40	41							
	37	39	40	41							
	38	40	41								
	39	40	41								
	40	41									
	41										

**Figure 3.** Example of the clinical algorithm: calcium supplementation recommendation.

### Algorithm Testing

Prior to the software's development, a test version was constructed (by SMH) using Excel resources that included diverse functions linking variables across different worksheets. The first version of the system design was outlined in the Excel testing form, creating an interactive spreadsheet (Figure 4). The testing form displayed real-time clinical diagnoses, the ANC recommendations, and additional guidance on the health care

process, all automatically elaborated by the system during data entry and guided by the logic underlying the clinical algorithms. Initial tests were performed using fabricated patient cases that explored data relative to all algorithms, probing all decision trees. The logic of the algorithm was thus reviewed and corrected, where necessary. Furthermore, the Excel prototype was helpful to visualize how the system could interact with the user in real time as a tool for support in clinical decisions in ANC.

**Figure 4.** Excel testing form.

First Contact			
Current Pregnancy Status			
Estimated Gestational Age (weeks)	12		
Number of fetuses	Multiple		
Previous pregnancy's Complications			
Yes	History of Pre-eclampsia or History of eclampsia		
No	Convulsions		
Yes	Tobacco		
Yes	Alcohol use or illicit substance use	If Drugs use, choose	Injectable drugs
Personal Background			
Pregestational weight (kg)	60		
Height (m)	1.6		
Body Mass Index (kg/m2)	23.4		
Disease/Risk/Complication prior to this pregnancy			
No	Diabetes		
No	Chronic hypertension		
No	Renal Disease		
No	Autoimmune disease		

Summary Panel - 1st Contact			
General Obstetric Diagnosis			
GA 12	Multiple		
Previous Pathological Obstetric Diagnosis			
History of Pre-eclampsia or History of eclampsia			
FALSE			
Tobacco			
Alcohol use or illicit substance use			
Current Pathological Obstetric Diagnosis			
Risk of Pre-eclampsia			
#N/A			
#N/A			
Clinical Diagnosis			
Normal weight		Weight gain expected 11.5-16kg	
		Total gain -60	
No anaemia detected			
FALSE			
#N/A			
FALSE			
Woman at risk of Intimate Partner Violence			
FALSE			
FALSE			
FALSE			
FALSE			
FALSE			
FALSE			
#N/A			
FALSE			
#N/A			
#N/A			
#N/A			
FALSE			
Preventive Measures Status			

## Development of the Data Dictionary

After a detailed review of the algorithms, the data elements which comprised them were separated into sections, including population and local setting characteristics, socio-demographic data, information on current pregnancy, medical (including immunization) and obstetric history, lifestyle and habits, physiological symptoms of pregnancy, physical examination and laboratory tests, counseling, and treatment. Initially, only data elements required to trigger recommendation-related algorithms were included in the data entry forms. As an example, basic variables for the early identification of sexually transmitted infections (STIs) were included, without further development of the complete algorithm for the differential diagnosis and management of these conditions (except syphilis and HIV, which are included in the ANC guidelines). For STIs as well as other issues, algorithms allow for the health worker to identify risk and subsequently support referral for further investigation or treatment.

From these efforts, the team constructed the data dictionary, which is a shared Google spreadsheet that contains all the clinical content and decision-support logic to be used to create

the data entry forms and decision-support workflows in the WHO digital ANC module (Figure 5). At this stage, the team also incorporated additional content in the data dictionary for the complete provision of routine ANC (eg, fetal heart rate, vaginal examination, maternal pulse, blood pressure, and temperature measurement), which are assumed standard clinical practices and therefore are not part of the ANC guidelines. The data dictionary served as the primary technical requirement for the software developers to build the MVP.

Additionally, all the variables in the data dictionary were mapped to standardized medical terminology codes, such as the International Classification of Diseases (ICD), Logical Observation Identifiers Names and Codes (LOINC), and Systematized Nomenclature of Medicine (SNOMED). Thus, the team also sought to standardize terminologies to ensure interoperability and ease of data aggregation for future integration with databases. For clinical care, terminologies are structured vocabularies covering health related concepts, such as diseases, diagnoses, laboratory tests, and treatments, to enable the storage, analysis, and exchange of data in a consistent and standard way [25,26].

**Figure 5.** Data Dictionary - Profile Tab. ANC: antenatal care; GA: gestational age; HIV: human immunodeficiency viruses, LMP: last menstrual period; MC: multiple choice; NA: not applicable; SFH: symphysis fundal height; STI: sexually transmitted infection.

Label	Info Icon	Due	Relevance	Name	Description	Type	Choices	Calculation
<b>CURRENT PREGNANCY</b>								
LMP known?	LMP = first day of Last Menstrual Period. If the exact date is unknown, but the period of the month is known, use day 5 for beginning of the month, day 15 for middle of the month and day 25 for end of the month. If completely unknown, select "No" and calculate GA from ultrasound (or SFH or abdominal palpation as a last resort).	First contact		lmp_known	Whether woman's LMP date is known or not.	MC (select one)	Yes No	
Specify date		First contact	[lmp_known] = "Yes"	lmp	The woman's Last Menstrual Period (LMP) date. This is defined as the first day of her most recent period.	Date		
NA		NA		lmp_edd	Expected Date of Delivery (EDD) calculated from LMP.	Calculation		[lmp] + 280 days
NA		NA		lmp_gest_age	Gestational age (GA) calculated from LMP.	Calculation		(today's date - [lmp]) / 7, shown in weeks and days
Ultrasound done?	An ultrasound is recommended for all women before 24 weeks gestation or even after if deemed necessary (e.g. to identify the number of fetuses, fetal presentation, or placenta location).	First contact		ultrasound_done	Whether or not the woman has had an ultrasound scan done at any point in this pregnancy.	MC (select one)	Yes No	
Ultrasound recommended	An ultrasound is recommended for all women before 24 weeks gestation or even after if deemed necessary (e.g. to identify the number of fetuses, fetal presentation, or placenta location).	NA	[ultrasound_done] = "No" and [site_ultrasound] = 1	NA	An ultrasound is recommended for all women before 24 weeks gestation or even after if deemed necessary (e.g. to identify the number of fetuses, fetal presentation, or placenta location). This informs the health worker to conduct the ultrasound at the health facility.	Toaster Message		
Refer for ultrasound in facility with U/S equipment	An ultrasound is recommended for all women before 24 weeks gestation or even after if deemed necessary (e.g. to identify the number of fetuses, fetal presentation, or placenta location).	NA	[ultrasound_done] = "No" and [site_ultrasound] = 0	NA	An ultrasound is recommended for all women before 24 weeks gestation or even after if deemed necessary (e.g. to identify the number of fetuses, fetal presentation, or placenta location). This informs the health worker to refer the woman to a facility where an ultrasound machine is available.	Toaster Message		
Ultrasound date		NA	[ultrasound_done] = yes	ultrasound_date	Date that the ultrasound was done.	Date		
GA from ultrasound - weeks		NA	[ultrasound_done] = yes	ultrasound_gest_age_wk	Gestational age in weeks from ultrasound.	Integer		
GA from ultrasound - days		NA	[ultrasound_done] = yes	ultrasound_gest_age_days	Gestational age in days from ultrasound.	Integer		
NA		NA	[ultrasound_done] = yes	ultrasound_gest_age_concept	Gestational age (in weeks) calculation used for GenIRS concept mapping. Decimal field.	Calculation		[ultrasound_gest_age_wk] * 7 + [ultrasound_gest_age_days] / 7

## Creation of User Interface and Application Development

As content was being developed in the data dictionary, the team also began creating mockups of the module's user interfaces, which allowed team members to collaboratively discuss, iterate on, and come to a consensus on the design and workflows required for the MVP. As the data dictionary content was completed over time, the team was able to enrich the interface designs with plausible clinical content, which further allowed for testing the designs with additional collaborators, including medically trained colleagues at the WHO and the Center for Research in Reproductive Health of Campinas (CEMICAMP), to elicit better feedback. During this iterative process, mockups were shared and discussed on Marvel [27]. After several rounds of feedback and iteration, the data dictionary and mockups were approved, which allowed the software development of the module to begin building it.

## Results

### Outputs

The methods described above resulted in 2 outputs. The first output is a reference software reflecting the generic WHO ANC guideline content, known as the WHO digital ANC module. The second output consists of the structured documentation of the different components which contributed to the development of the WHO digital ANC module, such as the data dictionary, clinical decision-support workflows, and mockups.

### WHO Digital ANC Module

The current version (1.0) of the WHO digital ANC module allows health workers to manage their patients' records as well as filter for women diagnosed with certain conditions, view attention flags, easily identify women overdue for contacts, and sort entries for gestational age, due date, and first and last name

(Figure 6A). Health workers can also see a summary of a woman's prior contacts and any attention flags; prompts noting conditions for close follow-up (eg, advanced maternal age, twin pregnancy) are highlighted in red (Figure 6B). For further details, all the information from the woman's prior contacts can also be accessed in the profile overview.

During a contact, the health worker can access the landing page (Figure 6C), which groups data entry forms based on the major clinical workflows that were identified during the design phase: Quick Check, Profile, Symptoms & Follow Up, Physical Exam, Tests, and Counselling & Treatment. Within these "containers," all ANC recommendations have a related question to record whether the health worker has performed the associated task. WHO recommendations that are not carried out are placed on a pending list of tasks, and those that are adequately fulfilled prompt a message with positive reinforcement. Other highlighted features include the following:

- Alerts to remind the health worker of a recommended action (eg, "An ultrasound is recommended for all women before 24-weeks gestation or even after if deemed necessary, to identify the number of fetuses, fetal presentation, or placenta location"), to provide positive reinforcement (eg, "Woman is fully immunized against tetanus!"), or to flag a risk (eg, "Pre-eclampsia risk counseling: the use of aspirin after 12 weeks gestation is recommended as well as calcium if low dietary intake area. Please also provide counseling")
- A task list of summarized actions to allow the health worker to identify potentially missed opportunities at the end of the contact, or to be reminded of pending actions (eg, test results that need to be recorded)
- Sparklines depicting trends in certain data elements (eg, weight, blood pressure, symphysis fundal height) (Figure 6D)
- Additional informative materials for health workers regarding nutrition counseling



**Figure 6.** Screenshots of the World Health Organization Digital Antenatal Care (WHO digital ANC) module. A: List of patients; B: Individual patient record summary; C: Home screen; D: Patient contact summary.



In addition, the MVP of the module also contains an interface for the manager at the health facility to set population levels (such as HIV prevalence) and site infrastructure parameters (such as the availability of ultrasound machines). For communication with the pregnant women, the MVP allows for the sending of SMS text messages (if the pregnant woman consents), and in a future version, there would be a complete interface sharing information as well as a virtual “pregnant woman card.” Finally, the MVP seeks to increase the use of data for accountability to drive quality service provision through the automated aggregation of key indicators and to increase the availability of dashboards for use by different stakeholders such as program managers.

### Test Cases

In order to assess the module’s components, 2 researchers (SMH and RS) elaborated 67 clinical test cases. The cases were designed to encompass the largest number of possible scenarios, testing all clinical algorithms. An online database was created using Google Forms, incorporating variables from the data dictionary. The team tested these cases to check the decision-support logic in the module and worked with the software developer to correct any issues discovered.

### Structured Documentation

Throughout the WHO digital ANC module’s development, the team meticulously documented the process for its creation. Firstly, during the definition of the MVP, targeted user personas were defined (PHC workers) and user stories were created (role-play of a typical ANC contact). Secondly, the aforementioned business processes, workflows, and contact schedule set the parameters for the module’s content (Figures 1 and 2). Next, the data dictionary aided in creating the core

data elements needed for the ANC service provision (Multimedia Appendix 1). The data elements linked together through the algorithms (based on the decision trees) and were also elaborated further in decision-support tables (Figure 7). Finally, the module provides a link between an individual’s data elements to the components (numerators and denominators), which constitute established ANC indicators to simplify the data aggregation process.

Beyond making the WHO digital ANC module available as a digital decision-support and client-record tool within OpenSRP, WHO is also making the structured documentation publicly available as digital accelerator kits [28]. This structured documentation is independent of the OpenSRP software and intended to accompany the WHO digital ANC module to assist countries in understanding the underlying data elements and algorithms to facilitate the customization of the ANC module to their context. Additionally, countries that may already have digital systems in place will be able to use the structured documentation (eg, the data dictionary and decision tables) to ensure that the content is aligned with the WHO guidelines, or at least to have a starting point from which they can design the digital system if they choose to use a separate software platform.

These different components of the structured documentation were derived from standardized methodologies in health informatics, most notably, business process mapping notation (BPMN) for the workflows, development of the data dictionary (which is itself a derivative of the XLS form standard), and decision model notation (DMN) for the decision logic. The formulation of the kits’ components was documented by the team, and the process of releasing structured documentation for digital systems is also being replicated across other health domains (eg, family planning, HIV).

Figure 7. Example decision-support table.

Decision ID	ANC.DT.31 Tetanus Toxoid-Containing Vaccination					
Business Rule	If the woman has not yet been fully immunized against tetanus, she should be immunized					
Trigger	ANC.B10.7 Immunizations					
Inputs				Output	Action	Annotations
'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'Under immunized'				'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'Under immunized'	Provide Tetanus toxoid immunizations - using the "Tetanus Toxoid-Containing Vaccine (TTCV) immunization - NO PREVIOUS" schedule (3 dose scheme)	Tetanus toxoid vaccination is recommended for all pregnant women who are not fully immunized against TT to prevent neonatal mortality from tetanus.
'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'No doses'				'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'No doses'	Provide Tetanus toxoid immunizations - using the "Tetanus Toxoid-Containing Vaccine (TTCV) immunization - NO PREVIOUS" schedule (3 dose scheme)	Tetanus toxoid vaccination is recommended for all pregnant women who are not fully immunized against TT to prevent neonatal mortality from tetanus.
'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'Unknown'				'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'Unknown'	Provide Tetanus toxoid immunizations - using the "Tetanus Toxoid-Containing Vaccine (TTCV) immunization - NO PREVIOUS" schedule (3 dose scheme)	Tetanus toxoid vaccination is recommended for all pregnant women who are not fully immunized against TT to prevent neonatal mortality from tetanus.
'Tetanus Toxoid-Containing Vaccine (TTCV) 1 immunization provided' = TRUE	'Tetanus Toxoid-Containing Vaccine (TTCV) 2 immunization provided' = TRUE	'Tetanus Toxoid-Containing Vaccine (TTCV) 3 immunization provided' = TRUE	'Tetanus Toxoid-Containing Vaccine (TTCV) 4 immunization provided' = TRUE	'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'Under immunized'	Provide Tetanus toxoid immunizations - using the "Tetanus Toxoid-Containing Vaccine (TTCV) immunization - WITH PREVIOUS" schedule (1 dose scheme)	1-4 doses of TTCV in the past: - 1 dose scheme: Woman should receive one dose of TTCV during each subsequent pregnancy to a total of five doses (five doses protect throughout the childbearing years).
'Tetanus Toxoid-Containing Vaccine (TTCV) 1 immunization provided' = TRUE	'Tetanus Toxoid-Containing Vaccine (TTCV) 2 immunization provided' = TRUE	'Tetanus Toxoid-Containing Vaccine (TTCV) 3 immunization provided' = TRUE		'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'Under immunized'	Provide Tetanus toxoid immunizations - using the "Tetanus Toxoid-Containing Vaccine (TTCV) immunization - WITH PREVIOUS" schedule (1 dose scheme)	1-4 doses of TTCV in the past: - 1 dose scheme: Woman should receive one dose of TTCV during each subsequent pregnancy to a total of five doses (five doses protect throughout the childbearing years).
'Tetanus Toxoid-Containing Vaccine (TTCV) 1 immunization provided' = TRUE	'Tetanus Toxoid-Containing Vaccine (TTCV) 2 immunization provided' = TRUE			'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'Under immunized'	Provide Tetanus toxoid immunizations - using the "Tetanus Toxoid-Containing Vaccine (TTCV) immunization - WITH PREVIOUS" schedule (1 dose scheme)	1-4 doses of TTCV in the past: - 1 dose scheme: Woman should receive one dose of TTCV during each subsequent pregnancy to a total of five doses (five doses protect throughout the childbearing years).
'Tetanus Toxoid-Containing Vaccine (TTCV) 1 immunization provided' = TRUE				'Tetanus Toxoid-Containing Vaccine (TTCV) immunization history' = 'Under immunized'	Provide Tetanus toxoid immunizations - using the "Tetanus Toxoid-Containing Vaccine (TTCV) immunization - WITH PREVIOUS" schedule (1 dose scheme)	1-4 doses of TTCV in the past: - 1 dose scheme: Woman should receive one dose of TTCV during each subsequent pregnancy to a total of five doses (five doses protect throughout the childbearing years).

## Discussion

There is a growing recognition of the multitude of software applications targeted at health workers [29,30]; however, the WHO digital ANC module reflects a systematic effort to ensure that the latest WHO clinical guidelines are available in digital form for routine use. As such, the distinguishing factor of the WHO digital ANC module is not necessarily the digital functionalities but rather the availability of the 2016 WHO ANC guideline content encoded in a format that countries can readily adapt and incorporate into service delivery [3].

Challenges in translating clinical guideline content to digital systems include potential misinterpretations that may result from not having the precise logic and data elements within the

guideline itself [3,31]. Furthermore, software development companies building digital decision support tools are often required to start from scratch in translating guideline content to digital workflows. With the creation of the WHO digital ANC module and accompanying structured documentation, the challenges associated with misinterpretation of the guideline content or unnecessary replication in distilling the recommendations are thus minimized. In settings where there are no digital systems for routine ANC in place, countries can leverage the WHO digital ANC module as an out-of-the-box generic digital system that can then be easily adapted to their local contexts instead of starting from scratch. Alternatively, in settings where electronic medical records and other digital systems already exist, the structured documentation in the WHO's forthcoming digital accelerator kits can be used to

inform the content within the digital system. Additionally, the ANC digital accelerator kit (upon which the module is based) aims to simplify and standardize individual-level and aggregate data collection, a key component to tracking guideline implementation at national and global levels.

The development of the WHO digital ANC module also resonates with an established approach developed by Boxawala et al [3] to manage the translation of narrative guideline recommendations to digital decision support tools. This framework consists of 4 sequential knowledge layers: the unstructured narrative recommendation as written in the guideline (Layer 1); a semi-structured derivative of the guideline recommendation in the form of decision trees, flowcharts, and data logic (Layer 2); a computable form of the recommendation represented in software code (Layer 3); and an executable digital decision-support tool that can be used directly by health workers (Layer 4) [3]. In our application of this framework, the WHO ANC recommendations represent Layer 1, the structured documentation of the digital accelerator kit corresponds to Layer 2, and the ANC module consists of Layer 3. Currently, Layer 4 is under development and seeks to translate the ANC recommendations into a computable and coded format that can be directly ingested by a digital decision-support system. This includes Fast Healthcare Interoperability Resources (FHIR), which will be embedded into the ANC module. This will facilitate the export of data in a standardized way for analysis by CQL and other methods optimized for FHIR, increasing interoperability.

Following the completion of the module, the team conducted user testing of the module with nurses and midwives in Indonesia and physicians in Brazil. These professionals were not part of the previous stages of the module's creation. A standardized protocol was developed to train health workers to use the module and elicit feedback from these health workers through focus group discussions and in-depth interviews. From these experiences, feedback was compiled, and the team made changes where deemed necessary prior to the generic version 1.0's release. The WHO digital ANC module (version 1.0) and underlying standardized documentation will be released for

global use to support countries and health care providers to adapt and adopt the WHO ANC recommendations.

We believe that this first version of the WHO digital ANC module is robust; however, future efforts are currently underway to support its implementation in various countries. As part of this process, the team is developing training materials (videos, handbooks, presentations, etc) for health workers at primary health centers. Additionally, a framework and indicators for monitoring, evaluating, and supervising the module's implementation and its impact on health worker performance will be created [32-34]. It is expected that the WHO digital ANC module will support the implementation of evidence-based practices and provide information for monitoring and surveillance; however, further evidence is needed to understand how the ANC module impacts the implementation of the WHO recommendations. The WHO is developing an implementation research study to assess whether the use of the module improves the fidelity to WHO ANC recommendations in various countries.

Ultimately, prescriptive and rigid recommendations are not appropriate or optimal, and over time, the WHO has moved toward context-specific recommendations and has developed tools that allow for the design of health care services based on local settings [35]. However, at the same time, doing so has also changed the complexity of interpreting and implementing recommendations. Digital tools have the potential to facilitate and accelerate the implementation of such guidelines on the ground. The WHO digital ANC module is a pioneering effort to create a roadmap for that process, which will inevitably be needed for policies across other health domains. The adaptive processes required for the implementation of the WHO ANC recommendations were enhanced at each step in the development of the WHO digital ANC module. This is a novel approach to facilitating the adoption and adaptation of guidelines through digital systems at the health delivery service level. The WHO will release the module for countries' adaptation and use. Further, the module's implementation will inform the WHO's ongoing efforts to create a pathway to adaptive and integrated (Smart) Guidelines in Digital Systems to improve health system quality, coverage, and accountability.

## Acknowledgments

This work was funded by the Sanofi Espoir Foundation, Department for International Development (DFID), as well as the WHO Department of Reproductive Health and Research through the Special Programme of Research, Development, and Research Training in Human Reproduction. The development of the clinical algorithms was supported by the Brazilian National Research Council (CNPq) through a grant for fellowship (435614/2016-9).

The authors alone are responsible for the views expressed in this article, and they do not necessarily represent the views, decisions, or policies of the institutions with which they are affiliated.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Overview of data element categories.

[DOCX File, 25 KB - [jmir\\_v22i10e16355\\_app1.docx](#)]

## References

1. World Health Organization. WHO handbook for guideline development. WHO. Geneva: The World Health Organization; 2014. URL: <https://apps.who.int/iris/handle/10665/145714> [accessed 2020-09-15]
2. Wang Z, Norris SL, Bero L. The advantages and limitations of guideline adaptation frameworks. *Implementation Sci* 2018 May 29;13(1):72. [doi: [10.1186/s13012-018-0763-4](https://doi.org/10.1186/s13012-018-0763-4)] [Medline: [29843737](https://pubmed.ncbi.nlm.nih.gov/29843737/)]
3. Boxwala AA, Rocha BH, Maviglia S, Kashyap V, Meltzer S, Kim J, et al. A multi-layered framework for disseminating knowledge for computer-based decision support. *Journal of the American Medical Informatics Association* 2011 Dec 01;18(Supplement 1):i132-i139. [doi: [10.1136/amiainl-2011-000334](https://doi.org/10.1136/amiainl-2011-000334)] [Medline: [22052898](https://pubmed.ncbi.nlm.nih.gov/22052898/)]
4. Nabyonga Orem J, Bataringaya Wavamunno J, Bakeera SK, Criel B. Do guidelines influence the implementation of health programs? — Uganda's experience. *Implementation Sci* 2012 Oct 15;7(1):98. [doi: [10.1186/1748-5908-7-98](https://doi.org/10.1186/1748-5908-7-98)] [Medline: [2306808](https://pubmed.ncbi.nlm.nih.gov/2306808/)]
5. Maaløe N, Meguid T, Housseine N, Tersbøl BP, Nielsen KK, Bygbjerg IC, et al. Local adaption of intrapartum clinical guidelines, United Republic of Tanzania. *Bull. World Health Organ* 2019 Mar 26;97(5):365-370. [doi: [10.2471/blt.18.220830](https://doi.org/10.2471/blt.18.220830)] [Medline: [31551633](https://pubmed.ncbi.nlm.nih.gov/31551633/)]
6. Bilici E, Despotou G, Arvanitis TN. The use of computer-interpretable clinical guidelines to manage care complexities of patients with multimorbid conditions: A review. *DIGITAL HEALTH* 2018 Oct 03;4:205520761880492. [doi: [10.1177/2055207618804927](https://doi.org/10.1177/2055207618804927)] [Medline: [30302270](https://pubmed.ncbi.nlm.nih.gov/30302270/)]
7. Puchalski Ritchie LM, Khan S, Moore JE, Timmings C, van Lettow M, Vogel JP, et al. Low- and middle-income countries face many common barriers to implementation of maternal health evidence products. *Journal of Clinical Epidemiology* 2016 Aug;76:229-237. [doi: [10.1016/j.jclinepi.2016.02.017](https://doi.org/10.1016/j.jclinepi.2016.02.017)] [Medline: [26931284](https://pubmed.ncbi.nlm.nih.gov/26931284/)]
8. Panisset U, Koehlmoos TP, Alkhatib AH, Pantoja T, Singh P, Kenney-Kayondo J, et al. Implementation research evidence uptake and use for policy-making. *Health Res Policy Sys* 2012 Jul 2;10(1):20. [doi: [10.1186/1478-4505-10-20](https://doi.org/10.1186/1478-4505-10-20)] [Medline: [22748142](https://pubmed.ncbi.nlm.nih.gov/22748142/)]
9. Vogel JP, Moore JE, Timmings C, Khan S, Khan DN, Defar A, et al. Barriers, Facilitators and Priorities for Implementation of WHO Maternal and Perinatal Health Guidelines in Four Lower-Income Countries: A GREAT Network Research Activity. *PLoS ONE* 2016 Nov 2;11(11):e0160020. [doi: [10.1371/journal.pone.0160020](https://doi.org/10.1371/journal.pone.0160020)] [Medline: [27806041](https://pubmed.ncbi.nlm.nih.gov/27806041/)]
10. Oxman A, Thomson M, Davis D, Haynes R. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *CMAJ* 1995;153(10):1423-1431. [Medline: [7585368](https://pubmed.ncbi.nlm.nih.gov/7585368/)]
11. World Health Organization. WHO Guideline Recommendations on Digital Interventions for Health System Strengthening. WHO Guideline Recommendations on Digital Interventions for Health System Strengthening. Geneva: WHO; 2019. URL: <https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/> [accessed 2020-09-15]
12. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005 Mar 04;330(7494):765. [doi: [10.1136/bmj.38398.500764.8f](https://doi.org/10.1136/bmj.38398.500764.8f)] [Medline: [15767266](https://pubmed.ncbi.nlm.nih.gov/15767266/)]
13. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes. *JAMA* 2005 Mar 09;293(10):1223-1238. [doi: [10.1001/jama.293.10.1223](https://doi.org/10.1001/jama.293.10.1223)] [Medline: [15755945](https://pubmed.ncbi.nlm.nih.gov/15755945/)]
14. World Health Organization, International Telecommunication Union. National eHealth strategy toolkit. World Health Organization and International Telecommunication Union. Geneva: World Health Organization and International Telecommunication Union; 2012. URL: <https://apps.who.int/iris/handle/10665/75211> [accessed 2020-09-15]
15. World Health Organization. Classification of digital health interventions v1.0. A shared language to describe the uses of digital technology for health. Geneva: WHO; 2018. URL: <http://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf>
16. World Health Organization. WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience. World Health Organization. Geneva: WHO; 2016. URL: [https://www.who.int/reproductivehealth/publications/maternal\\_perinatal\\_health/anc-positive-pregnancy-experience/en/](https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/anc-positive-pregnancy-experience/en/) [accessed 2020-09-15]
17. Tunçalp O, Pena-Rosas J, Lawrie T, Bucagu M, Oladapo O, Portela A, et al. WHO recommendations on antenatal care for a positive pregnancy experience-going beyond survival. *BJOG* 2017 Mar 09;124(6):860-862. [doi: [10.1111/1471-0528.14599](https://doi.org/10.1111/1471-0528.14599)] [Medline: [28190290](https://pubmed.ncbi.nlm.nih.gov/28190290/)]
18. International Telecommunication Union. ICT Facts & Figures: The world in 2015. ITU. Geneva: ICT Data and Statistics Division; 2015. URL: <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf> [accessed 2020-09-15]
19. World Health Organization. WHO recommendations: Optimizing health worker roles to improve access to key maternal and newborn health interventions through task shifting Geneva2012. World Health Organization. Geneva: WHO; 2012. URL: [https://www.who.int/reproductivehealth/publications/maternal\\_perinatal\\_health/978924504843/en/](https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/978924504843/en/) [accessed 2020-09-15]



20. World Health Organization. Pregnancy, childbirth, postpartum and newborn care: a guide for essential practice. WHO. Pregnancy, childbirth, postpartum and newborn care: a guide for essential practice – 3rd ed. Geneva; 2015. URL: [https://www.who.int/maternal\\_child\\_adolescent/documents/imca-essential-practice-guide/en/](https://www.who.int/maternal_child_adolescent/documents/imca-essential-practice-guide/en/) [accessed 2020-09-15]
21. World Health Organization. WHO recommendations for prevention and treatment of preeclampsia and eclampsia. WHO. Geneva: WHO; 2011. URL: [https://www.who.int/reproductivehealth/publications/maternal\\_perinatal\\_health/9789241548335/en/](https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/9789241548335/en/) [accessed 2020-09-15]
22. World Health Organization. Managing complications in pregnancy and childbirth: a guide for midwives and doctors. WHO. Geneva: WHO; 2017. URL: [https://www.who.int/maternal\\_child\\_adolescent/documents/managing-complications-pregnancy-childbirth/en/](https://www.who.int/maternal_child_adolescent/documents/managing-complications-pregnancy-childbirth/en/) [accessed 2020-09-15]
23. World Health Organization. Guidelines on Hepatitis B and C Testing. WHO. Geneva: WHO; 2017. URL: <https://www.who.int/hepatitis/publications/guidelines-hepatitis-c-b-testing/en/> [accessed 2020-09-15]
24. World Health Organization. Vaccines against influenza WHO position paper - November 2012. Wkly Epidemiol Rec 2012;87(47):461-476. [Medline: [23210147](#)]
25. Australian Government. Australian Digital Health Agency. URL: <https://www.digitalhealth.gov.au/get-started-with-digital-health/what-is-digital-health/clinical-terminology> [WebCite Cache ID <https://www.digitalhealth.gov.au/get-started-with-digital-health/what-is-digital-health/clinical-terminology>]
26. World Health Organization. Classifications: International Classification of Diseases (ICD) Information Sheet 2020. URL: <https://www.who.int/classifications/icd/factsheet/en/> [accessed 2020-09-16]
27. Marvel Enterprise. Marvel: The all-in-one platform powering design 2019. Marvel Enterprise 3. 2019. URL: <https://marvelapp.com/enterprise> [accessed 2020-09-15]
28. World Health Organization. Background on Accelerator Kits for Digital Client Records for Sexual, Reproductive, Maternal, Neonatal, Child and Adolescent Health (SRMNCAH). In: WHO. 2019 Presented at: Meeting presentation; 2019; Geneva.
29. Haddad SM, Souza RT, Cecatti JG. Mobile technology in health (mHealth) and antenatal care—Searching for apps and available solutions: A systematic review. International Journal of Medical Informatics 2019 Jul;127:1-8. [doi: [10.1016/j.ijmedinf.2019.04.008](https://doi.org/10.1016/j.ijmedinf.2019.04.008)] [Medline: [31128820](#)]
30. World Health Organization. Digital Health Atlas 2019. World Health Organization. Geneva: WHO; 2019. URL: <https://digitalhealthatlas.org/en/-/> [accessed 2020-09-15]
31. Codish S, Shiffman RN. A model of ambiguity and vagueness in clinical practice guideline recommendations. AMIA Annu Symp Proc 2005;2005:146-150. [Medline: [16779019](#)]
32. Lattof SR, Tunçalp O, Moran AC, Bucagu M, Chou D, Diaz T, et al. Developing measures for WHO recommendations on antenatal care for a positive pregnancy experience: a conceptual framework and scoping review. BMJ Open 2019 Apr 24;9(4):e024130. [doi: [10.1136/bmjopen-2018-024130](https://doi.org/10.1136/bmjopen-2018-024130)] [Medline: [31023748](#)]
33. Lattof S, Moran A, Moller A, Jayathilaka C, Diaz T, Tunçalp Ö. Implementation of the new WHO antenatal care model for a positive pregnancy experience: a monitoring framework 2020;5 [FREE Full text] [doi: [10.1136/bmjgh-2020-002605](https://doi.org/10.1136/bmjgh-2020-002605)]
34. Shankar AV, Asrilla Z, Kadha JK, Sebayang S, Apriatni M, Sulastris A, et al. Programmatic Effects of a Large-Scale Multiple-Micronutrient Supplementation Trial in Indonesia: Using Community Facilitators as Intermediaries for Behavior Change. Food Nutr Bull 2009 Jun 23;30(2\_suppl2):S207-S214. [doi: [10.1177/15648265090302s204](https://doi.org/10.1177/15648265090302s204)] [Medline: [20496613](#)]
35. Barreix M, Lawrie TA, Kidula N, Tall F, Bucagu M, Chahar R, et al. Development of the WHO Antenatal Care Recommendations Adaptation Toolkit: a standardised approach for countries. Health Res Policy Syst 2020 Jun 22;18(1):70 [FREE Full text] [doi: [10.1186/s12961-020-00554-4](https://doi.org/10.1186/s12961-020-00554-4)] [Medline: [32564777](#)]

## Abbreviations

**ANC:** antenatal care  
**EMR:** electronic medical record  
**FHIR:** Fast Healthcare Interoperability Resources  
**MVP:** minimum viable product  
**OpenSRP:** Open Smart Register Platform open-source software  
**PHC:** primary health care  
**STI:** sexually transmitted infection  
**WHO:** World Health Organization



*Edited by G Eysenbach; submitted 21.09.19; peer-reviewed by D Mohan, S Sarbadhikari; comments to author 12.06.20; revised version received 26.06.20; accepted 26.07.20; published 01.10.20.*

*Please cite as:*

Haddad SM, Souza RT, Cecatti JG, Barreix M, Tamrat T, Footitt C, Mehl GL, Syah IF, Shankar AH, Tunçalp Ö

*Building a Digital Tool for the Adoption of the World Health Organization's Antenatal Care Recommendations: Methodological Intersection of Evidence, Clinical Logic, and Digital Technology*

*J Med Internet Res* 2020;22(10):e16355

URL: <https://www.jmir.org/2020/10/e16355>

doi: [10.2196/16355](https://doi.org/10.2196/16355)

PMID: [33001032](https://pubmed.ncbi.nlm.nih.gov/33001032/)

©Samira M Haddad, Renato T Souza, Jose Guilherme Cecatti, Maria Barreix, Tigest Tamrat, Carolyn Footitt, Garrett L Mehl, Inraini F Syah, Anuraj H Shankar, Özge Tunçalp. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 01.10.2020. This is an open access article distributed under the terms of the Creative Commons Attribution IGO License (<https://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In any reproduction of this article there should not be any suggestion that WHO or this article endorse any specific organisation or products. The use of the WHO logo is not permitted. This notice should be preserved along with the article's original URL.

Original Paper

# Applying Digital Information Delivery to Convert Habits of Antibiotic Use in Primary Care in Germany: Mixed-Methods Study

Regina Poss-Doering<sup>1</sup>, MSc; Lukas Kuehn<sup>1</sup>, BSc; Martina Kamradt<sup>1</sup>, PhD; Katharina Glassen<sup>1</sup>, MD; Michel Wensing<sup>1</sup>, PhD

Department of General Practice and Health Services Research, University Hospital Heidelberg, Heidelberg, Germany

**Corresponding Author:**

Regina Poss-Doering, MSc

Department of General Practice and Health Services Research

University Hospital Heidelberg

Im Neuenheimer Feld 130.3

Heidelberg, 69120

Germany

Phone: 49 6221 56 ext 38643

Fax: 49 6221 56 1972

Email: [regina.poss-doering@med.uni-heidelberg.de](mailto:regina.poss-doering@med.uni-heidelberg.de)

## Abstract

**Background:** Antimicrobial resistance is an important global health issue. In Germany, the national agenda supports various interventions to convert habits of antibiotic use. In the CHANGE-3 (Converting Habits of Antibiotic Use for Respiratory Tract Infections in German Primary Care) study, digital tools were applied for information delivery: tablet computers in primary care practices, e-learning platforms for medical professionals, and a public website to promote awareness and health literacy among primary care physicians, their teams, and their patients.

**Objective:** This study is embedded in the process evaluation of the CHANGE-3 study. The aim of this study was to evaluate the acceptance and uptake of digital devices for the delivery of health-related information to enhance awareness and change habits of antibiotic use in primary care in Germany.

**Methods:** This study used a convergent-parallel mixed-methods design. Audio-recorded semistructured telephone interviews were conducted with physicians, nonphysician health professionals, and patients in the CHANGE-3 program. Pseudonymized verbatim transcripts were coded using thematic analysis. In-depth analysis was performed based on the inductive category of information provision via digital information tools. Identified themes were related to the main postulates of Diffusion of Innovations theory (DIT) to provide an explanatory frame. In addition, data generated through a structured survey with physicians and nonphysician health professionals in the program were analyzed descriptively and integrated with the qualitative data to explore the complementarity of the findings.

**Results:** Findings regarding the acceptance and uptake of digital devices were related to three postulates of DIT: innovation characteristics, communication channels, and unanticipated consequences. Participants considered the provided digital educative solutions to be supportive for promoting health literacy regarding conversion of habits of antibiotic use. However, health care professionals found it challenging to integrate these solutions into existing routines in primary care and to align them with their professional values. Low technology affinity was a major barrier to the use of digital information in primary care. Patients welcomed the general idea of introducing health-related information in digital formats; however, they expressed concerns about device-related hygiene and the appropriateness of the digital tools for older patients.

**Conclusions:** Patients and medical professionals in German primary care are reluctant to use digital devices for information and education. Using a Diffusion of Innovations approach can support assessment of existing barriers and provide information about setting-specific preconditions that are necessary for future tailoring of implementation strategies.

**Trial Registration:** International Standard Randomized Controlled Trial Number (ISRCTN) 15061174; <http://www.isrctn.com/ISRCTN15061174>.

(*J Med Internet Res* 2020;22(10):e18200) doi:[10.2196/18200](https://doi.org/10.2196/18200)

**KEYWORDS**

antimicrobial resistance; educative digital solutions; health literacy; diffusion of innovations

## Introduction

### Background

Inappropriate use and prescribing of antibiotics are among the key contributing factors to antimicrobial resistance, which remains an important global health issue. Even after decades of scientific research and a range of improvement programs on the rational use of antibiotics, substantial room for improvement remains in promoting awareness and inducing necessary changes. In recent years, health services research worldwide has focused on interventions aimed at promoting the rational use of antibiotics, usually by providing information and education to health care workers, patients, and the public [1]. Thus, an evidence base has been established regarding the effects of educational interventions targeted at physicians that use internet-based training [2], reflection on habits and behaviors and peer exchange of information [3], changes in provider-patient communication toward participatory decision making [4,5], and involvement of the complete practice team to optimize organizational processes and to ease stress and strain on individuals [6]. Previous studies showed significant effects on decreasing rehospitalization rates and increasing patient empowerment by providing health literacy via tablet computers [7-9]. A recent review concluded that educating patients through smartphone or tablet apps improves treatment adherence and clinical outcomes and has positive effects on health care economics [10]. These studies were conducted in inpatient care settings and primarily aimed to deliver information to patients with postoperative conditions; meanwhile, factors that influence the acceptance and uptake of digital information delivery for promoting awareness and health literacy regarding the rational use of antibiotics in primary care remain less well researched.

In Germany, the national agenda has reinforced policies to restrain the prescription of antibiotics [11]. As in many other countries, approximately 90% of antibiotics are prescribed in ambulatory care, mainly by general practitioners (GPs) [12] and most commonly during 41% of consultations for acute respiratory tract infections (ARTIs), of which only 52% were in accordance with guideline recommendations [13]. In this context, the CHANGE-3 study (Converting Habits of Antibiotic

Use for Respiratory Tract Infections in German Primary Care, 2017-2020) applied a set of educative intervention components, of which several were delivered in digital formats: tablet computers in primary care practices, e-learning platforms for medical professionals, and a public website to promote awareness and health literacy among primary care physicians, their teams, and their patients (see [Multimedia Appendix 1](#) for screenshots). A process evaluation carried out alongside the trial investigated the uptake and diffusion of these interventions, which were novel to the targeted physicians and practices. [Figures 1](#) and [2](#) show examples of the topics covered on the study-specific website.

The diffusion of many innovations starts slowly and then accelerates before slowing again [14]. Diffusion of Innovations theory (DIT) describes the process through which innovations diffuse and become adopted through social networks in populations [14,15]. DIT can support research and program development in assessing and understanding processes that lead to the desired translation of new ideas and technologies into widespread practice [16]. This classic middle-range theory offers concepts and approaches that can explain receptivity of health care practices and policies by individuals and organizations [17]. At the same time, the diffusion approach may help connect “research-based innovations with their potential users in a knowledge-utilization process [14].”

To evaluate the uptake and acceptance of implemented digital information delivery solutions used in the CHANGE-3 study by physicians, care teams, and patients in primary care in Germany, this study integrated an approach based on DIT that addresses three major diffusion postulates: (1) specific characteristics of innovations influence their diffusion (Relative advantage, Complexity, Compatibility, Trialability, Observability), (2) communication channels and key individuals embedded in social networks play an important role, and (3) consequences of the uptake of innovations can be unanticipated. The study primarily focused on the digital information delivery components, as these comprise a new approach to promote awareness and health literacy in primary care in Germany with regard to converting habits of antibiotic use for ARTI.

**Figure 1.** Screenshot of the study-specific website in German (“Fewer antibiotics...more health literacy”).



**Figure 2.** Screenshot of the study-specific website in German (“Fewer antibiotics...more hygiene”).



## Objectives

The aim of this mixed-methods study was to assess the acceptance and uptake of digital interventions for education and information delivery to patients, physicians, and care teams with regard to converting habits of antibiotic use for noncomplicated ARTIs in primary care.

## Methods

### Study Design

This study is part of the process evaluation of the CHANGE-3 study, which was embedded in a two-armed cluster-randomized controlled trial with 57 practices randomized into each group (N=114). Randomization for the trial was stratified by rate of antibiotics prescriptions for ARTIs at baseline per practice and was performed by the Institute of Medical Biometrics and Informatics at the University Hospital Heidelberg [18]. Within the process evaluation and for this study, a mix of quantitative and qualitative methods was used in a convergent-parallel design [19–21] to gain insights into the working mechanisms of the intervention program, which aimed to convert habits of antibiotic use and strengthen health literacy competencies. Complementary interview guides and survey questionnaires were developed and used. Mixed-methods research combines qualitative and quantitative research elements with the aim of expanding and strengthening the conclusion and validity of a study [22]. By applying this design, a broad thematic spectrum could be analyzed to assess and understand factors relevant to the acceptance of digital information delivery with regard to converting habits of antibiotic use in primary care in Germany. The intervention components focused in this study comprised tablet devices with educational contents for patients, an e-learning platform for medical professionals which offered a communication skills training, and the study-specific website on rational usage of antibiotics for both groups (see [Multimedia Appendix 1](#)). The content provided via the tablet presented guideline-based educational videos and information and was adapted from the study-specific website [18,23] and from a similar application developed for a different study [24].

### Context

The comprehensive implementation program composed for the CHANGE-3 study (ISRCTN 15061174) was based on published research and experience in quality improvement programs [25,26], and it was tested in primary care settings. The overall aim of the CHANGE-3 study was to sustainably hinder the progress of antimicrobial resistance by promoting the conversion of antibiotic use habits and to increase health literacy in practice teams, patients, and the general public. To achieve this, a regional public campaign with multimedia approaches was used to encourage the general public's engagement with rational antibiotic use and their active participation in a participatory decision-making process. In addition, an educational practice team intervention targeted internal process optimization in practices using educational intervention components, including digital information devices, individual feedback on prescribing habits, and outreach visits. The process evaluation conducted alongside the implementation program was embedded in the cluster-randomized trial, and we were particularly interested in

documenting the uptake of interventions and understanding the mechanisms and impacts of the intervention components. The study was approved by the Ethics Committee of the University of Heidelberg (reference number S-349/2018). Participants in the process evaluation all gave written informed consent. Confidentiality and anonymity were ensured throughout the study. The planned outcome evaluation, which is not part of the process evaluation, will provide information about effects of the interventions on the prescribing of antibiotics for patients with respiratory symptoms. As the CHANGE-3 study only recently concluded, analyses referring to the outcome evaluation have not yet been completed.

### Survey

After the start of the intervention, all GPs (n=132) and a sample of nonphysician health professionals (n=208), comparable to medical assistants (MAs) in the United States [27], were invited to engage in a survey (T1) in May 2019. The participation of the MAs was restricted to a maximum of 2 per practice with the intention to limit imbalance in the sample and to ensure that the reimbursement budget would not be overdrawn. To be eligible for inclusion, the GPs and MAs were required to be participants in the intervention or control group in the CHANGE-3 study and to have mastered the German language. Based on constructs of the Theoretical Domains Framework (TDF) [28], a study-specific questionnaire was developed and used. It included tailored items to facilitate investigation and understanding of the mechanisms and impact of the intervention components as well as the contextual factors. No patients were invited to participate in this survey, as it focused on the practice teams' perspectives on the applied intervention components. The questionnaire also included items referring to socio-demographic aspects and characteristics of the work environment. Email reminders were sent out after 4 weeks to increase the response rate. In March 2020, all participants who had returned T1 questionnaires were invited to participate in the follow-up survey (T2). No reminder was sent out.

### Interviews

Open-ended, semistructured, guide-based telephone interviews with the GPs and MAs participating in the CHANGE-3 study and a sample of their patients were conducted to explore their perspectives on the applied intervention components in general and the digital information delivery in particular. The interprofessional team of researchers (Health Services Research, Public Health, General Practice) developed study-specific interview guides for the three groups of interviewees (see [Multimedia Appendices 2–4](#) for translated versions). Interview guides were based on constructs of the TDF [28], a literature review, and predefined research questions. All recruits were required to be at least 18 years of age, legally fully competent, and in fluent command of German. The potential recruits considered were GPs and MAs who were participating in the CHANGE-3 intervention group and were working in a primary care practice in the German federal state of Baden-Wuerttemberg or Mecklenburg–Western Pomerania. Invitations to participate in an interview were sent out via the aQua Institute, Goettingen, Germany. Patients who sought treatment for an ARTI during the intervention period at one of the intervention or control



group practices were also eligible to participate in an interview. Using an opt-in approach, selected participating practices could support and initiate patient recruitment by addressing eligible patients. Support material was provided to care teams in these practices and facilitated a structured patient recruitment process.

All interested parties meeting the inclusion criteria received printed material as well as a telephone call from the research team at the Department of General Practice and Health Services Research, University Hospital Heidelberg, Germany, to provide further information. Participants were required to return a signed letter of intent to be included in the process evaluation and participate in the interview. In addition, the interviewees completed a one-time sociodemographic survey. All interviewees and care teams who supported patient recruitment received a small reimbursement fee. The first interview in each group served as a pilot. After that, minor adjustments were included where considered appropriate by the research team. No targeted sample size was set for the interviews, and data were collected until saturation of information was reached.

### Data Collection and Analysis

All T1 and T2 survey questionnaires returned to the Department of General Practice and Health Services Research at the University Hospital Heidelberg by June 2019 and April 20, 2020, respectively, were registered and pseudonymized by an experienced study nurse (T1) and the research team (T2). Support staff recorded all questionnaire data electronically in SPSS version 25 (IBM Corporation). Data were checked for plausibility by two study team members (RPD and LK), and typos were corrected where applicable. Subsequently, the data were analyzed descriptively in SPSS by the same two researchers. For this study, survey data referencing digital information delivery and sociodemographic characteristics were extracted from sets of items of both survey questionnaires (T1, T2) and included for analysis (see [Multimedia Appendices 5](#) and [6](#) for the translated questionnaire items). Survey data were interpreted within the frame of applicable DIT postulates.

Applying a purposive strategy with regard to equal distribution of sex and region, all GPs and MAs in the intervention group were invited to participate in an interview. A sample of 39 interview participants was recruited by the CHANGE-3 study team at the Department of General Practice and Health Services Research, University Hospital Heidelberg, between November 2018 and April 2019. This strategy facilitated identification of individuals who were especially experienced with regard to the phenomenon of interest and supported a detailed understanding

of key themes and relations to specific behaviors and roles [29]. All interviews were conducted via telephone by two members of the research team at the Department of General Practice and Health Services Research, University Hospital Heidelberg. Both researchers had profound experience in qualitative interviewing. All interviews were audio-recorded, pseudonymized, and transcribed verbatim. Data were managed with MAXQDA 2018.2 qualitative data management software (Verbi Software). The qualitative data were first thematically analyzed in a framework analysis based on constructs of the TDF and then combined with an inductive de novo approach according to concepts emerging from the data. To provide an explanatory model, and because most statements appeared to relate to social processes affecting diffusion, the identified key themes were subsequently categorized in accordance with three mirrored DIT postulates: (1) innovation characteristics, (2) communication channels, and (3) unanticipated consequences. All qualitative data generated from the interviews with the GPs, MAs, and patients, the field notes, and the sociodemographic survey were included for analysis. Quantitative and qualitative data were first analyzed separately and then brought together to complement each other.

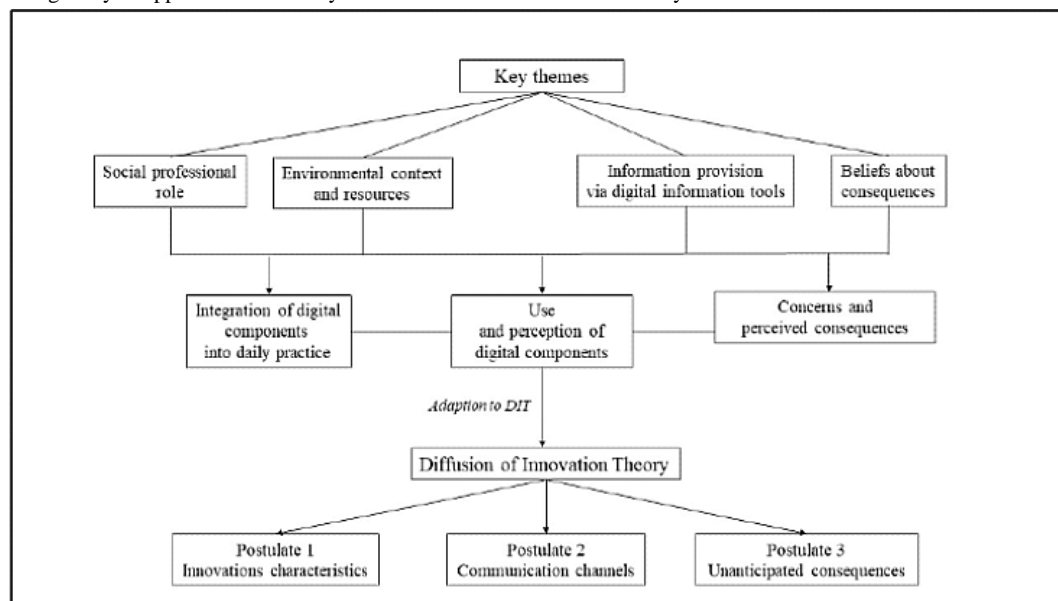
## Results

### Principal Findings

Findings derived from both quantitative and qualitative data are presented with a focus on digital information provision via the intervention components “tablet,” “e-learning platform,” and “website” and are categorized in accordance with selected postulates derived from DIT [13,14]. The postulates relate not only to relevant contextual factors expected to be identified from the survey data but also to perceptions regarding the professional role, beliefs about the consequences of the innovation, and emotions, as derived from the qualitative data.

The theorizing analytical approach of the analysis is shown in [Figure 3](#).

Key themes derived from the TDF provided the framework for the survey questionnaires and interview guides. Relevant contextual factors were identified from the survey data and qualitative data. During the qualitative analysis, the inductive domain “information provision via digital information tools” was added. All data were categorized according to the key themes and subsequently mirrored in the three DIT postulates where applicable to facilitate an explanatory framework for the findings.

**Figure 3.** Theorizing analysis approach of the study. DIT: Diffusion of Innovations theory.

### Sociodemographic Characteristics

A total of 80 GPs and 105 MAs returned the T1 questionnaire (N=185), and 63 GPs and 64 MAs returned the T2 questionnaire (N=127). In the intervention group, the mean age of the GPs was 53.2 years (SD 9.13) in T1 and 54.0 years in T2 (SD 9.70); in the control group, the mean age was 52.7 years (SD 9.6) in T1 and 56.5 years in T2 (SD 8.6). The mean age of the MAs was 40.9 years (SD 11.75) in both T1 and T2 in the intervention group; in the control group, the mean age was 42.0 years (SD 12.1) in T1 and 40.7 years in T2 (SD 11.4). The majority of the participating GPs in T1 were male (24/41, 59.5%, in the

intervention group and 25/39, 64%, in the control group), and the MAs were almost exclusively female (50/50, 100%, in the intervention group and 54/55, 98% in the control group). For T1, 28/41 (68%) of the GPs in the intervention group and 34/39 (82%) in the control group were reported to have implemented changes in their practice routines in the last two years. Additionally, 15/41 (37%) of GPs in the intervention group and 19/39 (49%) in the control group stated that they had participated in previous antibiotics studies. [Table 1](#) describes the characteristics of the participants in the intervention group and control group.

**Table 1.** Sociodemographic characteristics of the survey participants (T1) in this study (N=185).

Characteristic	GPs <sup>a</sup> (n=80)		MAs <sup>b</sup> (n=105)	
	Intervention group (n=41)	Control group (n=39)	Intervention group (n=50)	Control group (n=55)
Age, mean (SD)	53.2 (9.13)	52.7 (9.6)	40.9 (11.75)	42.0 (12.1)
Female sex, n (%)	17 (42.0)	14 (36.0)	50 (100.0)	54 (98.0)
Experience (years), mean (SD)	24.7 (9.6)	23.3 (9.2)	16.5 (11)	17.2 (11.2)
Implemented changes in the last 2 years, n (%)	28 (68)	34 (82)	34 (75)	38 (70)
Participated in another project to improve antibiotic prescribing, n (%)	15 (37)	19 (49)	14 (28)	32 (62)

<sup>a</sup>GPs: general practitioners.

<sup>b</sup>MAs: medical assistants.

[Table 2](#) describes the sociodemographic characteristics of the interview participants (N=39). The mean age of the GPs was 53 years (SD 8.29), with a mean of 24 years of expertise (SD 8.2). Of the 16 interviewed physicians, 13 (81%) had a background in general practice, 2 (13%) had a background in internal medicine, and 1 (6%) was still undergoing specialty training. All the MAs were female (7/7, 100%). The mean age of the MAs was 48 years (SD 11.8), with a mean of 22 years of

expertise (SD 11.8). The average years of employment in the GP practice in which they currently worked was 16.7 (SD 6.1). The mean size of the practice teams was 3 colleagues, GPs excluded. The mean age of the interviewed patients was 36 years (SD 12.2). On average, patients had been consulting their GP for eight years (SD 8). Of the 16 patients, 9 (56%) had ten years of school education.

**Table 2.** Sociodemographic characteristics of the interview participants in this study (N=39).

Characteristic	GPs <sup>a</sup> (n=16)	MAs <sup>b</sup> (n=7)	Patients (n=16)
Age, mean (SD)	53 (8.29)	48 (11.8)	36 (12.2)
Female sex, n (%)	9 (56)	7 (100)	10 (62.5)
Years of expertise, mean (SD)	24 (8.2)	22 (11.8)	N/A <sup>c</sup>
Years with current employer, mean (SD)	N/A	16.7 (6.1)	N/A
Years of consulting this GP, mean (SD)	N/A	N/A	8 (8)

<sup>a</sup>GPs: general practitioners.<sup>b</sup>MAs: medical assistants.<sup>c</sup>N/A: not applicable.

## Integration and Uptake of Digital Components Into Daily Practice

### Survey Data

In the intervention group, T1 survey questionnaires were completed and returned by 41 GPs and 50 MAs. In the control group, 39 GPs and 55 MAs returned completed questionnaires. In total, 80/132 (60.6%) of GPs and 105/208 (50.5%) of MAs responded.

The uptake of the digital interventions was mixed and was overall limited. In the intervention group, 33/41 GPs (81%) and 36/50 MAs (72%) reported that they had not used the e-learning platform. Of the GPs who had used it, 7/41 (17%) felt motivated to prescribe antibiotics according to clinical guidelines. It was found that 9/41 GPs (22%) and 16/50 MAs (32%) found the tablet to be a helpful device in addressing patients' expectations; meanwhile, 20/41 GPs (49%) and 19/50 MAs (38%) reported that they did not use the supplied tablets at all. When the tablets were offered to patients, 5/41 GPs (12%) and 13/50 MAs (26%) received the impression that patients actually used them and

browsed the provided information. Of the 41 GPs, the study-specific website motivated 15 (37%) to engage in guideline-oriented antibiotic prescribing, helped 19 (46%) address patient expectations, and supported 16 (39%) in their therapy decisions. Some GPs also reported that using the website influenced therapy decisions (12/41, 29%) and led to a reduction in prescriptions (11/41, 27%).

The T2 survey questionnaires were returned by 32 GPs and 32 MAs in the intervention group and by 31 GPs and 32 MAs in the control group. In total, 63/80 (79%) of eligible GPs and 64/105 (61%) of eligible MAs responded. With regard to the e-learning platform, 10/32 GPs (31%) and 2/32 MAs (6%) reported using it. The website was visited by 14/32 GPs (44%) and 13/32 MAs (41%). Tablet devices were used by 15/32 GPs (47%) and 20/32 MAs (63%). Regarding the uptake of the devices, the T2 data confirm the T1 findings.

**Table 3** provides information on the uptake of the three digital components as derived from the survey data. **Table 4** and **Table 5** provide information about the participants' subjective assessment of the intervention components.

**Table 3.** Uptake of digital components in the intervention group of the study, n (%).

Digital device	T1 survey		T2 survey	
	GPs <sup>a</sup> (n=41)	MAs <sup>b</sup> (n=50)	GPs (n=32)	MAs (n=32)
Visited website	21 (51)	22 (44)	14 (44)	13 (41)
Used e-learning	8 (20)	14 (28)	10 (31)	2 (6)
Used tablet	21 (51)	31 (62)	15 (47)	20 (63)

<sup>a</sup>GPs: general practitioners.<sup>b</sup>MAs: medical assistants.**Table 4.** Perceptions of the intervention component characteristics by the general practitioners, n (%).

Digital device	T1 survey (n=41)	T2 survey (n=32)
Website provides impulses for new behaviors, n (%)	N/A <sup>a</sup>	17 (56)
E-learning platform provides benefits in addressing patients' expectations, n (%)	7 (17)	11 (34)
Tablet provides benefits in addressing patients' expectations, n (%)	9 (22)	5 (16)

<sup>a</sup>N/A: not available.

**Table 5.** Perceptions of the intervention component characteristics by the medical assistants, n (%).

Digital device	T1 survey (n=50)	T2 survey (n=32)
Website is supportive for patient communication, n (%)	22 (44)	19 (59)
Website motivates me to support the GP <sup>a</sup> more intensely, n (%)	17 (34)	19 (59)
E-learning platform motivates me to support the GP in treating ARTI <sup>b</sup> infections, n (%)	11 (22)	6 (19)
Tablet is supportive of my daily routine, n (%)	18 (36)	16 (50)

<sup>a</sup>GP: general practitioner.

<sup>b</sup>ARTI: acute respiratory tract infection.

### Interview Data

In the qualitative study conducted as part of the process evaluation of the CHANGE-3 study, a total of 39 telephone-based interviews were carried out with GPs (n=16), MAs (n=7) and patients (n=16). The average interview duration was 22.1 minutes (SD 6.21). Selected quotes supporting key statements are provided below with indications of the participant group and transcript position. Additional quotes are provided in Appendix 7.

### Uptake of Digital Components in Daily Practice

During the analysis of the interview data and the identified key themes, reluctance toward integrating the tablet intervention component into daily practice became apparent. In relation to DIT Postulate 1, “innovation characteristics,” the following section links this reluctance to use tablet devices with the five innovation characteristics from DIT.

The GPs raised concerns regarding the *compatibility* of their professional values with the digital devices. Of the 16 GPs interviewed, 6 (38%) articulated a conscious decision not to use the digital information tools. They saw their medical practice as a place of tranquility, where patients have the opportunity to calm down during times when they are constantly flooded with information. Moreover, 4/16 GPs (25%) raised additional concerns regarding the usage of the tablets, as they perceived themselves as the prior channel of information for patients and strongly saw themselves in the role of an adviser. They assumed that even younger patients would prefer advice from human beings, which may reflect an unclear *relative advantage* of the tablets. The MAs shared these opinions; however, they also addressed aspects referring to the quality of generally available information that is obtainable via digital sources. The patients suggested that using a television screen could be more appropriate for waiting areas. A subjectively perceived want of *observability* appeared to also foster reluctance toward using the devices. Referring to the required comprehension skills, one GP raised *complexity* concerns of potentially excluding older patients by using tablets; this GP mentioned that with the aim of equal treatment, tablets should not be used regularly for information delivery in primary care until a complete generation change has been accomplished. *Trialability* considerations did not seem to be relevant. In summary, the GPs expressed that the tablets did not play a major role in their approach to patient-centered information delivery.

*I think people already are getting bombarded enough with this stuff. I don't want that in the practice here. We actually also have a ban on mobile phones here. So, of course you can play around on your mobile phone, but you must not talk on the phone [...] I believe that this is also quite good for the patients if they come to rest for a few minutes in the waiting area. In this respect, I don't see a place for it in my practice now. [GP 14, #64]*

*And I also don't want my older patients in particular to have the impression that “This is something that I can't go along with anymore[...].” [GP 09, #21]*

*In my opinion, we already have enough gadgets like computer, tablets and so forth where patients can inform themselves, and some of them show up here with preconceived opinions about their conditions and how to treat them. You have to be a little careful there. [MA 03, #46]*

*I believe, it rather makes sense to offer a TV or something bigger where information is permanently running, because otherwise, just one person at a time is able to use it [tablet], which does not make a lot of sense then, or a constantly changing display of information everybody can follow would make more sense. A tablet doesn't change much, I guess. [Patient 02, #32]*

### Use and Perception of Digital Intervention Components

Although the provision of the digital information tools was communicated to all participating GPs and MAs, their recollection of this was limited. The following section reflects statements from interviewed GPs and MAs who were aware of the presence of the digital information tools. We also include hypothetical statements from participants who did not encounter the intervention components. All these statements provide insight into the beliefs of the study participants about the appraisal of digital devices and the dissemination of information about these devices.

Awareness of the provided e-learning module was limited. Only 2 of the 16 interviewed GPs (13%) described in-depth use of the provided e-learning platform and considered it to be helpful in counselling situations. The communicative elements were seen to be helpful in dealing with patients' potentially unwarranted expectations. Only one GP categorized the illustrated situations as unrealistic; however, they still considered



them to be helpful to highlight specific situations. The MAs did not describe using the e-learning module at all.

*Yes, very staged scene, but to make it clear to you, I thought it was good. [GP 06, #42]*

*Ah yes, so you just say, let's say, not affront the patient somehow and so you just indirectly say that people who have certain expectations or who want to think an antibiotic now would just lead faster to their goal to get well again, so these communication-related aspects were very helpful. [GP 02, #18]*

The website was visited by only a few of the interviewees. Time constraints and a lack of awareness of the existence of the website were mentioned as the main reasons for not using it, pointing to potentially insufficient introduction and communication. However, the design of the website was described as visually appealing, yet overloaded with information. All the participants considered an educational website providing clear and relevant health-related information to be an asset; however, they assumed that age-related preferences were in place.

*Yes, I am aware of it, but I did not visit it [website] yet. [MA 01, #78]*

*No, somehow that [website] went right past me. [GP 18, #54]*

*You could click on different topics, short and concise, attractively constructed, but I would have arranged it in a little less cluttered way, less information then. Yes, but overall it is attractively interesting, you readily want to engage with it. Well, "Oh, what's this," have a brief look. [GP 09, #9]*

Patients who had not yet had the opportunity to use the tablets were asked how they would envision provision of health-related information to patients in general practices. Of the 16 patients, 7 (44%) named tablet applications as potential information tools and demonstrated an open-minded attitude toward the digital provision of health-related information. Four of the 16 patients (25%) did not share their opinion, and 6 (38%) demonstrated a somewhat hesitant attitude toward digital devices. One patient saw the opportunity to use the tablet as a tool to prospectively retrieve information about the recommended usage of antibiotics. Regarding a potential lack of knowledge, this would provide the opportunity to avoid uncomfortable situations during consultation with the GP, potentially shorten the consultation, and allow more time to address unclear aspects and remaining questions. In contrast to this, another patient stated having no interest in health information provided by a tablet solution when having a cold and feeling sick. Furthermore, 5/16 patients (31%) considered televisions in waiting areas of medical practices to be a suitable alternative information tool. The website was considered to be an option for younger adults, and suggestions were made as to where such a site could be referenced.

*I am not sure. I mean, when I'm sick, I am really sick and when I'm sitting there, I personally have no interest in reading information on tablets. [Patient 17, #77-78]*

*By the help of information leaflet, right. It has to be printed on them so people can find the website. That's a possibility that spontaneously popped into my head. [Patient 12, #73]*

In addition, another patient pointed out the difficulty of staying in line with the received information if a GP acts contrary to the recommendations. When in doubt about the treatment, standing up against a physician and refusing an antibiotics prescription was considered to be difficult, especially in situations where no laboratory results were available. In this scenario, to be on the reputed "safe side," the patient would prefer to go along with the GP's decision.

*Because you go to the doctor and when he says: "Ah, well antibiotics are necessary," then you say: "Yes, well, then I will take them." So, I just have to come back to the ENT especially, because he hasn't taken a smear or anything else, for example, he just determined that: Chronic sinuses: antibiotics. And yes, it didn't do anything either. [Patient 03, #42]*

GPs and MAs noticed implications for their daily routines, which they attributed to participating in the study in general and to using the digital information devices in particular. Of the 7 MAs, 4 (57%) saw an impact of using the tablets on their daily routines. The main influence was seen in communication with patients. With the help of the devices, the MAs felt it was easier to convey the importance of antibiotics-free treatment to patients. Additionally, they saw the opportunity for patients to inform themselves prospectively, which was considered to lead to a decline in the number of necessary consultations. Additionally, the MAs described a broad range of assumptions and beliefs about the tablets' influences on their own knowledge gain. This extended from the feeling of "no influence at all" to the feeling of having "increased confidence" regarding the usage of antibiotics. GPs saw changes in their communicative interactions with patients and considered their participation in the CHANGE-3 study to have led to notably more forthright debates about the relevance of antibiotics. They reported that they had started to address patients' perceptions regarding the prescription of antibiotics more directly. Patients only hypothetically contemplated the potential effects of the digital provision of health-related information in primary care practices.

*Well, you actually ask the patients more often, right? "Do you expect an antibiotic?", you also say right away "viruses are not antibiotics accessible!", you discuss that simply more offensively, right? You also assume somehow that it is clear, but to most people it is not clear at all, perhaps one is too much in one's own world and perception, right? - you simply integrate this more obviously in the conversation. [GP 01, #22]*

*So, I think definitely better than before. I think you can always learn something and I definitely feel confident and literate [enough] in this topic to be able to also give the best possible advice to the patients. [MA 05, #34]*



## Concerns and Perceived Consequences

Unanticipated concerns and consequences which had not been considered beforehand emerged from the data and appeared to have a major impact on the participants' reluctance to use digital information tools. To begin with, the influencing factors were organizational in nature and addressed daily routines in general practices; hygiene issues were mentioned as a major factor that obstructed the use of tablets. GPs as well as patients articulated concerns about using tablets in a public waiting area with regard to keeping them clean and disinfected. While patients' concerns referred to the potential risk of infection, the GPs also contemplated the proper provision of disinfection wipes next to the devices. A noticeably low technology affinity led to reduced use of the tablets, the e-learning module, and the study-specific website. The e-learning platform provided video-based situations to all GPs in the intervention group to illustrate potential difficulties in provider-patient communication and in addressing patients' expectations; however, it was only used by a small number of participants. The GPs and MAs did not extensively familiarize themselves with the study-specific website. Thus, it was referred to only on occasion, and patients were not systematically made aware of it.

*Yes, I would ask myself: "How many people touched it before me?" I wouldn't touch that honestly. Well, and I don't know how often [...] it might get disinfected once a day, but if there are five people alone with a flu virus somehow and then it's not the same strain, yes, well [...] well, I do not support it. [Patient 11, #54]*

*So, on the one hand I find it problematic, also because everyone is tapping on it and also from the hygienic aspect I should actually put disinfectant wipes next to it [...] [GP 09, #21]*

Another unanticipated aspect was the perception of intensified time constraints in the usage of tablets. The MAs felt pressured to find time to incorporate the tablets into their daily routines. For them, the intervention component represented an additional task for which they felt responsible. The GPs and MAs both contemplated whether the amount of time a patient spends in the waiting area was sufficient for proper knowledge transfer. From their perspective, the value of tablets as an information tool did not justify investing the scarce resource of time into efforts to brief patients about the proper usage of the devices.

*Well, that would tie up far too much time at the front desk, if a MA first has to explain what to do with the tablet, then she has to collect a deposit for it and then return it, disinfect it, that would tie up far too much MA working time, so we didn't use it. [GP 04, #22]*  
*[...] at our place, patients often don't have much time in the waiting area to deal with it [tablet] because they don't sit [and wait] that long [...] [MA 07, #30]*

One further unexpected organizational aspect stemmed from a general concern of the GPs and MAs about the potential theft of the tablets. One MA recalled a moment when even a painting had been stolen from a wall in their practice. Under such circumstances, a tablet, as an object of higher value, was not considered to be the right information tool to distribute to

patients in a medical practice. As a consequence, due to hygiene issues, a perception of intensified time constraints, and concerns about potential theft, the tablets were rarely used as intended or were not offered to patients at all.

*Yes, and something has been stolen before, somehow a painting even from the wall, and we just didn't want to induce this stress with this tablet now, [...] we had this conversation in the practice [...] [MA 01, #22]*

## Discussion

### Overview

The findings of this study in German primary care provide insights into the uptake of digital devices, medical professionals' beliefs about digital educational interventions, and how these may have collided with the professionals' understanding of their professional roles. The implementation met with challenges and hesitation, with the main barriers being perceived incompatibility with existing routines, low technology affinity, and concerns regarding the complexity of digital tools.

Active use of digital devices for health-related purposes appears to be more common in other high-income countries in Europe, such as Denmark and the Netherlands, where web-based platforms for patient education and e-learning for GPs have been established. Based on research evidence and prior experiences, the tablet application, the e-learning platform, and the website were implemented to strengthen health literacy, encourage guideline fidelity, and provide a subsequent change in antibiotic use habits for ARTI in German primary care. Prior studies used tablet computers to specifically disseminate health literacy in adults [30-32], for self-management programs for chronic diseases [33], diagnostics [34,35], interventions for parents and children [36-40], coping strategies after diverse surgeries [8,41], or as tracking functions for aims of weight reduction [42] and general fitness [43,44]. However, the findings of this study demonstrate a variety of challenges and a reluctance toward using digital information tools in German primary care practices for both patients and health professionals. This may be due to incompatibility of the digital devices with practice routines, patient expectations, and the perceptions of GPs and MAs about their medical practice and professional values. The GPs emphasized that they wanted to create a safe harbor for patients in a complex technological world, and they still saw themselves as the primary source of valid information for patients. As this attitude may stem from professional self-conception and may even be intensified by the age factor, future implementation programs should take these factors into account. MAs attempted to incorporate the digital information provision but considered it to be of little importance. Thus, not all study participants came into contact with the digital intervention components, which may also be explained by the choice of communication channels used in the public campaign and the practice team intervention for the distribution of the innovation. Patients considered the provision of health-related information via a trustworthy website to be a good option for younger patients and named options for creating awareness of such a website. They also saw waiting area televisions, which were not offered in the participating practices, as a better

alternative to tablet devices. Adopters and nonadopters of the innovations voiced concerns about the applicability and adequacy of the digital components. In the next section, the findings will be discussed in relation to the selected DIT postulates to provide potential explanations for the findings.

### Postulate 1: Innovation Characteristics

Participants in this study found it challenging to implement the use of tablets into their organizational routines in primary care. The perceived lack of compatibility of this new intervention component with existing routines as well as the simple and convenient option to withdraw them appeared to be a major factor contributing to the absence of success. More importantly, a limited technology affinity and willingness to use digital solutions in the practice setting remained a major barrier. A qualitative study by Lyles et al [45] investigated tablets as a waiting room tool, where they were used as visit planners that focused on specific patient needs and were perceived as a “safe place” by patients to bring up sensitive topics. This supports our findings regarding the opportunity to avoid uncomfortable situations with GPs during consultations. Therefore, to disperse compatibility concerns, it may be beneficial to highlight the potential of tablet devices to maintain and support the desired “safe harbor” for patients.

Complexity concerns were also raised regarding the confrontation of older patients with modern technical devices. This is in line with the findings of a mixed-methods study by Patel et al [46] in which interviewed health care providers raised concerns about confronting older people with health literacy devices based on digital platforms. They surveyed 84 patients in a community health center waiting area in Massachusetts and identified a high level of interest in tablet-based health literacy solutions. Likewise, Stribling and Richardson [47] used tablets to provide educative health-related information in clinical waiting areas and surveyed patients afterward regarding their satisfaction with this method of digital information delivery. On average, the patients were satisfied with the usability and the educative input via the devices. However, in our study, complexity concerns for older people were primarily mentioned by health care providers or younger patients, not by the older people themselves. Future research should investigate perceptions of older patients more closely to evaluate if digital solutions are a reasonable intervention to strengthen their health literacy competencies.

The general possibility to observe the trialability effects of the new intervention [14] was not postulated sufficiently. It was assumed that the GPs would recognize the potential influence of the intervention on the health literacy competencies of patients. However, in light of known time constraints in primary care, it can be assumed that the GPs did not extensively evaluate the potential effects on patient health literacy and were not sufficiently aware of the inherent relative advantages. This leads to the assumption that there was no clear understanding of how the digital solutions could have been integrated into daily care routines and of the type of added value that could be expected. Instead, perceptions about innovation characteristics were more dominant, leading to a preference for more familiar and traditional solutions. Prior studies also used multimedia

approaches to provide information in a digitally enhanced manner and offered printed information, videos, or individualized risk assessments [31,32]. In contrast to our findings, those studies showed significant effects of the use of tablets as educational devices for patients. It can be assumed that in the absence of significant impediments, broader adoption and use of the technology-mediated information delivery approaches could have been possible.

### Postulate 2: Communication Channels

The value of a new innovation must be defined and communicated. With Postulate 2, DIT suggests promoting new interventions with the help of personal approaches and the use of gatekeepers and opinion leaders. Rogers [14] stated that people make decisions based not only on rational considerations, but even more on their personal beliefs. Thus, in the primary care setting, it could have been more effective to demonstrate the value of the digital information tools using peer support instead of providing informative material that was based on rational facts alone. Also, peer patient coaching efforts could have been considered, as a lack of knowledge about antibiotic resistance may affect patients' ability to understand the importance of the topic. Klingenberg et al [48] identified that less than 60% of patients in German primary care practices were aware of the possibility of being infected by resistant bacteria or the fact that antibiotics have no effects on viruses. To sufficiently transport knowledge and demonstrate the effectiveness and value of digital devices in achieving this transport, a more personal information delivery approach appears to be a relevant choice. However, with the exception of the offered outreach visits, such approaches remained largely unconsidered in the CHANGE-3 study, and gatekeepers were of no relevance.

### Postulate 3: Unanticipated Outcomes

The analysis of the qualitative data facilitated the identification of a number of unanticipated outcomes. Time constraints, insecurities, unclear procedures, hygienical aspects, and maintenance responsibilities were identified as barriers to sustainable application of the tablets in daily care. Reduced technology affinity greatly influenced the will and ability to use the digital information devices, the e-learning module, and the study-specific website. The e-learning platform was provided to all GPs in the intervention group. It consisted of video-based situations illustrating potential communicative difficulties between GPs and patients. The main focus was on addressing patients' expectations, especially if they wished to receive antibiotic-based treatment without indication. Although this situation is often claimed to occur frequently, the e-learning module was only used by a small number of participants.

Schreiweis et al [49] conducted a systematic review in which they aimed to identify barriers to and facilitators of the implementation of e-health devices. The main barriers were found to be concerns about theft, absence of motivation, added workload, and general issues in adopting eHealth devices into organizational routines. Facilitators were found in the ease of use, staff motivation, involvement of all stakeholders, and the availability of resources. The findings in our study seem to match these results.

The reduction of antibiotic resistance is a goal that is preventive in nature. Noncompliant usage or prescription of antibiotics appears to have no directly perceived influence on individuals. Thus, this subjectively perceived lack of observability can lead to slower adoption of innovations. The effects of a desired behavior change become visibly lagged in time, which increases the difficulty of achieving satisfactory adoption of new intervention components with preventive characteristics. Therefore, the relative advantage of preventive innovations must be prioritized in communication [16]. Under such circumstances, it is even more important to point out the relevance of new and, here, digital intervention components, as awareness and responsibility help ensure a desired rate of implementation [50]. However, although the adoption of digital information devices was slow in this study, our findings show that the GPs and MAs noticed changes in communication and interaction with patients, and they attributed these changes to participation in the study and the use of its components.

Rogers [14] defined five adopter categories (innovators, early adopters, early majority, late majority, and laggards) that describe types of individuals and their willingness to adopt innovations. The criterion for building these categories was the degree of innovativeness at which an individual feels ready to implement an innovation into their professional routine. For sufficient contrast, a closer look at the characteristics of innovators and laggards is worthwhile. While innovators are defined by a more encouraged attitude toward change, laggards are driven by traditional values and suspicion. Laggards and late majority adopters require the removal of uncertainties about new ideas to decrease their skepticism and help them feel safe to adopt those ideas. According to the findings of this study, MAs and GPs who considered unanticipated consequences and compatibility concerns may have been driven by uncertainties or tradition and, thus, can potentially be categorized as laggards. According to Rogers, uncertainties should be addressed by individualized messages and involvement of opinion leaders. Because innovators are crucial to reach a critical mass and to accelerate the adoption of new ideas, it may also be a purposive approach to identify innovators and early adopters beforehand and let them act as role models in their social network of peers [14]. However, opinion leaders were not involved or identified in this study, and early adopters were not used to accelerate the adoption of the digital information devices.

This mixed-methods study aimed to assess the acceptance and impact of educative digital information delivery to patients, physicians, and care teams. The intention was to promote and facilitate a conversion of habits of antibiotic use for noncomplicated acute respiratory tract infections in German primary care. However, identified barriers impeded a broader use of digital solutions, thus pointing to a need to tailor and strengthen future implementation strategies in the field. This can minimize potential perceptions of burden on care teams and enable increased use, as potential barriers would be identified prior to implementation.

## Strengths and Limitations

The purposive sample of interview participants facilitated a detailed exploration and understanding of the central themes and underlying perceptions of the three participant groups. Structural variance in the qualitative data was ensured through age, gender, and years of working experience. Qualitative interviews are an important research tool to enable and foster understanding of perspectives of targeted groups. Telephone interviews can be seen as a valuable method of collecting information on sensitive topics [51]. All participants felt comfortable with the chosen method, as it accommodated their busy schedules. No participant opted for a face-to-face interview. Rapport was built without effort and was supported by frequent vocalized acknowledgments during the interviews. The participant-centered approach enabled consideration of verbal prompts, follow-up questions, and note-taking without distracting or influencing the interviewees. Applying this method resulted in rich data. Notably, DIT was developed before internet and digitalization efforts gained momentum in daily life and work, providing countless options for rapid dissemination of innovations. This may indicate a need to reassess the current applicability of DIT. However, applying DIT provided a useful explanatory frame for the hesitant integration of digital educational devices. Analysis of all data was guided by adequate methodological strategies aiming to minimize research bias and reduce the risk of losing relevant content. Typicality of the observations was shown by providing simple counts where their support of the findings can be expected. This also addresses potential issues of anecdotalism and exoticism. Reporting of the qualitative findings followed the recommendations of the COREQ (COConsolidated criteria for REporting Qualitative research) checklist [52].

Some limitations must be reported. Some members of the study population participated in a similar study that ran almost concurrently. This may have contributed to a stronger awareness of the topic but also to a clouded perception that impeded stronger engagement with the digital solutions, which were used by only a small number of participants. Social desirability of answers in the data cannot be excluded, although it can be considered to be less probable in light of the presented findings. Health literacy interventions have a preventive purpose, which may reduce fidelity rates. All presented findings must be interpreted with caution in terms of their representativeness.

## Conclusion

Patients and medical professionals in primary care can be reluctant to use digital devices for information and education that aim to strengthen the rational use of antibiotics in ARTIs. To encourage a higher rate of intervention fidelity, future studies should consider identified barriers and facilitators in a detailed manner, and more tailored interventions should be developed accordingly. Interventions that are in line with professionals' perceptions will be crucial for success.

## Acknowledgments

The authors would like to thank all participating interview partners of this study and the aQua Institute, which gave overall support to conducting the CHANGE-3 study and the recruitment of the participating physicians and medical assistants. We also thank the staff at the Department of General Practice and Health Services Research at the University Hospital, Heidelberg, who were involved in transcribing the qualitative data and digitalizing the survey data.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Screenshots of the tablet app, e-learning, and website.

[[DOCX File , 1743 KB](#) - [jmir\\_v22i10e18200\\_app1.docx](#) ]

### Multimedia Appendix 2

Translated interview guide for GPs.

[[DOCX File , 15 KB](#) - [jmir\\_v22i10e18200\\_app2.docx](#) ]

### Multimedia Appendix 3

Translated interview guide for MAs.

[[DOCX File , 14 KB](#) - [jmir\\_v22i10e18200\\_app3.docx](#) ]

### Multimedia Appendix 4

Translated interview guide for patients.

[[DOCX File , 14 KB](#) - [jmir\\_v22i10e18200\\_app4.docx](#) ]

### Multimedia Appendix 5

Translated survey questionnaires for GPs.

[[DOCX File , 69 KB](#) - [jmir\\_v22i10e18200\\_app5.docx](#) ]

### Multimedia Appendix 6

Translated survey manuscripts for MAs.

[[DOCX File , 73 KB](#) - [jmir\\_v22i10e18200\\_app6.docx](#) ]

### Multimedia Appendix 7

Additional quotations.

[[DOCX File , 16 KB](#) - [jmir\\_v22i10e18200\\_app7.docx](#) ]

## References

1. Kyaw BM, Tudor Car L, van Galen LS, van Agtmael MA, Costelloe CE, Ajuebor O, et al. Health Professions Digital Education on Antibiotic Management: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *J Med Internet Res* 2019 Sep 12;21(9):e14984. [doi: [10.2196/14984](#)] [Medline: [31516125](#)]
2. Little P, Stuart B, Francis N, Douglas E, Tonkin-Crine S, Anthierens S, GRACE consortium. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. *Lancet* 2013 Oct 05;382(9899):1175-1182. [doi: [10.1016/S0140-6736\(13\)60994-0](#)] [Medline: [23915885](#)]
3. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012 Jun 13(6):CD000259. [doi: [10.1002/14651858.CD000259.pub3](#)] [Medline: [22696318](#)]
4. Altiner A, Brockmann S, Sielk M, Wilm S, Wegscheider K, Abholz HH. Reducing antibiotic prescriptions for acute cough by motivating GPs to change their attitudes to communication and empowering patients: a cluster-randomized intervention study. *J Antimicrob Chemother* 2007 Sep;60(3):638-644. [doi: [10.1093/jac/dkm254](#)] [Medline: [17626023](#)]
5. Cals JWL, Butler CC, Hopstaken RM, Hood K, Dinant G. Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in lower respiratory tract infections: cluster randomised trial. *BMJ* 2009 May 05;338:b1374. [doi: [10.1136/bmj.b1374](#)] [Medline: [19416992](#)]
6. Butler CC, Simpson SA, Dunstan F, Rollnick S, Cohen D, Gillespie D, et al. Effectiveness of multifaceted educational programme to reduce antibiotic dispensing in primary care: practice based randomised controlled trial. *BMJ* 2012 Feb 02;344:d8173 [[FREE Full text](#)] [doi: [10.1136/bmj.d8173](#)] [Medline: [22302780](#)]



7. Melin M, Hägglund E, Ullman B, Persson H, Hagerman I. Effects of a Tablet Computer on Self-care, Quality of Life, and Knowledge: A Randomized Clinical Trial. *J Cardiovasc Nurs* 2018;33(4):336-343. [doi: [10.1097/JCN.0000000000000462](https://doi.org/10.1097/JCN.0000000000000462)] [Medline: [29369123](https://pubmed.ncbi.nlm.nih.gov/29369123/)]
8. Dallimore R, Asinas-Tan ML, Chan D, Hussain S, Willett C, Zainuldin R. A randomised, double-blinded clinical study on the efficacy of multimedia presentation using an iPad for patient education of postoperative hip surgery patients in a public hospital in Singapore. *Singapore Med J* 2017 Sep;58(9):562-568. [doi: [10.11622/smedj.2016084](https://doi.org/10.11622/smedj.2016084)] [Medline: [27121921](https://pubmed.ncbi.nlm.nih.gov/27121921/)]
9. Wischer JL, Oermann MH, Zadvinskis IM, Kinney KC. Effects of iPad Video Education on Patient Knowledge, Satisfaction, and Cardiac Rehabilitation Attendance. *Qual Manag Health Care* 2018;27(4):204-208. [doi: [10.1097/QMH.0000000000000185](https://doi.org/10.1097/QMH.0000000000000185)] [Medline: [30260927](https://pubmed.ncbi.nlm.nih.gov/30260927/)]
10. Timmers T, Janssen L, Kool RB, Kremer JA. Educating Patients by Providing Timely Information Using Smartphone and Tablet Apps: Systematic Review. *J Med Internet Res* 2020 Apr 13;22(4):e17342 [FREE Full text] [doi: [10.2196/17342](https://doi.org/10.2196/17342)] [Medline: [32281936](https://pubmed.ncbi.nlm.nih.gov/32281936/)]
11. DART 2020 | 2nd Interim Report 2017. The Federal Government of Germany. Berlin: German Federal Ministry of Health. Federal Ministry of Food and Agriculture. Federal Ministry of Education and Research; 2017 May. URL: [https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3\\_Downloads/D/DART\\_2020/DART2020\\_Interim\\_Report\\_2017.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/D/DART_2020/DART2020_Interim_Report_2017.pdf) [accessed 2020-09-29]
12. Bundesamt für Verbraucherschutz und Lebensmittelsicherheit, Paul-Ehrlich-Gesellschaft für Chemotherapie, Zentrum Infektiologie und Reisemedizin. Germap 2015: Antibiotika-Resistenz und -Verbrauch; Bericht über den Antibiotikaverbrauch und die Verbreitung von Antibiotikaresistenzen in der Human- und Veterinärmedizin in Deutschland. Rheinbach, Germany: Antiinfectives Intelligence; 2016:180-184.
13. Kraus EM, Pelzl S, Szecsenyi J, Laux G. Antibiotic prescribing for acute lower respiratory tract infections (LRTI) - guideline adherence in the German primary care setting: An analysis of routine data. *PLoS One* 2017;12(3):e0174584 [FREE Full text] [doi: [10.1371/journal.pone.0174584](https://doi.org/10.1371/journal.pone.0174584)] [Medline: [28350820](https://pubmed.ncbi.nlm.nih.gov/28350820/)]
14. Rogers E. Diffusion of Innovations, 5th Edition. New York, NY: Simon and Schuster; 2003.
15. Dearing JW. Applying Diffusion of Innovation Theory to Intervention Development. *Res Soc Work Pract* 2009 Sep 01;19(5):503-518 [FREE Full text] [doi: [10.1177/1049731509335569](https://doi.org/10.1177/1049731509335569)] [Medline: [20976022](https://pubmed.ncbi.nlm.nih.gov/20976022/)]
16. Murray CE. Diffusion of Innovation Theory: A Bridge for the Research - Practice Gap in Counseling. *J Couns Dev* 2009;87(1):108-116. [doi: [10.1002/j.1556-6678.2009.tb00556.x](https://doi.org/10.1002/j.1556-6678.2009.tb00556.x)]
17. Dearing JW, Cox JG. Diffusion Of Innovations Theory, Principles, And Practice. *Health Aff (Millwood)* 2018 Feb;37(2):183-190. [doi: [10.1377/hlthaff.2017.1104](https://doi.org/10.1377/hlthaff.2017.1104)] [Medline: [29401011](https://pubmed.ncbi.nlm.nih.gov/29401011/)]
18. Wollny A, Altiner A, Brand T, Garbe K, Kamradt M, Kaufmann-Kolle P, et al. Converting habits of antibiotic use for respiratory tract infections in German primary care - study protocol of the cluster-randomized controlled CHANGE-3 trial. *Trials* 2019 Feb 06;20(1):103 [FREE Full text] [doi: [10.1186/s13063-019-3209-7](https://doi.org/10.1186/s13063-019-3209-7)] [Medline: [30728043](https://pubmed.ncbi.nlm.nih.gov/30728043/)]
19. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research, Third Edition. Thousand Oaks, CA: Sage Publications; 2017.
20. Tashakkori A, Teddlie C. Thousand Oaks, California, United States of America: Sage Publications; 2010.
21. Shorten A, Smith J. Mixed methods research: expanding the evidence base. *Evid Based Nurs* 2017 Jul;20(3):74-75. [doi: [10.1136/eb-2017-102699](https://doi.org/10.1136/eb-2017-102699)] [Medline: [28615184](https://pubmed.ncbi.nlm.nih.gov/28615184/)]
22. Schoonenboom J, Johnson RB. How to Construct a Mixed Methods Research Design. *Kolner Z Soz Sozpsychol* 2017;69(Suppl 2):107-131 [FREE Full text] [doi: [10.1007/s11577-017-0454-1](https://doi.org/10.1007/s11577-017-0454-1)] [Medline: [28989188](https://pubmed.ncbi.nlm.nih.gov/28989188/)]
23. CHANGE-3. weniger Antibiotika. 2020. URL: <https://www.weniger-antibiotika.de/> [accessed 2020-04-28]
24. Patienten-Info. ARENA. 2020. URL: <https://www.arena-info.de/patienten-info/> [accessed 2020-04-28]
25. Anthierens S, Tonkin-Crine S, Cals JW, Coenen S, Yardley L, Brookes-Howell L, GRACE/CHAMP INTRO team. Clinicians' views and experiences of interventions to enhance the quality of antibiotic prescribing for acute respiratory tract infections. *J Gen Intern Med* 2015 Apr;30(4):408-416 [FREE Full text] [doi: [10.1007/s11606-014-3076-6](https://doi.org/10.1007/s11606-014-3076-6)] [Medline: [25373834](https://pubmed.ncbi.nlm.nih.gov/25373834/)]
26. Meeker D, Linder JA, Fox CR, Friedberg MW, Persell SD, Goldstein NJ, et al. Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices: A Randomized Clinical Trial. *JAMA* 2016 Feb 09;315(6):562-570 [FREE Full text] [doi: [10.1001/jama.2016.0275](https://doi.org/10.1001/jama.2016.0275)] [Medline: [26864410](https://pubmed.ncbi.nlm.nih.gov/26864410/)]
27. Freund T, Everett C, Griffiths P, Hudon C, Naccarella L, Laurant M. Skill mix, roles and remuneration in the primary care workforce: who are the healthcare professionals in the primary care teams across the world? *Int J Nurs Stud* 2015 Mar;52(3):727-743 [FREE Full text] [doi: [10.1016/j.ijnurstu.2014.11.014](https://doi.org/10.1016/j.ijnurstu.2014.11.014)] [Medline: [25577306](https://pubmed.ncbi.nlm.nih.gov/25577306/)]
28. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implement Sci* 2017 Jun 21;12(1):77 [FREE Full text] [doi: [10.1186/s13012-017-0605-9](https://doi.org/10.1186/s13012-017-0605-9)] [Medline: [28637486](https://pubmed.ncbi.nlm.nih.gov/28637486/)]
29. Ritchie J, Lewis J. Qualitative Research Practice: A Guide for Social Science Students and Researchers. London, UK: Sage Publications; Feb 18, 2003.
30. Tan SL, Whittal A, Lippke S. Testing a Photo Story Intervention in Paper Versus Electronic Tablet Format Compared to a Traditional Brochure Among Older Adults in Germany: Randomized Controlled Trial. *JMIR Aging* 2018 Dec 06;1(2):e12145 [FREE Full text] [doi: [10.2196/12145](https://doi.org/10.2196/12145)] [Medline: [31518254](https://pubmed.ncbi.nlm.nih.gov/31518254/)]



31. Heinrich K, Sanchez K, Hui C, Talabi K, Perry M, Qin H, et al. Impact of an electronic medium delivery of warfarin education in a low income, minority outpatient population: a pilot intervention study. *BMC Public Health* 2019 Aug 05;19(1):1050 [FREE Full text] [doi: [10.1186/s12889-019-7370-4](https://doi.org/10.1186/s12889-019-7370-4)] [Medline: [31382942](https://pubmed.ncbi.nlm.nih.gov/31382942/)]
32. Robinson JK, Friedewald JJ, Desai A, Gordon EJ. Response Across the Health-Literacy Spectrum of Kidney Transplant Recipients to a Sun-Protection Education Program Delivered on Tablet Computers: Randomized Controlled Trial. *JMIR Cancer* 2015 Aug 18;1(2):e8 [FREE Full text] [doi: [10.2196/cancer.4787](https://doi.org/10.2196/cancer.4787)] [Medline: [28410176](https://pubmed.ncbi.nlm.nih.gov/28410176/)]
33. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. *J Med Internet Res* 2016 May 16;18(5):e97 [FREE Full text] [doi: [10.2196/jmir.4883](https://doi.org/10.2196/jmir.4883)] [Medline: [27185295](https://pubmed.ncbi.nlm.nih.gov/27185295/)]
34. Fujita K, Watanabe T, Kuroiwa T, Sasaki T, Nimura A, Sugiura Y. A Tablet-Based App for Carpal Tunnel Syndrome Screening: Diagnostic Case-Control Study. *JMIR mHealth uHealth* 2019 Sep 13;7(9):e14172 [FREE Full text] [doi: [10.2196/14172](https://doi.org/10.2196/14172)] [Medline: [31586365](https://pubmed.ncbi.nlm.nih.gov/31586365/)]
35. Yimtae K, Israsena P, Thanawirattananit P, Seesutas S, Saibua S, Kasemsiri P, et al. A Tablet-Based Mobile Hearing Screening System for Preschoolers: Design and Validation Study. *JMIR mHealth uHealth* 2018 Oct 23;6(10):e186 [FREE Full text] [doi: [10.2196/mhealth.9560](https://doi.org/10.2196/mhealth.9560)] [Medline: [30355558](https://pubmed.ncbi.nlm.nih.gov/30355558/)]
36. Khokhar D, Nowson CA, Margerison C, West M, Campbell KJ, Booth AO, et al. The Digital Education to Limit Salt in the Home Program Improved Salt-Related Knowledge, Attitudes, and Behaviors in Parents. *J Med Internet Res* 2019 Feb 25;21(2):e12234 [FREE Full text] [doi: [10.2196/12234](https://doi.org/10.2196/12234)] [Medline: [30801255](https://pubmed.ncbi.nlm.nih.gov/30801255/)]
37. Jesus LMT, Martinez J, Santos J, Hall A, Joffe V. Comparing Traditional and Tablet-Based Intervention for Children With Speech Sound Disorders: A Randomized Controlled Trial. *J Speech Lang Hear Res* 2019 Nov 22;62(11):4045-4061. [doi: [10.1044/2019\\_JSLHR-S-18-0301](https://doi.org/10.1044/2019_JSLHR-S-18-0301)] [Medline: [31644381](https://pubmed.ncbi.nlm.nih.gov/31644381/)]
38. Ruggiero KJ, Bunnell BE, Andrews Iii AR, Davidson TM, Hanson RF, Danielson CK, et al. Development and Pilot Evaluation of a Tablet-Based Application to Improve Quality of Care in Child Mental Health Treatment. *JMIR Res Protoc* 2015 Dec 30;4(4):e143 [FREE Full text] [doi: [10.2196/resprot.4416](https://doi.org/10.2196/resprot.4416)] [Medline: [26717906](https://pubmed.ncbi.nlm.nih.gov/26717906/)]
39. Hammersley ML, Okely AD, Batterham MJ, Jones RA. An Internet-Based Childhood Obesity Prevention Program (Time2bHealthy) for Parents of Preschool-Aged Children: Randomized Controlled Trial. *J Med Internet Res* 2019 Feb 08;21(2):e11964 [FREE Full text] [doi: [10.2196/11964](https://doi.org/10.2196/11964)] [Medline: [30735139](https://pubmed.ncbi.nlm.nih.gov/30735139/)]
40. Shorey S, Ng YPM, Ng ED, Siew AL, Mörelius E, Yoong J, et al. Effectiveness of a Technology-Based Supportive Educational Parenting Program on Parental Outcomes (Part 1): Randomized Controlled Trial. *J Med Internet Res* 2019 Feb 13;21(2):e10816 [FREE Full text] [doi: [10.2196/10816](https://doi.org/10.2196/10816)] [Medline: [30758289](https://pubmed.ncbi.nlm.nih.gov/30758289/)]
41. den Bakker CM, Huirne JA, Schaafsma FG, de Geus C, Bonjer HJ, Anema JR. Electronic Health Program to Empower Patients in Returning to Normal Activities After Colorectal Surgical Procedures: Mixed-Methods Process Evaluation Alongside a Randomized Controlled Trial. *J Med Internet Res* 2019 Jan 29;21(1):e10674 [FREE Full text] [doi: [10.2196/10674](https://doi.org/10.2196/10674)] [Medline: [30694205](https://pubmed.ncbi.nlm.nih.gov/30694205/)]
42. Pagoto S, Tulu B, Agu E, Waring ME, Oleski JL, Jake-Schoffman DE. Using the Habit App for Weight Loss Problem Solving: Development and Feasibility Study. *JMIR mHealth uHealth* 2018 Jun 20;6(6):e145 [FREE Full text] [doi: [10.2196/mhealth.9801](https://doi.org/10.2196/mhealth.9801)] [Medline: [29925496](https://pubmed.ncbi.nlm.nih.gov/29925496/)]
43. Mehra S, Visser B, Cila N, van den Helder J, Engelbert RH, Weijs PJ, et al. Supporting Older Adults in Exercising With a Tablet: A Usability Study. *JMIR Hum Factors* 2019 Feb 01;6(1):e11598 [FREE Full text] [doi: [10.2196/11598](https://doi.org/10.2196/11598)] [Medline: [30707106](https://pubmed.ncbi.nlm.nih.gov/30707106/)]
44. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. *JMIR Mhealth Uhealth* 2015 May 21;3(2):e43 [FREE Full text] [doi: [10.2196/mhealth.4003](https://doi.org/10.2196/mhealth.4003)] [Medline: [25998158](https://pubmed.ncbi.nlm.nih.gov/25998158/)]
45. Lyles CR, Altschuler A, Chawla N, Kowalski C, McQuillan D, Bayliss E, et al. User-Centered Design of a Tablet Waiting Room Tool for Complex Patients to Prioritize Discussion Topics for Primary Care Visits. *JMIR Mhealth Uhealth* 2016 Sep 14;4(3):e108 [FREE Full text] [doi: [10.2196/mhealth.6187](https://doi.org/10.2196/mhealth.6187)] [Medline: [27627965](https://pubmed.ncbi.nlm.nih.gov/27627965/)]
46. Patel V, Hale TM, Palakodeti S, Kvedar JC, Jethwani K. Prescription Tablets in the Digital Age: A Cross-Sectional Study Exploring Patient and Physician Attitudes Toward the Use of Tablets for Clinic-Based Personalized Health Care Information Exchange. *JMIR Res Protoc* 2015 Oct 19;4(4):e116 [FREE Full text] [doi: [10.2196/resprot.3806](https://doi.org/10.2196/resprot.3806)] [Medline: [26481906](https://pubmed.ncbi.nlm.nih.gov/26481906/)]
47. Stribling JC, Richardson JE. Placing wireless tablets in clinical settings for patient education. *J Med Libr Assoc* 2016 Apr;104(2):159-164 [FREE Full text] [doi: [10.3163/1536-5050.104.2.013](https://doi.org/10.3163/1536-5050.104.2.013)] [Medline: [27076806](https://pubmed.ncbi.nlm.nih.gov/27076806/)]
48. Klingenberg A, Brand T, Andres E, Kaufmann-Kolle P, Wambach V, Szecsenyi J. Was wissen Patienten über Antibiotika, und wie häufig erwarten sie deren Verordnung? *ZfA* 2019(5):198-202. [doi: [10.3238/zfa.2019.0198-0202](https://doi.org/10.3238/zfa.2019.0198-0202)]
49. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and Facilitators to the Implementation of eHealth Services: Systematic Literature Analysis. *J Med Internet Res* 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: [10.2196/14197](https://doi.org/10.2196/14197)] [Medline: [31755869](https://pubmed.ncbi.nlm.nih.gov/31755869/)]
50. Rogers E. Diffusion of Innovations, Fourth Edition. New York, NY: Simon and Schuster; Jul 06, 2010:518.
51. Cachia M, Millward L. The telephone medium and semi - structured interviews: a complementary fit. *Qual Res Organ Manag* 2011 Nov 15;6(3):265-277. [doi: [10.1108/17465641111188420](https://doi.org/10.1108/17465641111188420)]

52. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007 Dec;19(6):349-357. [doi: [10.1093/intqhc/mzm042](https://doi.org/10.1093/intqhc/mzm042)] [Medline: [17872937](https://pubmed.ncbi.nlm.nih.gov/17872937/)]

## Abbreviations

**ARTI:** acute respiratory tract infection

**CHANGE-3:** Converting Habits of Antibiotic Use for Respiratory Tract Infections in German Primary Care Facilities

**COREQ:** CONsolidated criteria for REporting Qualitative research

**DIT:** Diffusion of Innovations theory

**GP:** general practitioner

**MA:** medical assistant

**TDF:** Theoretical Domains Framework

*Edited by G Eysenbach; submitted 11.02.20; peer-reviewed by S Hewage, V Osadchiy, T Muto, M Nomali; comments to author 23.03.20; revised version received 29.04.20; accepted 11.08.20; published 07.10.20.*

*Please cite as:*

*Poss-Doering R, Kuehn L, Kamradt M, Glassen K, Wensing M*

*Applying Digital Information Delivery to Convert Habits of Antibiotic Use in Primary Care in Germany: Mixed-Methods Study*  
*J Med Internet Res* 2020;22(10):e18200

URL: <http://www.jmir.org/2020/10/e18200/>

doi: [10.2196/18200](https://doi.org/10.2196/18200)

PMID: [32960773](https://pubmed.ncbi.nlm.nih.gov/32960773/)

©Regina Poss-Doering, Lukas Kuehn, Martina Kamradt, Katharina Glassen, Michel Wensing. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 07.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Help-Seeking Behaviors of Transition-Aged Youth for Mental Health Concerns: Qualitative Study

Chelsea Stunden<sup>1,2</sup>, BSc, MPH; Julie Zasada<sup>2,3</sup>, BHSc, MPH; Nicole VanHeerwaarden<sup>2,4</sup>, BSc, MHI; Elisa Hollenberg<sup>2</sup>, MSW; Alexxa Abi-Jaoudé<sup>2</sup>, MPH; Gloria Chaim<sup>5,6</sup>, MSW, RSW; Kristin Cleverley<sup>5,6,7</sup>, RN, PhD; Joanna Henderson<sup>5,6</sup>, PhD; Andrew Johnson<sup>2</sup>, BA; Andrea Levinson<sup>6,8</sup>, MD; Brian Lo<sup>2,4,9</sup>, BHSc, MHI; Janine Robb<sup>8</sup>, BScN, MSc; Jenny Shi<sup>2</sup>, BSc, MPH; Aristotle Voineskos<sup>6,10</sup>, PhD, MD; David Wiljer<sup>2,4,6,11</sup>, PhD

<sup>1</sup>School of Population and Public Health, Simon Fraser University, Burnaby, BC, Canada

<sup>2</sup>Office of Education, Centre for Addiction and Mental Health, Toronto, ON, Canada

<sup>3</sup>Schulich School of Medicine & Dentistry, Western University, London, ON, Canada

<sup>4</sup>Institute of Health Policy, Management & Evaluation, University of Toronto, Toronto, ON, Canada

<sup>5</sup>Margaret and Wallace McCain Centre for Child, Youth & Family Mental Health, Centre for Addiction and Mental Health, Toronto, ON, Canada

<sup>6</sup>Department of Psychiatry, University of Toronto, Toronto, ON, Canada

<sup>7</sup>Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

<sup>8</sup>Health and Wellness, University of Toronto, Toronto, ON, Canada

<sup>9</sup>Information Management Group, Centre for Addiction and Mental Health, Toronto, ON, Canada

<sup>10</sup>Slaight Family Centre for Youth in Transition, Centre for Addiction and Mental Health, Toronto, ON, Canada

<sup>11</sup>Education, Technology & Innovation, University Health Network, Toronto, ON, Canada

**Corresponding Author:**

David Wiljer, PhD

Education, Technology & Innovation

University Health Network

R. Fraser Elliot Building RFE 3E-411

190 Elizabeth Street

Toronto, ON, M5G 2C4

Canada

Phone: 1 416 340 4800 ext 6322

Email: [David.Wiljer@uhn.ca](mailto:David.Wiljer@uhn.ca)

## Abstract

**Background:** Transition-aged youth are particularly vulnerable to mental health problems, yet they are one of the least likely demographic groups to seek help.

**Objective:** The aim of this study is to explore the influences on and patterns in help-seeking for mental health concerns among transition-aged youth who attend postsecondary schools in Canada.

**Methods:** A qualitative research design was used, involving 12 semistructured focus groups with transition-aged youth (17-29 years) who attended postsecondary schools in Canada. A thematic analysis was conducted to code the transcripts and develop themes.

**Results:** Four main themes and subthemes regarding the process and experience of help-seeking were generated: (1) the influence of formal service providers (accessibility and experiences), (2) the influence of social factors (system navigation and stigma), (3) the influence of health literacy (symptom recognition, acting on symptoms, digital tools and the internet, and mental health awareness campaigns), and (4) the influence of low-intensity sources of support, namely, self-help.

**Conclusions:** Transition-aged youth seek help for mental health problems in different ways. Despite efforts to improve access to mental health services, transition-aged youth continue to face barriers to accessing these services, especially formal sources of support. The factors identified in this study that either hinder or facilitate help-seeking have pragmatic implications for developing help-seeking interventions and delivering mental health services for this population. In addition to other facilitators, family physicians are an important resource in the help-seeking process. Furthermore, digital help-seeking tools have unique characteristics that may make them an important source of support for transition-aged youth.

**KEYWORDS**

mental health; students; adolescent; substance abuse; eHealth; mHealth; mobile apps; help-seeking behavior; social stigma; social support

## Introduction

### Help-Seeking Behaviors of Transition-Aged Youth

The transition between adolescence and adulthood can be a particularly challenging time. Unique factors during this period, such as high levels of stress and anxiety, being introduced to alcohol and other drugs, and navigating new experiences and responsibilities, can make transition-aged youth vulnerable to mental health problems, which include substance use concerns [1]. Yet, they are one of the least likely demographic groups to seek help [2,3].

There is no universal definition of transition-aged youth. In the mental health literature, the age range spans anywhere between 14 and 29 years: some studies place it at 14-21 years, whereas others indicate 16-29 years [4]. For this paper, we adopted a broad definition of transition-aged youth of 17-29 years.

### Barriers to Help-Seeking

For this population, the process of seeking help for mental health concerns within the current social and health care system is arduous [5-7]. Barriers at the systemic, sociocultural, and personal levels can prevent or delay help-seeking among transition-aged youth [8,9]. A survey conducted in the United Kingdom found that 35% of youth with mental health concerns did not try to find help because of barriers such as poor access to care, perceived stigma, difficulty expressing concerns, and challenges navigating help-seeking processes or because they preferred to be self-reliant [10]. System-wide fragmentation makes it difficult to access appropriate services, information, and advice. These barriers have serious consequences, including self-medication and social, academic, and vocational difficulties [11]. They are compounded by other barriers such as long wait lists, stigma, high cost, lack of anonymity, and poor mental health literacy [12]. The result is often a failure to receive adequate mental health treatment [11-14]. The many barriers to help-seeking that transition-aged youth face need to be addressed to better serve the mental health needs of this population [8,9].

Eisenberg et al [15] identified 2 types of stigma associated with mental illness: perceived public stigma, which is defined as negative attitudes held by others, and self-stigma, which is defined as negative attitudes toward oneself. These 2 types of stigma interact and play a significant role in hindering help-seeking behaviors in transition-aged youth. Culture is another factor. Lack of cultural competence and sensitivity in mental health services, as well as the stigma attached to mental illness in many cultures, may make youths reluctant to seek help [15,16]. Together, these stigmas influence self-reliant behaviors, and transition-aged youths often prefer to cope with mental health problems on their own [1,2,16,17].

The nature of mental health problems can also be a barrier to seeking help. As these problems may have symptoms that are not easily recognized and often have a gradual onset, they might not be identified as issues that require medical attention; rather, the decision to consult a health care provider is influenced more by a voluntary help-seeking process [13]. Symptom recognition can be a factor in determining when to seek help.

### Facilitators to Help-Seeking

Although barriers to help-seeking for transition-aged youth have been well established in the literature, little is known about what facilitates help-seeking in this population [12,18]. Facilitators who have been identified include people having positive past experiences with formal and informal help-seeking as well as encouragement from others to connect with mental health services [12,19].

### Conceptual Help-Seeking Frameworks

In this paper, we use the conceptual framework by Rickwood and Thomas [13], which suggests that help-seeking is an adaptive coping process used to obtain external support to address a mental health concern. Within this framework, formal help is diverse and comes in the form of professional advice, support, and treatment delivered by professionals whose role it is to provide mental health care [13]. Semiformal help is offered by service providers who do not have a specified role in mental health care, such as teachers, coaches, and some community workers [13]. The third type of help, informal sources of support, such as friends and family, is also outlined in the framework and has been found to be important in encouraging youths to seek help [2,13,17,20,21]. Health Canada found that 27% of youth (aged 15-24 years) with mental, emotional, or substance use problems consulted informal sources compared with only 12% who consulted formal health professionals [20]. Self-help resources are a fourth type of support that includes sources such as self-help books and unguided website use. The framework also invites researchers to identify the stage within the process of help-seeking, the time frame, the type of help sought, and the type of mental health concern. Using a consistent framework to examine help-seeking is thought to increase the transferability of findings across studies and to better inform policy and practice decisions [13].

Confidentiality, trust, and positive relationships are important factors that encourage youths to pursue help from both formal and informal support groups [12]. Emotional competence (the ability to identify, describe, and manage emotions) can also predict help-seeking [17,22]. A survey of 300 undergraduate students found that those with low emotional competence were less willing to seek professional help and less likely to have positive experiences with mental health professionals [22]. More research is required to understand what factors promote help-seeking in transition-aged youth [8,9,23].



## Study Purpose

This study is part of a larger study that aims to evaluate and optimize a web-based and mobile health intervention called Thought Spot [24]. This crowdsourced digital platform enables transition-aged youth in postsecondary settings to look for and access mental health and wellness services. The app is a student-led project that prioritizes inclusion through steering committees, working groups, and focus groups. Thought Spot invites students to share their knowledge about services, discover wellness options on the web and in their geographical area, and post and read service reviews [24-26]. Although efforts have been made to improve access to mental health care for transition-aged youth, evidence suggests that barriers to care still exist [8]. The objective of this part of the larger study is to explore influences on and patterns in the help-seeking behaviors of transition-aged youth.

## Methods

### Setting and Participants

As part of the Thought Spot study [24], this paper reports the qualitative investigation of help-seeking behaviors that were generated through focus groups conducted throughout the study. A total of 17 focus groups were conducted on various topics as part of a participatory design research process to optimize the Thought Spot app. A subset of 12 focus groups that addressed help-seeking were examined for this paper. Participants were recruited using convenience sampling through web and offline methods that included posters, postsecondary school recruitment presentations, word of mouth, and social media posts (Facebook, Twitter, Instagram, and Slack). English-speaking youth aged 17 years or older who were enrolled in full- or part-time studies at a Canadian university or college were eligible to participate.

### Data Collection Procedures

Recruitment for this study took place over 3 years (July 2016-August 2018). Participants were enrolled on a first-come, first-served basis, with up to 10 participants per focus group. Priority was given to individuals who had not yet participated in the study. However, if there were fewer than 10 participants in a group, participants who had already attended another group were permitted to participate again if they requested to do so in response to study recruitment advertisements for subsequent groups. Participants were allowed to engage in more than one focus group, as each focus group was structured with different activities and modified help-seeking questions (Multimedia Appendix 1) [26]. In addition, we attempted to build engagement and capacity in the participatory design research process by having students participate and contribute more than once. Contributions of repeat participants were therefore expected to vary from group to group.

All participants signed informed consent forms and filled out a demographic form when they arrived for each focus group. A unique identifier was created based on participant demographics. Those who participated in multiple focus groups filled out a new demographic form for each group to avoid breaking confidentiality about having attended a previous group. Repeat participants were expected to use the same unique

identifier each time. Participants received Can \$40 (US \$31) and bus tokens as remuneration.

A semistructured interview guide (Multimedia Appendix 1) was developed for the overall study, with questions focusing on seeking help and health information.

Interviews were conducted in English and audio-recorded. Nine master's-level practicum students co-designed and facilitated the focus groups with the study team. There were 3 practicum student facilitators per group—1 acting as a note-taker, 1 as a flipchart assistant, and 1 as the main facilitator. These students included the authors (BL, CS, JS, JZ, and NV) and 4 students listed in the Acknowledgments section of this paper (AK, PD, SS, and VT). A staff research coordinator (AA or GF) and research analyst (EH) also attended the focus groups to provide general oversight and guide facilitation. Choosing to use practicum students to co-design and facilitate the groups reflected the decision to have transition-aged youth co-design the broader project. The focus group questions were broad and allowed participants to describe and discuss their experiences in as little or as much detail as they wanted.

### Data Analysis

Data collection continued until researchers determined that information saturation had been reached, as outlined in the published protocol [24]. Audio-recordings were transcribed verbatim, and data were entered into NVivo 12 (QSR International Pty Ltd) software after anonymizing and quality-checking transcripts for accuracy. Data were then coded using the 6 phases of thematic analysis by Braun and Clarke [27]. Phase 1 began with 1 researcher (JZ) reviewing the data and noting ideas as preliminary codes. In phase 2, 2 researchers (JZ and CS) generated initial codes based on all the transcripts and organized data relevant to each code. These codes included attributes of resources, types of service delivery, recommendations for future services, facilitators to use, barriers to use, outcomes of help-seeking, and help-seeking processes. In phase 3, 2 researchers (JZ and CS) organized data into potential themes based on the researchers' understanding of the data and in relation to any categories that may apply from the conceptual framework by Rickwood and Thomas on help-seeking [13]. The themes were reviewed with a wider research team for input and feedback around grouping and conceptualizing the codes into themes. Updates to the themes included further refinements of the types of facilitators (eg, internal or external) and barriers to help-seeking (eg, concern with health care providers, costs, lack of social support, stigma, system navigation, time). The themes were then reviewed in phase 4 by 1 researcher (CS) to ensure that they worked in relation to the coded extracts, and the researcher also defined the final themes. The themes were reviewed and verified by the second researcher (JZ) in phase 5. Phase 6 involved finalizing the selection of quotes for analysis (CS) and relating the analysis back to the research question and then reviewing and confirming the analysis (JZ). The study received Research Ethics Board approval by the ethics boards at the Centre for Addiction and Mental Health, University of Toronto, Ryerson University, and George Brown College.



## Results

### Demographics

Participants were male and female aged 17 to 29 years who attended postsecondary schools in Canada. A total of 110 requests were received from students to attend the 12 focus groups examined for this paper (some students requested taking part in more than 1 focus group). Of these requests, 24 did not result in focus group attendance for various reasons. In 13 requests, participants signed up for a group but then canceled their subscription, 5 (39%) were ineligible, 3 (23%) were not available on the dates of the groups, 2 (15%) could not

participate as the groups were full, and 1 (8%) did not respond after initial contact. The remaining requests resulted in 86 focus group attendees and 69 unique participants, based on our attendance data. A total of 10 participants attended more than 1 focus group, 2 attended 4 focus groups, 3 attended 3 focus groups, and 5 attended 2 focus groups. There was an average of 7 participants per focus group (range 4-9).

We were unable to remove 4 misidentified duplicate demographic surveys from our data. Therefore, although there were only 69 unique participants in the focus groups, we report the demographics of 73 respondents in [Tables 1](#) and [2](#). The demographics of the focus group participants are listed in [Table 1](#).

**Table 1.** The demographics of the focus group participants (N=73).

Demographics	Values, n (%)
<b>Gender</b>	
Male	20 (27)
Female	53 (73)
Nonbinary	0 (0)
<b>Age (years)</b>	
17-21	27 (37)
22-24	32 (44)
25-29	14 (19)
<b>Highest level of education achieved</b>	
Some university but no university degree	27 (37)
College diploma	6 (8)
Bachelor's degree	23 (31)
Some graduate school but no graduate degree	6 (8)
Master's degree	4 (5)
Other education or no response	7 (10)
<b>Ethnicity<sup>a</sup></b>	
Aboriginal (eg, First Nations, Métis, Inuit); Arab or Middle-Eastern; and Black (eg, origins include Canadian, American, Caribbean, African)	7 (9)
East and Southeast Asian (eg, Vietnamese, Cambodian, Korean)	4 (5)
Chinese	19 (24)
South Asian (eg, East Indian, Sri Lankan)	18 (23)
White or European	27 (34)
Don't know; other group	4 (5)
<b>Experience with mental health and substance use problems</b>	
Yes	38 (52)
No	26 (36)
Don't know or no response	9 (12)

<sup>a</sup>Some respondents answered in more than one category.

**Table 2.** Self-reported types of mental health or substance use problems (N=73).

Mental health problem <sup>a</sup>	Total, n (%)
Anxiety	19 (26)
Mood disorder (eg, depression, bipolar)	17 (23)
Comorbidity	16 (22)
Has a family member or friend with mental health or substance use problems	7 (10)
Borderline personality disorder, eating disorder, psychosis, posttraumatic stress disorder, and suicidal ideation	8 (11)

<sup>a</sup>Participants were able to indicate more than one answer.

All participants were identified as either male or female on the demographic form. *Trans* and *other* were also listed as gender options; however, no participants chose either of them. A subset (39/73, 53%) of participants elaborated on the types of mental health concerns they had (Table 2). To protect confidentiality and anonymity, if there were less than 4 responses for any of the categories in the demographics form, the responses were combined with other categories for the same question that also had less than 4 responses (Tables 1 and 2). This approach resulted in combined responses for some ethnicity categories and some types of mental health concerns.

Four main themes related to help-seeking were generated through transcript analysis and coding: formal services and providers influence help-seeking, social factors influence help-seeking, health literacy influences help-seeking, and self-help–low-intensity sources of support influence help-seeking. These themes and subthemes are listed in Multimedia Appendix 2.

### Formal Services and Providers Influence Help-Seeking

Formal sources of support are service providers with a specific professional role in delivering mental health (including addiction) care, such as family physicians, psychiatrists, and psychologists [12,13]. Participants discussed 2 main influences on seeking help from formal service providers: accessibility and empathy and trust.

#### Accessibility of Formal Service Providers

The ability of participants to access a formal service provider when they needed one was a common topic among the participants. Several participants reported contacting multiple formal service providers while help-seeking. They had difficulty accessing this type of support for various reasons. Participants often expressed concern about the instrumental costs (money and time) of seeking help from formal service providers. They also indicated that it was important during initial contact for service providers to create an empathetic and trusting environment. Establishing this connection made young people more willing to share their mental health concerns during subsequent encounters and seek treatment, services, or pathways to support in the future.

#### Wait Times Affect Access to Formal Services

Long wait times to be seen by formal service providers frustrated many participants and influenced their help-seeking behavior. One participant stated:

*The wait times for getting any support are ridiculous no matter where you go. Oh, a year, two years, six months. I'm already in crisis right now and I want it right now or like a week from now, not a year from now.* [Focus Group 1]

Participants were also discouraged by the reality of having to consult multiple service providers to address specific aspects of their mental health concerns, a fragmentation that precludes holistic support and increases wait times. One student told us:

*Everywhere you go...they want to only help you with one issue, it's just not helpful.* [Focus Group 1]

When the desired treatment or support was not available in a timely way, participants sought alternative types of assistance, such as crisis services. One participant spoke about a friend:

*If they were going through a crisis but they were waiting for the next appointment...they would always just go to the emergency department.* [Focus Group 7]

#### Financial Burdens Affect Access to Formal Services

Participants often described the financial burden associated with formal service providers (eg, psychologists, psychotherapists, counselors) in the community as a barrier to accessing help and pursuing optimal mental health. Some participants noted that although mental health services at postsecondary schools were generally more affordable than services in the community, these services sometimes make referrals to specialists in the community for students with serious mental health problems. Participants identified private psychotherapy as being inaccessible as it is expensive. They perceived pharmacological treatments to be more affordable. Accessibility was also limited by some health plans in Canada that did not cover participants' preferred treatments. One student told us:

*You often have to pay. And if you have no money or minimal money, there are not a lot of places that do scaling or they don't inform you that they do scaling. So, if you want to see a social worker or a psychologist or psychiatrist, it's often you need to pay \$150 an hour.* [Focus Group 3]

#### Ease of Access Affects Access to Formal Services

The ease of access supported participants' help-seeking efforts. Factors that increased access included convenient locations and times, anonymity and confidentiality, affordability, appointment booking, clearly displayed eligibility criteria, and short wait times. One participant stated:

*I felt too anxious to really talk about things, so it really helped that they had an email address and I didn't have to call in to make appointments, I could just schedule them over email.* [Focus Group 7]

Services and resources that provided a variety of options simplified the help-seeking process. Participants spoke highly of services that offered various entry points, multiple services, and several ways to learn about mental health. Some participants described their frustration with receiving siloed treatment for comorbid conditions and liked having services colocated in one place. One participant told us:

*Agencies are becoming hubs now, so most places have everything on-site, so people don't have to travel everywhere to get what they need.* [Focus Group 5]

Many participants identified family physicians as an important initial contact in the help-seeking process. Physicians helped them to navigate the mental health system and gave them information about other services or provided referrals to them. Family physicians typically made referrals to specialized mental health services. One student told us:

*I think the easiest way is probably to go through a doctor or some sort of professional.* [Focus Group 3]

### Experiences With Formal Service Providers: Empathy and Trust

Participants discussed both positive and negative experiences with formal service providers that influenced their help-seeking behaviors. Negative experiences with health care providers and fear of similar experiences in the future deterred help-seeking. Many participants described a lack of empathy or understanding from health care professionals or had heard about others' bad experiences, which made them hesitant to connect with these service providers. For several participants, the lack of connection made it difficult to talk about their problems. A student told us:

*I don't feel like I can trust another doctor. . . [They tell you] "It's temporary. It will pass. It's part of your life. You went through a crisis. It will go away. Don't worry about it." Or "I don't believe you. Maybe come in a week and you'll feel different." That type of stuff. There's a lot of, oh, oh. And if you do have mental illness, oh, maybe [you're] lying. Maybe [you] actually don't know what [you're] talking about. So there's that sometimes as well.* [Focus Group 1]

Negative past encounters decreased motivation to pursue help. Participants who had tried many times without success to access mental health support were skeptical about getting the help they needed or wanted. Some also lacked confidence in their current health care provider's abilities or were dissatisfied with their care. Participants described feeling hopeless in these situations and being unable to seek help elsewhere because of long wait times or high costs. One student mentioned:

*It's hard to start over because if you lose the person you have, you're going to have to start all over in the line. They're not going to just give you someone else. You're going to have to accept whatever you've got, and good luck trying to find someone else, because*

*it took six months to a year to find whoever you got in the first place.* [Focus Group 1]

Participants also mentioned the importance of receiving feedback from peers before taking action. They wanted to hear about the experiences of others with a particular support before they invested their own time and resources. One participant said:

*It kind of motivates you ... If someone else did this and found it helpful, I'm going to do the same thing as well.* [Focus Group 3]

### Social Factors Influence Help-Seeking

Social factors influenced help-seeking in both positive and negative ways. Participants discussed how informal supports (eg, friends, partners, parents) and semiformal supports (eg, teachers, work supervisors, academic supervisors, youth workers, coaches) could either facilitate or hinder help-seeking. Moreover, social stigma made it difficult for participants to speak about their mental health concerns and needs in their social networks.

### Social Support Affects System Navigation

Social support networks were critical during help-seeking. Participants felt empowered when they received social support or encouragement or when they saw someone else succeed in getting help. One student mentioned:

*I find a lot of times when someone describes an experience that they've had, I find that much more trustworthy than just looking at the idealized version of what it's supposed to be.* [Focus Group 3]

Participants identified different chosen social support networks, and how useful they found these different support networks also varied. Some participants felt that they had limited support from friends or family. Some attributed this lack of support to attending school far from home. Others mentioned that social isolation was connected to experiencing more severe symptoms of mental illness. Both situations hindered help-seeking. Complying with treatment was particularly difficult for participants who did not have family support. One participant told us:

*Sometimes...you ask for help and your family says we're not going to help you with that...You're prescribed medication and...your parents won't give you the money you need to buy them. But then your counselors are also like, "Well, you can't not take the medication."* [Focus Group 7]

Other participants had friends or family who were able to recognize and respond constructively to their mental health problems and facilitated the help-seeking process on their behalf. One participant said:

*So, normally in high school...at least in my experience, if you're struggling with mental health issues, your parents will kind of handle it. Like, your parents will make your doctor's appointments and find referrals and find you a psychologist or a psychiatrist. And by [post-secondary school], you have cycled through the system enough times that*

*you're fairly familiar and you know what to seek out. However, I know a lot of people who started university and, in their first year, developed some mental health issues. A lot of friends of mine would ask me, because they knew that I had experience with that, and they just feel lost.* [Focus Group 8]

For some, informal social supports were the first step to finding help. These people helped participants navigate the process and shared their own experiences. A student told us:

*We went online, but we found the online help to not be very specific or applicable and [we] had to take it one step further and consult either a friend or somebody who has been through it to get more of the personal touch.* [Focus Group 7]

Semiformal supports who do not have a specified role in mental health care (eg, coaches, residence fellows, professors) also helped participants navigate the system and checked in with them to see whether they had followed through on their intended action. One student told us:

*I've seen profs facilitate with counselors, making sure that the student is able to find someone or know where to go. . . If you tell someone and they're able to help you through the process and just support you through it, then makes it a lot easier.* [Focus Group 1]

### Stigma Affects Help-Seeking

Stigma was reported as a deterrent to seeking help from formal and informal supports. Not all participants had social networks that offered support around seeking help for mental health concerns. One participant mentioned:

*Within my group of friends. . . we don't talk about mental health that much, if at all.* [Focus Group 3]

Participants discussed the stigma of mental health problems and their fear of discrimination if they disclosed their mental health concerns. Internalized stigma, feelings of shame and embarrassment, and being in denial about their situation were common experiences. One participant said:

*I think sometimes seeking help can also be like a sign of weakness and that kind of stops you from going to seek help. Like, you feel like you should be able to deal with these things by yourself.* [Focus Group 9]

Participants were often afraid of being judged by friends, peers, and family. One student told us:

*Being involved in a group of friends that actually make fun of mental health problems, you can't even talk about it.* [Focus Group 7]

Another student said:

*Before, I had to go through my parent's insurance plans and they would know what I was taking all the time and it would make me feel uncomfortable and judged.* [Focus Group 7]

Participants also feared discrimination by their professors and peers and the potential consequences of disclosure for their career or academic ambitions. One participant said:

*I'm not sure if telling my personal or health issue to my prof will actually. . . what if I need a recommendation letter from this person later? Will this color their opinion of me and my work abilities?* [Focus Group 1]

One participant described how the stigma attached to pharmacological treatment made a friend hesitant to seek help:

*It's not just the side effects that he was worried about. He was also worried about the stigma associated with having to take medications for his depression.* [Focus Group 1]

Participants who were trained to be mental health service providers (eg, social workers, medical students, residents) also spoke about the unique stigma associated with help-seeking as a mental health care provider. One student in social work told us:

*I'm in social work. You're supposed to be able to help people. You're not supposed to be the sick one too... You're not supposed to be your clients... even though maybe those that need the help are those that are providing the help as well. And it's sometimes fearful for those that are providers just because they are told that they're not supposed to [seek help from formal health service providers] . . . because then they won't want to hire you.* [Focus Group 3]

As a result of these concerns, many participants wanted to keep their mental health problems private and seek help on their own. One participant said:

*I think it's still taboo to talk about getting help for mental health or addictions. There's more access now, but I think. . . people don't like asking for help.* [Focus Group 1]

Although societal and internalized stigma made some participants reluctant to seek support from their social networks, several participants spoke about the importance of finding people they could relate to. One student told us about a friend's experience:

*After she was diagnosed and everything, she told one friend about it and from there it just started a chain reaction, and one by one, everybody started saying. . . "Oh, we have that problem too, but we were too scared to tell anybody about it because it's such a taboo with our parents."* [Focus Group 1]

### Health Literacy Influences Help-Seeking

Participants discussed factors related to health literacy that could influence help-seeking. Mental health literacy is an awareness and maintenance of one's own mental health needs, becoming knowledgeable about formal diagnoses, treatments, and the effect of stigma, and being able to effectively seek help when needed [28]. In this study, steps toward health literacy by participants included recognizing symptoms, acting on symptoms, learning about mental health through web-based sources or awareness campaigns, and finding out what services are available. Participants used multiple entry points to learn



about mental health problems and the types of services they could access.

### **Health Literacy Affects Symptom Recognition**

Participants described difficulty in identifying mental health concerns and attaining and maintaining positive mental health. They noted that they needed first to be self-aware and recognize that their psychological state was out of the ordinary before they would access services. One participant said:

*I think a big thing was. . .self-awareness and this realization that maybe. . .this is not how you should be feeling every single day. That's like a very big turning point where you realize that. . .it's not normal to feel this way every day and you shouldn't have to feel this way every day. [Focus Group 3]*

Lacking this insight or not understanding how to find help, which includes getting a diagnosis, made it difficult to search for appropriate services. One participant told us:

*The very first step is to get a diagnosis. However, it's very difficult for people to take that step. Like, I genuinely do not know what services or programs are available for people who are just kind of experiencing symptoms but don't have a diagnosis. . .because. . .a lot of places. . .only provide these programs and services for people who already have a proper diagnosis. [Focus Group 8]*

Several participants also explained that youths might not know what types of assistance are available and need help to find appropriate resources or support. One participant said:

*Yes, it's hard to define what mental health help actually is. And so people tend to [run] in circles, avoid conversation around it because they don't really know how to describe what it is. [Focus Group 2]*

Some participants described not always being able to articulate their mental health history and symptoms, which made it difficult to solicit support from their networks. One participant said:

*I feel like maybe sometimes you're sitting at the doctor's office and you're probably experiencing some. . .mental health problems that you also want to talk about, [but] you don't know how to talk about it, and then time's up and the doctor needs to move on to the next patient, and then it kind of falls through the cracks. [Focus Group 1]*

### **Symptom Recognition Does Not Always Lead to Help-Seeking**

Although recognizing symptoms was the first step in seeking help, self-awareness did not always lead to action. Some participants described being aware of their distress but downplaying it or not knowing when to seek help. One participant told us:

*I find I always do this thing where I end up looking at someone that's worse off and I'm like, I can probably handle it or I can probably deal with it. [Focus Group 9]*

Another participant said:

*Participants also described lacking the motivation to pursue help, because of the symptoms of their mental health problem and "feeling so hopeless that you don't even want to seek [help]." [Focus Group 2]*

Participants who recognized that something was wrong described weighing the pros and cons of getting help. Some participants eventually accessed services as certain stressors made them feel that their situation had evolved into something more serious and they could no longer avoid getting help. One participant said:

*I'll never be feeling fine and thinking, "Oh, I should probably worry about my mental health right now." It's usually when I'm already at that low point, where I'm just desperately seeking mental health services. [Focus Group 7]*

### **Health Literacy Is Affected by Digital Health Tools and the Internet**

Accessing digital health resources on the web increased participants' mental health literacy. Participants often sought information through web-based media to validate their feelings and identify what kind of help was available for their particular concerns. One student told us:

*I'd go on [website]. . .it's not a diagnosis, but it can help you seek out one. I found that what I got was essentially what the doctor told me. [Focus Group 8]*

Using mental health resources effectively required being able to match information or services to identified needs and then feeling comfortable and confident reaching out to these supports. Participants often described web-based resources as an important entry point to help-seeking. One student said:

*I feel like the internet can be a good stepping stone. It's not really somewhere that you find conclusive answers, but it's a good starting point. [Focus Group 7]*

Participants often described using the internet and social networking websites or mobile apps (Facebook, Instagram, and YouTube) to look for support and mental health information. One participant said:

*There's a lot of Facebook group communities that sometimes are available. Not really services, but, like, groups of people that are like, "Yeah, I have the same problems." [Focus Group 8]*

Participants spoke about learning from the experiences of others. One participant said:

*I've also seen different videos, where a lot of YouTubers that I used to watch or I sort of watch, they've also gone through mental health issues. So a lot of them post about their experiences...Seeing someone open up about an experience and maybe you can relate, then it really takes off a lot of stress. [Focus Group 9]*



However, web-based information about mental health could be confusing, so participants sometimes reached their own conclusions. One student told us:

*A lot of different mental illnesses have the same symptoms... You end up self-diagnosing yourself with a bunch of different illnesses, which you might not have all or any of them.* [Focus Group 9]

Participants who sought support or resources on the web also described the information as ambiguous and the amount of information as overwhelming. One student said:

*It definitely looks like there's a lot of resources out there for a variety of needs and lots of stress[ors] people might be feeling, but... first being able to find and decide upon which one you actually want to pursue takes a lot of time. Then, when you actually try to access it, it seems really difficult and daunting.* [Focus Group 9]

Despite the issues that participants highlighted about using web-based resources, digital media was a popular source of help-seeking.

### **Health Literacy Is Affected by Mental Health Campaigns**

Participants reported that mental health campaigns improved their mental health literacy. Social marketing and other mental health campaigns increased motivation to seek help. Participants described feeling empowered by hearing other people's stories about finding support. One student told us:

*Bell Let's Talk Day comes around, and [people] start posting all over Facebook... being like, "I started out on this and got more comfortable with it, and now I'm better and I want to help other people get better."* [Focus Group 7]

Other participants recalled learning about support through promotional events, guest speakers, emails, posters, and pamphlets in high school. One student said:

*We had a social worker in high school that would come to all of the classrooms, so you were getting to see a social worker regardless.* [Focus Group 4]

Some participants recalled learning about mental health services before they personally experienced mental health symptoms. They became aware of these services through information sessions, word of mouth, emails, posters in university common rooms, pamphlets at community centers or health offices, advertisements on websites, or in-course syllabi offering advice or support for mental health and wellness. Many participants mentioned a phone helpline that was advertised on cereal boxes. Knowing about mental health resources before they were even needed supported navigation by providing a good starting point. One student told us:

*Even before I had to use the service, I was aware of what it was. Even before I had an inkling that I would ever need to call, I just knew what it was.* [Focus Group 4]

### **Self-Help: Low-Intensity Sources of Support Influence Help-Seeking**

Participants reported trying to take care of their mental health concerns on their own and discussed the importance of resources that enable self-management. They described self-help and wellness activities for managing mental health concerns as generally low-cost, self-guided, and immediately available. Participants who were not comfortable accessing crisis or formal services found self-help particularly useful.

Participants often experimented with making small changes in their lives through self-help apps that do not involve live support (eg, e-learning, web searching, computer-mediated therapy, mindfulness sessions). One student said:

*On YouTube... it's not actually going to see someone for cognitive behavioral therapy, but there's resources online of learning some of the techniques that they use and maybe they can implement, even if they don't actually go seek help [formally].* [Focus Group 4]

Another student mentioned:

*They had these online Instagram things where every day of the week it was a day of self-care.* [Focus Group 2]

Participants also accessed various digital mental health resources, such as Headspace [29], Meditation Studio App [30], 7 Cups [31], Buddify [32], and Virtual Hope Box [33]. Participants used these apps to manage their mental health every day, track patterns in their mood, or chat anonymously with people about their mental health concerns. One participant said:

*People who are in situations that don't have a lot of time – students, mature students, parents, etcetera – use online stuff like 7 Cups.* [Focus Group 5]

Unlike participants with more urgent mental health care needs, those who had only mild or moderate symptoms felt that crisis-based services were not a good fit for their needs, although they seemed to be the most readily available and accessible. One participant said:

*When I think of a hotline, I think of like an emergency or dire situation. It's not something that comes as a first resource for me.* [Focus Group 9]

Instead, participants with less urgent concerns tried to meet their needs through other accessible and affordable support for overall health and wellness (eg, fitness, meditation and mindfulness, journaling, nature walks, playing music, listening to music, playing with pets). One participant told us:

*[I do] hobbies that require a lot of concentration. If you play a musical instrument, you're forced to focus on the instrument and the music, and then your mind just completely forgets what it was stressed about at the time, so it kind of takes your mind off that, and when you come back to that situation, you kind of see things in a new perspective.* [Focus Group 3]

Other wellness resources were also common. Some participants temporarily distracted themselves from stress and unpleasant tasks with *guilty pleasures*, such as watching YouTube or

Netflix, or playing video games to clear their minds. However, they recognized that these activities provided only a temporary reprieve. One participant explained that Netflix sometimes reduced productivity as it was easy to spend a lot of time on it and neglect important tasks. One student said:

*Netflix is a common thing now. I think everyone watches a lot of TV. . . But, if you're obsessive, like watching massive amounts of TV, then it can really make things worse by procrastination of everything.*

[Focus Group 3]

Although participants identified many low-intensity sources of support, they recognized that some activities that youths use to cope with difficult feelings or situations may not be healthy. One participant said:

*Professionals, they're not really going to recommend going out for drinks, but sometimes going out for drinks just helps you deal with it.* [Focus Group 8]

Other participants pursued wellness by accessing low- or no-cost community- or school-based mental health resources. Alternatives to crisis services or professional support are listed in [Multimedia Appendix 3](#).

## Discussion

### Principal Findings

Despite the high prevalence of mental health problems that develop in adolescence and early adulthood, many young people still do not seek help [10,34-36]. This study identified many of the barriers to and facilitators of help-seeking that previous research has found, such as accessibility of sources of support, experiences, system navigation, stigma, symptom recognition, and behavioral intention. It also offers new insights for developing strategies and services to meet the mental health needs of transition-aged youth by discussing the role of digital resources in the help-seeking pathway and the types of self-help resources that young people look for during this process. We found that help-seeking differs significantly among transition-aged youth in terms of the types and forms of support they prefer. Many use a combination of support, which can be categorized into formal, informal, and self-help support, using the conceptual framework by Rickwood and Thomas [13]. Our study contributes to the existing literature by identifying a range of self-help options beyond websites that were accessed by youth. We added a new category for digital self-help apps that offer peer support, therapeutic techniques, or other forms of self-help ([Multimedia Appendix 3](#)) because they represent a fairly recent distinctive self-help category not included in previous classifications. Identifying these types of support and categorizing them according to a conceptual framework is an area for ongoing development within the help-seeking literature, in which most types of assistance are not adequately described or classified [13].

### Frequent Use of Low-Intensity Services During Help-Seeking

Our study suggests the frequent use of low-intensity services, such as the web or mobile apps, among transition-aged youth.

These digital services were noted as particularly helpful because of their ease of access. Our findings support other research that identifies web-based mental health resources as a primary source of support for transition-aged youth [13]. Many participants in our study reported using educational websites about therapeutic techniques, peer-to-peer active listening services, interactive self-help mobile apps, and social media (eg, Facebook, Instagram, Yik Yak). Recreational activities, accessed through websites such as Groupon and Meetup were also mentioned as ways to support mental health and overall wellness. The range of digital self-help apps that the youth in our study accessed are listed in [Multimedia Appendix 3](#). Our findings add to the current literature by suggesting that popular digital mental health and wellness apps may be unique access points for reaching young help-seekers. Improving and expanding these resources should become a priority, particularly where traditional offline mental health services are difficult to access [37]. More research is needed to determine the effectiveness of these emerging types of support.

### Web-Based Health Information May Hinder Help-Seeking

Self-directed efforts to increase health literacy also influenced the help-seeking process. Digital health solutions and the internet can facilitate access to information about mental health problems and services and increase mental health literacy. However, participants in our study cautioned that web-based self-help resources can also be counterproductive if users are unable to decide which resources match their needs. We reiterate other researchers' cautions that the web can cause information overload and increase the potential for harm (such as incorrect self-diagnoses) when help-seekers lack the skills and experience to evaluate the information they find on the web [12,38,39]. New coordinated public health initiatives should focus on teaching transition-aged youth how to evaluate the quality and credibility of web-based health information.

Although seeking help on the web was common, it was not always helpful, and the credibility of sources was sometimes uncertain. Youth also accessed emotional support on the web where they could anonymously engage with a community of peers who offered support by sharing points of view and helped them decide what help to pursue. Web-based help-seekers read testimonials and web-based reviews about what to expect from each service provider.

### Social Marketing Campaigns and Educational Outreach May Catalyze the Help-Seeking Process

In addition to actively pursuing help, transition-aged youth are also learning about mental health supports through social marketing campaigns and educational outreach. Research has found that youth who are exposed to social marketing or educational outreach around mental health are more likely to reach out for help and get treatment [40,41]. Our research adds to these findings, suggesting that early educational intervention has long-term positive effects on help-seeking in young people. More research is needed to determine the most effective way to encourage transition-aged youth to seek help for mental health problems.

## Instrumental Barriers to Help-Seeking

The findings of our study are also consistent with research that shows that general mental health care is largely inaccessible because of instrumental barriers that include distance, cost, and wait lists [42,43]. Although on-campus services at postsecondary schools in Canada can mitigate some of these barriers, for example, they are usually supported by tuition and ancillary fees or are covered by student insurance, these services also often have long wait times. Additionally, walk-in appointments and drop-in sessions are not always offered. Despite the advantages of campus-based services, our findings suggest that barriers remain for transition-aged youth who seek services on and off campus.

The extensive wait times reported in the literature were echoed in our findings. Only about 8.6% of agencies in Canada do not have a wait list for child and adolescent programs or services. However, timely access to treatment is critical because long wait times not only prolong psychological and physiological discomfort but can also increase attrition rates, reduce motivation to seek help, and decrease expectations and positive treatment outcomes [44,45]. Barriers to accessing care are compounded by other barriers, such as service fragmentation, poor health literacy, low motivation to change, and lack of social supports. Together, these barriers lead to underused or inappropriate services, poor follow-up rates, decreased service quality, and high health care costs.

## Mental Health Service Providers Influence Help-Seeking

Previous experiences with formal service providers were either deterrents or facilitators to help-seeking, particularly given the challenges of navigating the fragmented mental health care system. Past help-seeking experiences that were empathetic and instilled trust encouraged future help-seeking. Similarly, negative, stigmatizing experiences discouraged help-seeking. Our findings contribute to existing evidence that positive experiences with formal health providers, which focus on building a trusting relationship, are important to youth [12].

Family physicians are increasingly being recognized for their role in the mental health help-seeking process by supporting transitions between youth and adulthood [46], providing treatment [47], and making referrals to specialists [17]. Transition-aged youth in our study primarily viewed family physicians as gatekeepers for referrals to mental health specialists. Other studies have also found family physicians to be the first point of contact and to be responsible for making referrals, despite facing numerous barriers themselves (eg, lack of time, knowledge, resources, training) [47-50]. Our findings reiterate the unique role of family physicians in help-seeking and point to the need to train and support family physicians in guiding young patients through system navigation and treatment. Significant benefits could result from developing supports for family physicians to manage and diagnose young people with mental health concerns by offering timely access and continuity of care.

## Help-Seeking for Mental Health Providers in Training

Transition-aged youth in our study who were undergoing training to become formal mental health service providers, such as social workers or physicians, described a unique barrier to help-seeking in the form of stigma. As future professionals whose responsibility is to help others, they feared that disclosing their own mental health problems would damage their reputation and career prospects. However, mental health problems among health care professionals are well-documented and include depression, anxiety, substance abuse, and suicide risk [51-55]. Other research suggests that emotional distress during training (eg, medical residency) can persist into professional practice and lead to burnout [56]. To our knowledge, no research has been conducted on the help-seeking behaviors of this population; however, our findings underscore the importance of offering formal mental health supports tailored to the needs of mental health care providers in training and point to the need for further research in this area.

## Limitations

The qualitative approach of this study enriches our understanding and insights into the mental health help-seeking behaviors of transition-aged youth; however, the findings may not be generalizable beyond the Greater Toronto Area, where the study took place. Help-seeking behaviors likely differ across postsecondary and cultural contexts, particularly at postsecondary schools that do not have access to on-campus support services. In this study, it was not possible to determine the effects of location, sex, gender, or culture on help-seeking behavior. Our findings are useful for health care providers who are creating digital mental health programs and services to improve access to care. Using focus groups provided an opportunity for participants to express their personal experiences while navigating the help-seeking process. However, a limitation of focus groups is the potential toward normative discourse, especially in the context of mental health where people may not be comfortable talking about certain issues in a group setting. Another limitation was that certain groups were underrepresented in this study: depression and anxiety were the mental health problems reported by the vast majority of participants, and most participants were female. Youths with serious mental health problems (eg, psychotic disorders) and transition-aged males are known to be less likely to access any sort of health care or to participate in research [57]. These groups may be better engaged in their help-seeking behaviors through other channels, such as technology, health clinics, schools, and social media [58].

Multiple focus group attendance by participants who requested it was another possible limitation in this study (10 participants attended more than 1 group) as it has the potential to amplify some themes in relation to others. The rationale for agreeing to these requests was to increase participants' overall, sustained engagement with the participatory design research process throughout the study. As described in the Methods section, because of the differences between focus group topics, the contribution of participants was expected to be different from group to group. Another limitation related to repeat participants, who filled out a demographics form each time they attended a

focus group, was that we were unable to remove 4 misidentified duplicate demographic surveys from our data. This resulted in reporting on demographics for 73 participants, although only 69 unique participants were involved.

## Conclusions

Transition-aged youth are influenced by various barriers and facilitators when they seek information and support for mental health problems. Future investments should focus on reducing barriers associated with accessibility, system navigation, stigma, and symptom recognition. This study also identified digital health tools, the internet, and low-intensity sources of support

as important parameters in the help-seeking process. Family physicians were found to be important gatekeepers to other mental health services. Digital and web-based technologies and the additional supports or features they offer may also be powerful tools for promoting help-seeking. Young people who are trained to become mental health care providers may experience a unique barrier to help-seeking in the form of stigma, an area that warrants future research. There is also a need for research on the effectiveness and unique features of low-intensity supports in seeking help for mental health problems.

## Acknowledgments

The authors wish to thank all the study participants, the Thought Spot Student Group (a student research advisory committee), research team members, research coordinator Genevieve Ferguson, and the 4 practicum students who co-designed and facilitated groups: Amandeep Kaur, Paulysha De Gannes, Vivetha Thambinathan, and Shehab Sennah. This project was approved and funded by the Canadian Institute for Health Research. The authors are also grateful for the collaboration and support of the University of Toronto, Ryerson University, and George Brown College during this research.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Semistructured interview guide.

[DOCX File, 15 KB - [jmir\\_v22i10e18514\\_app1.docx](#)]

### Multimedia Appendix 2

Themes from qualitative analysis.

[DOCX File, 15 KB - [jmir\\_v22i10e18514\\_app2.docx](#)]

### Multimedia Appendix 3

Mental health supports discussed by study participants.

[DOCX File, 17 KB - [jmir\\_v22i10e18514\\_app3.docx](#)]

## References

1. Rickwood DJ, Mazzer KR, Telford NR. Social influences on seeking help from mental health services, in-person and online, during adolescence and young adulthood. *BMC Psychiatry* 2015 Mar 7;15:40 [FREE Full text] [doi: [10.1186/s12888-015-0429-6](#)] [Medline: [25886609](#)]
2. Rickwood DJ, Deane FP, Wilson CJ. When and how do young people seek professional help for mental health problems? *Med J Aust* 2007 Oct 1;187(S7):S35-S39. [Medline: [17908023](#)]
3. Pearson C, Janz T, Ali J. Health at a Glance - Mental and Substance Use Disorders in Canada. Statistics Canada. 2013 Sep. URL: <https://www150.statcan.gc.ca/n1/en/pub/82-624-x/2013001/article/11855-eng.pdf?st=2n15aWjr> [accessed 2020-09-22]
4. Mandarino K. Transitional-age youths: barriers to accessing adult mental health services and the changing definition of adolescence. *J Hum Behav Soc Environ* 2014 Apr 30;24(4):462-474 [FREE Full text] [doi: [10.1080/10911359.2013.835760](#)]
5. Arnett J. Emerging adulthood: the winding road from the late teens through the twenties. In: *Adolescence and Emerging Adulthood*. Oxford, UK: Oxford Scholarship; Jan 2019.
6. MacKinnon N, Colman I. Factors associated with suicidal thought and help-seeking behaviour in transition-aged youth versus adults. *Can J Psychiatry* 2016 Dec;61(12):789-796 [FREE Full text] [doi: [10.1177/0706743716667417](#)] [Medline: [27578771](#)]
7. Pryce J, Napolitano L, Samuels GM. Transition to Adulthood of Former Foster Youth: Multilevel Challenges to the Help-Seeking Process. *Emerging Adulthood* 2017 Jan 17;5(5):311-321. [doi: [10.1177/2167696816685231](#)]
8. Dunley P, Papadopoulos A. Why is it so hard to get help? Barriers to help-seeking in postsecondary students struggling with mental health issues: a scoping review. *Int J Ment Health Addiction* 2019 Feb 22;17(3):699-715 [FREE Full text] [doi: [10.1007/s11469-018-0029-z](#)]



9. Li W, Dorstyn DS, Denson LA. Predictors of mental health service use by young adults: a systematic review. *Psychiatr Serv* 2016 Sep 1;67(9):946-956. [doi: [10.1176/appi.ps.201500280](https://doi.org/10.1176/appi.ps.201500280)] [Medline: [27079988](https://pubmed.ncbi.nlm.nih.gov/27079988/)]
10. Salaheddin K, Mason B. Identifying barriers to mental health help-seeking among young adults in the UK: a cross-sectional survey. *Br J Gen Pract* 2016 Oct;66(651):e686-e692. [doi: [10.3399/bjgp16X687313](https://doi.org/10.3399/bjgp16X687313)] [Medline: [27688518](https://pubmed.ncbi.nlm.nih.gov/27688518/)]
11. Shandley K, Austin D, Klein B, Kyrios M. An evaluation of 'reach out central': an online gaming program for supporting the mental health of young people. *Health Educ Res* 2010 Aug;25(4):563-574. [doi: [10.1093/her/cyq002](https://doi.org/10.1093/her/cyq002)] [Medline: [20150530](https://pubmed.ncbi.nlm.nih.gov/20150530/)]
12. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry* 2010 Dec 30;10:113 [FREE Full text] [doi: [10.1186/1471-244X-10-113](https://doi.org/10.1186/1471-244X-10-113)] [Medline: [21192795](https://pubmed.ncbi.nlm.nih.gov/21192795/)]
13. Rickwood D, Thomas K. Conceptual measurement framework for help-seeking for mental health problems. *Psychol Res Behav Manag* 2012;5:173-183 [FREE Full text] [doi: [10.2147/PRBM.S38707](https://doi.org/10.2147/PRBM.S38707)] [Medline: [23248576](https://pubmed.ncbi.nlm.nih.gov/23248576/)]
14. Vloet MA, Davidson S, Cappelli M. 'We suffer from being lost': formulating policies to reclaim youth in mental health transitions. *Healthc Q* 2011;14(Spec No 2):32-38. [doi: [10.12927/hcq.2011.22361](https://doi.org/10.12927/hcq.2011.22361)] [Medline: [24956424](https://pubmed.ncbi.nlm.nih.gov/24956424/)]
15. Eisenberg D, Hunt J, Speer N. Help seeking for mental health on college campuses: review of evidence and next steps for research and practice. *Harv Rev Psychiatry* 2012;20(4):222-232. [doi: [10.3109/10673229.2012.712839](https://doi.org/10.3109/10673229.2012.712839)] [Medline: [22894731](https://pubmed.ncbi.nlm.nih.gov/22894731/)]
16. Golberstein E, Eisenberg D, Gollust SE. Perceived stigma and help-seeking behavior: longitudinal evidence from the healthy minds study. *Psychiatr Serv* 2009 Sep;60(9):1254-1256 [FREE Full text] [doi: [10.1176/ps.2009.60.9.1254](https://doi.org/10.1176/ps.2009.60.9.1254)] [Medline: [19723742](https://pubmed.ncbi.nlm.nih.gov/19723742/)]
17. Rickwood D, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. *Aus e-J Adv Mental Health* 2014 Dec 17;4(3):218-251 [FREE Full text] [doi: [10.5172/jamh.4.3.218](https://doi.org/10.5172/jamh.4.3.218)]
18. Mitchell C, McMillan B, Hagan T. Mental health help-seeking behaviours in young adults. *Br J Gen Pract* 2017 Jan;67(654):8-9. [doi: [10.3399/bjgp17X688453](https://doi.org/10.3399/bjgp17X688453)] [Medline: [28034926](https://pubmed.ncbi.nlm.nih.gov/28034926/)]
19. Downs MF, Eisenberg D. Help seeking and treatment use among suicidal college students. *J Am Coll Health* 2012;60(2):104-114. [doi: [10.1080/07448481.2011.619611](https://doi.org/10.1080/07448481.2011.619611)] [Medline: [22316407](https://pubmed.ncbi.nlm.nih.gov/22316407/)]
20. Findlay LC, Sunderland A. Professional and informal mental health support reported by Canadians aged 15 to 24. *Health Rep* 2014 Dec;25(12):3-11 [FREE Full text] [Medline: [25517935](https://pubmed.ncbi.nlm.nih.gov/25517935/)]
21. Reavley NJ, Yap MB, Wright A, Jorm AF. Actions taken by young people to deal with mental disorders: findings from an Australian national survey of youth. *Early Interv Psychiatry* 2011 Nov;5(4):335-342. [doi: [10.1111/j.1751-7893.2011.00292.x](https://doi.org/10.1111/j.1751-7893.2011.00292.x)] [Medline: [21951823](https://pubmed.ncbi.nlm.nih.gov/21951823/)]
22. Ciarrochi JV, Deane FP. Emotional competence and willingness to seek help from professional and nonprofessional sources. *Br J Guid Counsell* 2001;29(2):233-246 [FREE Full text] [doi: [10.1080/03069880020047157](https://doi.org/10.1080/03069880020047157)]
23. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychol Med* 2015 Jan;45(1):11-27. [doi: [10.1017/S0033291714000129](https://doi.org/10.1017/S0033291714000129)] [Medline: [24569086](https://pubmed.ncbi.nlm.nih.gov/24569086/)]
24. Wiljer D, Abi-Jaoude A, Johnson A, Ferguson G, Sanches M, Levinson A, et al. Enhancing self-efficacy for help-seeking among transition-aged youth in postsecondary settings with mental health and/or substance use concerns, using crowd-sourced online and mobile technologies: the thought spot protocol. *JMIR Res Protoc* 2016 Nov 4;5(4):e201 [FREE Full text] [doi: [10.2196/resprot.6446](https://doi.org/10.2196/resprot.6446)] [Medline: [27815232](https://pubmed.ncbi.nlm.nih.gov/27815232/)]
25. Kaur A, Isaranuwatthai W, Jaffer A, Ferguson G, Abi-Jaoude A, Johnson A, et al. A web- and mobile-based map of mental health resources for postsecondary students (Thought Spot): protocol for an economic evaluation. *JMIR Res Protoc* 2018 Mar 29;7(3):e83 [FREE Full text] [doi: [10.2196/resprot.9478](https://doi.org/10.2196/resprot.9478)] [Medline: [29599106](https://pubmed.ncbi.nlm.nih.gov/29599106/)]
26. VanHeerwaarden N, Ferguson G, Abi-Jaoude A, Johnson A, Hollenberg E, Chaim G, et al. The optimization of an ehealth solution (thought spot) with transition-aged youth in postsecondary settings: participatory design research. *J Med Internet Res* 2018 Mar 6;20(3):e79 [FREE Full text] [doi: [10.2196/jmir.8102](https://doi.org/10.2196/jmir.8102)] [Medline: [29510970](https://pubmed.ncbi.nlm.nih.gov/29510970/)]
27. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101 [FREE Full text] [doi: [10.1191/1478088706qp0630a](https://doi.org/10.1191/1478088706qp0630a)]
28. Kutcher S, Wei Y, Coniglio C. Mental health literacy: past, present, and future. *Can J Psychiatry* 2016 Mar;61(3):154-158 [FREE Full text] [doi: [10.1177/0706743715616609](https://doi.org/10.1177/0706743715616609)] [Medline: [27254090](https://pubmed.ncbi.nlm.nih.gov/27254090/)]
29. Headspace - Meditation and Sleep Made Simple. 2019. URL: <https://www.headspace.com/> [accessed 2020-02-01]
30. Meditation Studio App. URL: <https://meditationstudioapp.com/> [accessed 2020-02-01]
31. 7 Cups. 2020. URL: <https://www.7cups.com/> [accessed 2020-02-01]
32. Buddhify. 2020. URL: <https://buddhify.com/> [accessed 2020-02-01]
33. Bush NE, Smolenski DJ, Denneson LM, Williams HB, Thomas EK, Dobscha SK. A virtual hope box: randomized controlled trial of a smartphone app for emotional regulation and coping with distress. *Psychiatr Serv* 2017 Apr 1;68(4):330-336. [doi: [10.1176/appi.ps.201600283](https://doi.org/10.1176/appi.ps.201600283)] [Medline: [27842473](https://pubmed.ncbi.nlm.nih.gov/27842473/)]
34. Vanheusden K, Mulder CL, van der Ende J, van Lenthe FJ, Mackenbach JP, Verhulst FC. Young adults face major barriers to seeking help from mental health services. *Patient Educ Couns* 2008 Oct;73(1):97-104. [doi: [10.1016/j.pec.2008.05.006](https://doi.org/10.1016/j.pec.2008.05.006)] [Medline: [18584997](https://pubmed.ncbi.nlm.nih.gov/18584997/)]



35. O'Connor PJ, Martin B, Weeks CS, Ong L. Factors that influence young people's mental health help-seeking behaviour: a study based on the health belief model. *J Adv Nurs* 2014 Nov;70(11):2577-2587. [doi: [10.1111/jan.12423](https://doi.org/10.1111/jan.12423)] [Medline: [24720449](https://pubmed.ncbi.nlm.nih.gov/24720449/)]
36. Bergeron E, Poirier L, Fournier L, Roberge P, Barrette G. Determinants of service use among young Canadians with mental disorders. *Can J Psychiatry* 2005 Sep;50(10):629-636. [doi: [10.1177/070674370505001009](https://doi.org/10.1177/070674370505001009)] [Medline: [16276854](https://pubmed.ncbi.nlm.nih.gov/16276854/)]
37. Wetterlin FM, Mar MY, Neilson EK, Werker GR, Krausz M. eMental health experiences and expectations: a survey of youths' web-based resource preferences in Canada. *J Med Internet Res* 2014 Dec 17;16(12):e293 [FREE Full text] [doi: [10.2196/jmir.3526](https://doi.org/10.2196/jmir.3526)] [Medline: [25519847](https://pubmed.ncbi.nlm.nih.gov/25519847/)]
38. Morahan-Martin J, Anderson CD. Information and Misinformation Online: Recommendations for Facilitating Accurate Mental Health Information Retrieval and Evaluation. 2000 Oct;3(5):731-746 [FREE Full text] [doi: [10.1089/10949310050191737](https://doi.org/10.1089/10949310050191737)]
39. Pretorius C, Chambers D, Cowan B, Coyle D. Young people seeking help online for mental health: cross-sectional survey study. *JMIR Ment Health* 2019 Aug 26;6(8):e13524 [FREE Full text] [doi: [10.2196/13524](https://doi.org/10.2196/13524)] [Medline: [31452519](https://pubmed.ncbi.nlm.nih.gov/31452519/)]
40. Yorgason JB, Linville D, Zitzman B. Mental health among college students: do those who need services know about and use them? *J Am Coll Health* 2008;57(2):173-181. [doi: [10.3200/JACH.57.2.173-182](https://doi.org/10.3200/JACH.57.2.173-182)] [Medline: [18809534](https://pubmed.ncbi.nlm.nih.gov/18809534/)]
41. Andreason AR. A social marketing approach to changing mental health practices directed at youth and adolescents. *Health Mark Q* 2004;21(4):51-75. [doi: [10.1300/J026v21n04\\_04](https://doi.org/10.1300/J026v21n04_04)] [Medline: [15774369](https://pubmed.ncbi.nlm.nih.gov/15774369/)]
42. Yap MB, Reavley N, Jorm AF. Where would young people seek help for mental disorders and what stops them? Findings from an Australian national survey. *J Affect Disord* 2013 May;147(1-3):255-261. [doi: [10.1016/j.jad.2012.11.014](https://doi.org/10.1016/j.jad.2012.11.014)] [Medline: [23228570](https://pubmed.ncbi.nlm.nih.gov/23228570/)]
43. Gibson V. As More Students Seek Mental-health Care, They Face Long Waits – or Pay Out of Pocket – as Universities Struggle With Demand. *The Globe and Mail*. 2019. URL: <https://www.theglobeandmail.com/canada/article-faced-with-long-waits-for-mental-health-care-students-and/> [accessed 2020-03-01]
44. Kowalewski K, McLennan JD, McGrath PJ. A preliminary investigation of wait times for child and adolescent mental health services in Canada. *J Can Acad Child Adolesc Psychiatry* 2011 May;20(2):112-119 [FREE Full text] [Medline: [21541100](https://pubmed.ncbi.nlm.nih.gov/21541100/)]
45. Schraeder KE, Reid GJ. Why wait? The effect of wait-times on subsequent help-seeking among families looking for children's mental health services. *J Abnorm Child Psychol* 2015 Apr;43(3):553-565. [doi: [10.1007/s10802-014-9928-z](https://doi.org/10.1007/s10802-014-9928-z)] [Medline: [25178864](https://pubmed.ncbi.nlm.nih.gov/25178864/)]
46. Toulany A, Stukel TA, Kurdyak P, Fu L, Guttmann A. Association of primary care continuity with outcomes following transition to adult care for adolescents with severe mental illness. *JAMA Netw Open* 2019 Aug 2;2(8):e198415 [FREE Full text] [doi: [10.1001/jamanetworkopen.2019.8415](https://doi.org/10.1001/jamanetworkopen.2019.8415)] [Medline: [31373654](https://pubmed.ncbi.nlm.nih.gov/31373654/)]
47. Clatney L, Macdonald H, Shah S. Mental health care in the primary care setting: family physicians' perspectives. *Can Fam Physician* 2008 Jun;54(6):884-889. [Medline: [18556498](https://pubmed.ncbi.nlm.nih.gov/18556498/)]
48. Steele M, Shapiro J, Davidson B, Floyd G, Johnston J, Stretch N, et al. Survey comparing criteria used by rural and urban primary care physicians for referrals to child and adolescent psychiatrists and children's mental health agencies in Ontario. *J Can Acad Child Adolesc Psychiatry* 2010 Nov;19(4):284-289 [FREE Full text] [Medline: [21037919](https://pubmed.ncbi.nlm.nih.gov/21037919/)]
49. Steele M, Zayed R, Davidson B, Stretch N, Nadeau L, Fleisher W, et al. Referral Patterns and Training Needs in Psychiatry among Primary Care Physicians in Canadian Rural/Remote Areas. *J Can Acad Child Adolesc Psychiatry* 2012 May;21(2):111-123 [FREE Full text] [Medline: [22548108](https://pubmed.ncbi.nlm.nih.gov/22548108/)]
50. O'Brien D, Harvey K, Howse J, Reardon T, Creswell C. Barriers to managing child and adolescent mental health problems: a systematic review of primary care practitioners' perceptions. *Br J Gen Pract* 2016 Oct;66(651):e693-e707 [FREE Full text] [doi: [10.3399/bjgp16X687061](https://doi.org/10.3399/bjgp16X687061)] [Medline: [27621291](https://pubmed.ncbi.nlm.nih.gov/27621291/)]
51. Issue Brief - Workplace Mental Health. Mental Health Commission of Canada. 2016. URL: [https://www.mentalhealthcommission.ca/sites/default/files/2017-01/Issue\\_Brief\\_workplace\\_mental\\_health\\_eng.pdf](https://www.mentalhealthcommission.ca/sites/default/files/2017-01/Issue_Brief_workplace_mental_health_eng.pdf) [accessed 2020-03-01]
52. Lloyd C, King R, Chenoweth L. Social work, stress and burnout: a review. *J Ment Health* 2009 Jul 6;11(3):255-265 [FREE Full text] [doi: [10.1080/09638230020023642](https://doi.org/10.1080/09638230020023642)]
53. Sánchez-Moreno E, de La Fuente Roldán I, Gallardo-Peralta LP, Barrón López de Roda A. Burnout, Informal Social Support and Psychological Distress among Social Workers. *Br J Soc Work* 2014 Sep 02;45(8):2368-2386. [doi: [10.1093/bjsw/bcu084](https://doi.org/10.1093/bjsw/bcu084)]
54. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, et al. Prevalence of burnout among physicians: a systematic review. *J Am Med Assoc* 2018 Sep 18;320(11):1131-1150 [FREE Full text] [doi: [10.1001/jama.2018.12777](https://doi.org/10.1001/jama.2018.12777)] [Medline: [30326495](https://pubmed.ncbi.nlm.nih.gov/30326495/)]
55. Lee FJ, Stewart M, Brown JB. Stress, burnout, and strategies for reducing them: what's the situation among Canadian family physicians? *Can Fam Physician* 2008 Feb;54(2):234-235 [FREE Full text] [Medline: [18272641](https://pubmed.ncbi.nlm.nih.gov/18272641/)]
56. Raimo J, LaVine S, Spielmann K, Akerman M, Friedman KA, Katona K, et al. The correlation of stress in residency with future stress and burnout: a 10-year prospective cohort study. *J Grad Med Educ* 2018 Oct;10(5):524-531 [FREE Full text] [doi: [10.4300/JGME-D-18-00273.1](https://doi.org/10.4300/JGME-D-18-00273.1)] [Medline: [30386477](https://pubmed.ncbi.nlm.nih.gov/30386477/)]
57. Galdas PM, Cheater F, Marshall P. Men and health help-seeking behaviour: literature review. *J Adv Nurs* 2005 Mar;49(6):616-623. [doi: [10.1111/j.1365-2648.2004.03331.x](https://doi.org/10.1111/j.1365-2648.2004.03331.x)] [Medline: [15737222](https://pubmed.ncbi.nlm.nih.gov/15737222/)]

58. Dunne T, Bishop L, Avery S, Darcy S. A review of effective youth engagement strategies for mental health and substance use interventions. *J Adolesc Health* 2017 May;60(5):487-512. [doi: [10.1016/j.jadohealth.2016.11.019](https://doi.org/10.1016/j.jadohealth.2016.11.019)] [Medline: [28087268](https://pubmed.ncbi.nlm.nih.gov/28087268/)]

*Edited by G Eysenbach; submitted 09.03.20; peer-reviewed by A Kaur, C Pretorius; comments to author 21.04.20; revised version received 23.07.20; accepted 27.07.20; published 05.10.20.*

*Please cite as:*

Stunden C, Zasada J, VanHeerwaarden N, Hollenberg E, Abi-Jaoudé A, Chaim G, Cleverley K, Henderson J, Johnson A, Levinson A, Lo B, Robb J, Shi J, Voineskos A, Wiljer D

*Help-Seeking Behaviors of Transition-Aged Youth for Mental Health Concerns: Qualitative Study*

*J Med Internet Res* 2020;22(10):e18514

URL: <https://www.jmir.org/2020/10/e18514>

doi: [10.2196/18514](https://doi.org/10.2196/18514)

PMID: [33016882](https://pubmed.ncbi.nlm.nih.gov/33016882/)

©Chelsea Stunden, Julie Zasada, Nicole VanHeerwaarden, Elisa Hollenberg, Alexxa Abi-Jaoudé, Gloria Chaim, Kristin Cleverley, Joanna Henderson, Andrew Johnson, Andrea Levinson, Brian Lo, Janine Robb, Jenny Shi, Aristotle Voineskos, David Wiljer. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Psychometric Evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): Evaluation Study

Saskia Marion Kelders<sup>1,2</sup>, PhD; Hanneke Kip<sup>1</sup>, MSc; Japie Greeff<sup>2,3</sup>, MSc

<sup>1</sup>Center for eHealth and Wellbeing Research, Department of Psychology, Health, and Technology, University of Twente, Enschede, Netherlands

<sup>2</sup>Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa

<sup>3</sup>School of Computer Science and Information Systems, Faculty of Natural and Agricultural Sciences, North-West University, Vanderbijlpark, South Africa

**Corresponding Author:**

Saskia Marion Kelders, PhD

Center for eHealth and Wellbeing Research

Department of Psychology, Health, and Technology

University of Twente

PO Box 217

Enschede, 7500AE

Netherlands

Phone: 31 534896055

Email: [s.m.kelders@utwente.nl](mailto:s.m.kelders@utwente.nl)

## Abstract

**Background:** Engagement emerges as a predictor for the effectiveness of digital health interventions. However, a shared understanding of engagement is missing. Therefore, a new scale has been developed that proposes a clear definition and creates a tool to measure it. The TWente Engagement with Ehealth Technologies Scale (TWEETS) is based on a systematic review and interviews with engaged health app users. It defines engagement as a combination of behavior, cognition, and affect.

**Objective:** This paper aims to evaluate the psychometric properties of the TWEETS. In addition, a comparison is made with the experiential part of the Digital Behavior Change Intervention Engagement Scale (DBCI-ES-Ex), a scale that showed some issues in previous psychometric analyses.

**Methods:** In this study, 288 participants were asked to use any step counter app on their smartphones for 2 weeks. They completed online questionnaires at 4 time points: T0=baseline, T1=after 1 day, T2=1 week, and T3=2 weeks. At T0, demographics and personality (conscientiousness and intellect/imagination) were assessed; at T1-T3, engagement, involvement, enjoyment, subjective usage, and perceived behavior change were included as measures that are theoretically related to our definition of engagement. Analyses focused on internal consistency, reliability, and the convergent, divergent, and predictive validity of both engagement scales. Convergent validity was assessed by correlating the engagement scales with involvement, enjoyment, and subjective usage; divergent validity was assessed by correlating the engagement scales with personality; and predictive validity was assessed by regression analyses using engagement to predict perceived behavior change at later time points.

**Results:** The Cronbach alpha values of the TWEETS were .86, .86, and .87 on T1, T2, and T3, respectively. Exploratory factor analyses indicated that a 1-factor structure best fits the data. The TWEETS is moderately to strongly correlated with involvement and enjoyment (theoretically related to cognitive and affective engagement, respectively;  $P<.001$ ). Correlations between the TWEETS and frequency of use were nonsignificant or small, and differences between adherers and nonadherers on the TWEETS were significant ( $P<.001$ ). Correlations between personality and the TWEETS were nonsignificant. The TWEETS at T1 was predictive of perceived behavior change at T3, with an explained variance of 16%. The psychometric properties of the TWEETS and the DBCI-ES-Ex seemed comparable in some aspects (eg, internal consistency), and in other aspects, the TWEETS seemed somewhat superior (divergent and predictive validity).

**Conclusions:** The TWEETS performs quite well as an engagement measure with high internal consistency, reasonable test-retest reliability and convergent validity, good divergent validity, and reasonable predictive validity. As the psychometric quality of a scale is a reflection of how closely a scale matches the conceptualization of a concept, this paper is also an attempt to conceptualize and define engagement as a unique concept, providing a first step toward an acceptable standard of defining and measuring engagement.

**KEYWORDS**

engagement; attrition; eHealth; adoption; adherence; questionnaire; scale validation; digital health interventions

## Introduction

In eHealth, the use of technology to support health and wellbeing (a well-documented issue) is nonadherence. Especially within digital health interventions (DHIs), users often do not use the offered technology the way that the developers of the technology intended, which is referred to as nonadherence [1,2]. Examples are participants not completing all lessons within a mental health intervention, not using a step counter app on a daily basis, or not using all functions within a diabetes management system. Developers and researchers often assume that there is a dose-response relationship: for people who use a technology more and who adhere to the technology, the positive effects are greater. There is some evidence to support this assumption [3], but many DHIs do not show this relationship [4]. It has been argued that this has to do with the way adherence is conceptualized and measured (eg, with just the number of logins) [4,5]. However, more importantly, it has also been posited that the reasons why people use a technology might be more important in predicting effectiveness than the frequency or duration of use. Research shows that when people use DHIs because they feel involved or are able to identify themselves with the intervention, these DHIs are more likely to be effective for these people [6,7].

When looking at the reasons behind DHI use, the concept of engagement often emerges as a predictor for effectiveness [8-10]. In a broad sense, engagement is often seen as how involved or occupied someone is with something, and as something that is related to a positive outcome, such as effectiveness. For example, in health care, patient engagement has been shown to be related to better health outcomes [11], and in organizations, work engagement is related to better performance [12]. In relation to DHIs, engagement has also been posited as related to better outcomes in terms of more effective interventions [9,13,14].

It is important to note the conceptual difference between adherence and engagement. Adherence only says something about the objective usage of a DHI and indicates whether or not a participant uses the DHI as was intended by the developers [5]. In some, but not all, conceptualizations of engagement, the usage of a DHI is part of engagement [8]. Moreover, in all conceptualizations of engagement, it encompasses other aspects as well, which refer more to the reasons behind using a DHI. This implies that a participant might be engaged but not adherent. For example, someone believes that using an app is very helpful to reach one's goals (highly engaged) but feels that using the app once a week rather than the intended usage of once a day already helps to reach those goals (nonadherent). A participant may also be adherent but not engaged when, for example, using an app as intended because a researcher or therapist has asked them to, rather than due to a feeling that the app is personally beneficial. While the conceptual differences between engagement and adherence might be understood, within

the context of DHIs, a shared understanding and definition of engagement is missing.

In order to gain more insight into how engagement can be defined and conceptualized within the context of DHIs, a recent review looked at the definitions and components of engagement in various domains such as students, health, and digital engagement [8]. This review concluded that engagement contains emotional, behavioral, and cognitive components. Until now, this is not fully reflected in the definitions for DHI engagement, where in many cases, engagement is defined as only a behavioral component (for example, only as the usage of a DHI). This problem of a very narrow definition of engagement that does not encompass the full breadth of the concept is acknowledged by others. For example, a review on engagement in DHIs concluded that engagement should be characterized as the extent of usage and a subjective experience characterized by attention, interest, and affect [9]. However, that review also found that the majority of articles included only viewed engagement in behavioral terms, that is, as usage. Yet the definition chosen in that review also does not seem to fully encompass the emotional, behavioral, and cognitive components of engagement. The field of engagement to DHIs still needs a "clear, tailored, and domain-specific definition of the construct, which captures the associated emotional, behavioral, and cognitive components present within the given context" [8].

In order to gain insight into the content of these components, we interviewed self-proclaimed engaged users of health apps to study what they view as being engaged [13,15]. Behavioral engagement seemed to focus more on having a routine in using the technology and making it a part of daily life than about the frequency of using it. In this case, the quality of the behavior (being a routine and costing little effort) was more important than the frequency of the behavior (usage), which also shows the difference between adherence and engagement. Cognitive engagement with a DHI was found to be prominently related to the goals of the users. People have to think that the technology is useful for them and that it increases their ability to achieve their own goals. This is related to the importance of personal relevance for engagement, as stated in earlier studies on DHIs [6,7,16], and is also similar to the "attention and interest" part of the definition of engagement of Perski et al [9]. Lastly, the study found that affective engagement played a role for every participant but was less salient in most cases. Interestingly, affect was not only focused on feelings towards the technology itself but also on achieving goals, differing from the concept of user engagement [17]. Furthermore, although most participants mentioned positive affect when achieving goals, negative affect also seemed to play a role. Participants experienced frustration when not achieving their goals, and for some, this may enhance their motivation to go on. Finally, identity seemed to play a role in affective engagement: users need to be able to identify in some way with the technology and what it stands for.



A recent systematic review has identified various ways to assess (a form of) engagement to DHIs, such as using qualitative methods, self-report scales, ecological momentary assessments, and system usage data [14]. However, most of these methods assess only a specific form or component of engagement. The study posited that self-report scales might be the most accessible way to gain a more nuanced view of engagement in larger samples, especially when attempting to include the subjective experience. However, most existing self-report scales are created for measuring user engagement with technologies such as e-commerce websites or video games [14]. In these contexts, the goals of users are different than in DHIs. Engagement with DHIs seems to be needed at 2 levels: engagement with the technology itself, and engagement with the health behavior the technology aims to improve [10]. This is in contrast to, for example, user engagement with a shopping website, which involves only engagement with the technology and not with another offline behavior. This makes these existing scales less applicable to eHealth technologies such as DHIs. Another issue with many existing user-engagement scales is that they often include attributes that predict engagement but are not, in itself, part of engagement; this applies, for example, to aesthetic appeal or usability, which raises validity concerns [14].

Within the aforementioned systematic review paper on measuring engagement, 2 scales specifically targeted at engagement with eHealth technologies such as DHIs were identified [14]. The first scale, the eHealth engagement scale, showed adequate internal reliability and predictive validity on, amongst other things, retention of information and intentions to change [18]. However, conceptually, this scale seems to be more focused on engagement with health-related information than acting on that information, and thus on behavior change, as is our focus [9,13,14]. Moreover, this scale is an example of a scale that includes attributes that might predict engagement but are not part of the engagement itself. For example, the scale includes an assessment of the credibility of the technology, which may be a predictor of engagement but is not something that is seen as part of engagement itself. The second scale, the Digital Behavior Change Intervention Engagement Scale (DBCI-ES), was developed based on a broader definition of engagement, including both behavior and a subjective experience that might encompass cognitive and affective components of engagement. However, 2 validation studies showed that the psychometric properties of this scale are somewhat problematic [19,20]. The main issues seem to lie in the combination of the subjective and objective engagement measure, and in the discriminant and criterion validity. The experiential part of the scale (DBCI-ES-Ex) did show predictive validity, but this was not assessed on outcomes (such as changed drinking behavior) but on subsequent login. If engagement is seen as important for effectiveness, even more so than usage and adherence, then it should predict outcomes and not usage.

Based on our previous work on defining the concept of engagement, we developed another scale: the TWEnt Engagement with Ehealth and Technologies Scale (TWEETS) [15]. This scale was developed based on the interview study with engaged health app users that was discussed earlier [13], and on the systematic review of the concept of engagement in

different domains [8]. The scale employs a definition of engagement that incorporates behavior, cognition, and affect, as is common in other fields of research where engagement is used [8,21]. In the TWEETS, engaged behavior includes the existence of a routine in which individuals use the technology, low effort required to use the technology, and technology usage that is not fixed but may fluctuate to fit with the needs of the current moment. Cognitive engagement is related to the technology being able to support and motivate people in reaching their goals, such as the goal of improving one's wellbeing. Moreover, it entails that engaged users are willing to spend mental effort in using a DHI because it helps them achieve their goals, and they are intrinsically motivated. Affective engagement is related to emotions that people feel when seeing their progress in the DHI, or a lack thereof, and related to emotions such as enjoyment felt when using the technology itself. Lastly, it entails identity: engaged users seem to identify themselves in some way with the technology or with the goal of the technology.

In order for this new scale to be of use, its internal consistency and reliability should be examined. Moreover, assessment is needed for whether the scale measures a unique concept and is sufficiently different from other concepts such as involvement and adherence. As engagement is a relatively new concept without an existing gold-standard assessment, there is no criterion to relate it to, contrary to how the validation of the DBCI-ES was set up [19,20]. In that study, usage was seen as a criterion; however, since engagement is more than just usage [9,10,14,15], we feel that this does not do justice to the full concept of engagement. At the same time, there are concepts that, while different, are related and should correlate with a measure of engagement, and there are concepts that should reflect something different than engagement, and therefore should not be correlated. Specifically, engagement is conceptually related to concepts like involvement (how meaningful a product is for an individual) and enjoyment (the action or state of deriving gratification from an object) [7,22,23]. Therefore, it is expected that engagement correlates with these aspects, but that this correlation is moderate.

Another question that bears discussion is whether engagement is, and should be, related to the usage of a system. Based on the conceptualization of engagement as employed in the TWEETS, it is expected that engagement is not strongly related to the number of times individuals use a DHI. This is similar to what was found in the psychometric evaluation of the DBCI-ES [20]. Furthermore, engagement should reflect a more goal-oriented or intrinsically motivating reason for using a DHI, instead of, for example, using something because you are inclined to do so due to your personality [9,13]. Therefore, engagement should not be correlated to personality traits as conscientiousness and intellect/imagination. Lastly, it is important to assess whether the measure of engagement predicts future outcomes such as behavior change and clinical measures, as it is theorized that engagement influences outcomes and not the mere usage of a system.

The current study was carried out in the context of students using step counter apps. The main goal is to evaluate the psychometric properties of the TWEETS (internal consistency



and reliability; and convergent, divergent, and predictive validity). As a secondary exploratory objective, the psychometric properties of the TWEETS were compared to those of the experiential part of the DBCI-ES (DBCI-ES-Ex). We chose not to include the behavioral subscale due to the finding that this behavioral subscale may not be a valid indicator of (behavioral) engagement [20].

## Methods

### Design

For this study, participants were asked to use any step counter app on their smartphones for 2 weeks. Because the study aims at exploring the psychometric properties of the TWEETS and not, for example, how engaging a specific DHI is, the focus on students and a step counter app was deemed feasible and appropriate. Furthermore, we aimed at studying the TWEETS in an ecologically valid way and, therefore, opted to use existing step counter apps instead of, for example, a dedicated research app. Participants were asked to fill out an online questionnaire at 4 time points: T0=baseline, T1=after 1 day, T2=1 week, and T3=2 weeks. This study was approved by the ethical committee of the University of Twente (application number 18881).

At T0, demographics (gender, age, nationality, and employment status) and personality were assessed; at the other 3 time points, aspects related to the use of the step counter app were measured (engagement, involvement, enjoyment, usage, and perceived behavior change). This allowed us to perform the analyses needed to assess the internal consistency and reliability, and the convergent, divergent, and predictive validity of the engagement scales. Convergent validity refers to the degree to which measures of constructs that theoretically *should* be related, *are* related. To assess convergent validity, we chose to correlate the engagement scales to measures that are deemed to be related to behavioral, cognitive, and affective engagement. For behavioral engagement, we decided to use frequency of use and adherence as theoretically related concepts, as these are often posited as part of, or related to, engagement [9,14]. For cognitive engagement, involvement was chosen as the theoretically related concept, as involvement captures the personal relevance of an object (in this case, a DHI), which is often seen as an important part of engagement [6,7,9,16]. For affective engagement, enjoyment was chosen as the theoretically related concept because affective engagement is often seen as a reflection on how much users enjoy using the DHI [22,24].

Divergent validity refers to whether concepts that should theoretically not be related are not related. As engagement should reflect the position of an individual toward a DHI (and not a general trait or an individual), we chose to use personality traits as the concepts for divergent validity. Specifically, the personality traits of conscientiousness and intellect/imagination were used because they might reflect a somewhat similar disposition as engagement but are still theoretically unrelated. Conscientiousness, or being diligent, might lead participants to take the task of using the DHI very seriously and to, therefore, possibly use a DHI more [25]. However, this is a different reason for using a DHI and should thus not be related to engagement. Intellect/imagination—a person's preference for imaginative,

artistic, and intellectual activities [26]—could also lead to participants using a DHI more, as many DHIs also involve some degree of intellectual activity. However, as a personality trait, it should, theoretically, not be related to engagement with a specific DHI.

Predictive validity refers to the extent to which a score on a scale (in this case, engagement) predicts a score on a criterion measure in the future. As it is theorized that engagement influences outcomes like behavior change and clinical measures, we used the effectiveness of the DHI as this criterion measure.

### Participants

Participants were eligible for the study if they were willing to use a step counter app on their smartphones for 2 weeks. Recruitment took place via the participant pool of the University of Twente, where students receive credit for participating in research, and through the personal networks of the researchers. A total of 313 participants completed the baseline survey from December 2019 to April 2019; of the 313 participants, 288 filled out T1, 279 filled out T2, and 269 filled out T3. The 288 participants who filled out the TWEETS at least once (at T1, T2, or T3) were included in this study. Few data were missing of the included participants: T0 of all participants was complete; at T1, the number of participants that completed each measure ranged from 285 and 286 on the different measures (for example, 285 participants filled out the adherence measure while 286 filled out the TWEETS measure); at T2, the number of participants ranged from 270 and 274; and at T3, participants ranged from 259 and 265. Of the 288 included participants, most were female (228/288, 79%) and students (253/288, 88%). Most participants were German (202/288, 70%); 16% (46/288) were Dutch and 5% (14/288) were South African. In total, participants of 19 different nationalities were included. The mean age of participants was 22 (SD 7.1; range 18-70) years.

### Procedure

After participants signed up for the study and gave informed consent, they filled out the online T0 questionnaire with demographic information. After completing this questionnaire, they were asked to choose a step counter app to use for the following 2 weeks. While some apps were suggested (such as pre-installed step counter apps), participants were allowed to use any app that they preferred or already used. They were asked to open the app at least once a day, but it was suggested that it might be helpful to use the app more often, for example, during the day to check whether or not they were on track to reach their goal and then again at the end of the day to see how they did. Furthermore, participants were informed that they would receive 3 online follow-up surveys: after the first day, after a week, and after 2 weeks. It was explained that these surveys would cover their experience with using the step counter app. To gain credits for their education, participants had to complete all surveys.

### Materials

The following describes the different constructs—engagement, personality, involvement, enjoyment, usage, and perceived behavior change—that were measured in the questionnaires.

## Engagement

Engagement was assessed with the newly developed TWente Engagement with Ehealth Technologies Scale (TWEETS; [Table 1](#)) [15]. The TWEETS consists of 9 items on a 5-point Likert scale (strongly disagree=0, disagree=1, neutral=2, agree=3, strongly agree=4). Of the 9 items, 3 are aimed at assessing behavioral engagement, 3 on cognitive engagement, and 3 on affective engagement. The full scale is presented in [Table 1](#).

**Table 1.** The Twente Engagement with Ehealth Technologies Scale (TWEETS).

Item	Thinking about using [the technology] the last week, I feel that:	Construct
1	[this technology] is part of my daily routine	Behavior
2	[this technology] is easy to use <sup>a</sup>	Behavior
3	I'm able to use [this technology] as often as needed (to achieve my goals)	Behavior
4	[this technology] makes it easier for me to work on [my goal]	Cognition
5	[this technology] motivates me to [reach my goal]	Cognition
6	[this technology] helps me to get more insight into [my behavior relating to the goal]	Cognition
7	I enjoy using [this technology]	Affect
8	I enjoy seeing the progress I make in [this technology]	Affect
9	[This technology] fits me as a person	Affect

<sup>a</sup>Based on the outcomes of this study, this item was later changed to “[this technology] takes me little effort to use.”

As a second measure of engagement, the experiential subscale of the DBCI-ES (DBCI-ES-Ex) was used, consisting of 8 items on a 7-point answering scale with anchored end and middle points: not at all=1, moderately=3, extremely=7 [20]. The items were set up in the following manner: “Please answer the following questions with regard to your most recent use of the step counter app. How strongly did you experience the following?” The items were (1) Interest, (2) Intrigue, (3) Focus, (4) Inattention, (5) Distraction, (6) Enjoyment, (7) Annoyance, and (8) Pleasure, with items 4, 5, and 7 reverse-scored. In this study, Cronbach alpha values were .72, .80, and .84 on T1, T2, and T3, respectively.

## Personality

The personality of participants was assessed with the Mini-International Personality Item Pool (Mini-IPIP), a 20-item short form of the 50-item International Personality Item Pool–5-Factor Model measure [27]. For this study, the subscales on conscientiousness and intellect/imagination were used (both consisting of 10 items with a 7-point answering scale), which have been shown to have good psychometric properties [26]. In this study, personality was assessed at T0. The Cronbach alpha value was .78 for conscientiousness and .77 for intellect/imagination.

## Involvement

Involvement was assessed using the short version of the Personal Involvement Inventory (10 items, mean score 1-7; a higher score means more involvement), which has been shown to have good psychometric properties [23]. In this study, Cronbach alpha values were .90, .93, and .94 at T1, T2, and T3, respectively.

The scale was adapted for each time point: after one day of usage, items were posed as expectations (for example, “*I think using this app can become part of my daily routine*”); after 1 and 2 weeks, items were posed as looking back at using the app. Furthermore, the scale allows for adaption to the studied technology by adding the technology, the goal, and the behavior relating to the goal. For the current study, this was implemented as “this app” and “increasing the number of steps I take each day.”

## Enjoyment

Enjoyment was assessed using the enjoyment subscale of the Intrinsic Motivation Inventory (IMI, 7 items, mean score 1-7; a higher mean score means more enjoyment), which has been shown to have good psychometric properties [28]. The Cronbach alpha values in this study were .87, .91, and .92 at T1, T2, and T3, respectively.

## Usage

As participants were allowed to use any step counter app for reasons of feasibility and ecological validity, it was not possible to gather objective usage data. Therefore, subjective usage was assessed by asking participants whether or not they opened the app at least once a day, as was advised (*yes*=adherent; *no*=nonadherent), and how often they opened the app on a regular day that they opened it (frequency). In this case, the frequency of use and adherence are largely independent of each other because the frequency relates to a regular day when they adhered (ie, opened the app at least once). Only when a user did not use the app on any given day would the user be both nonadherent and have a use frequency of 0.

## Perceived Behavior Change

At T2 and T3, as an indication for the effectiveness of the app, we assessed whether the participants perceived behavior change due to using the step counter app. This was assessed using 1 item (“Do you feel that you have changed your behavior because of using the step counter app?”), with 3 answer options (“yes,” “maybe,” and “no”).

## Data Analyses

Analyses were performed using SPSS Statistics (version 25; IBM Corp). Analyses were conducted with the data of the

participants who completed the measures that were used in each specific analysis. Due to the limited amount of missing data, no imputation of missing values was required. Engagement measured by the TWEETS and by the DBCI-ES-Ex violated the assumption of normality at all time points by being left-skewed. However, due to the relatively large sample size and the robustness of analyses such as Pearson correlations and analyses of variance (ANOVAs), parametric tests were used [29-31]. Furthermore, we performed exploratory nonparametric analyses on our data to see whether the results from these analyses would lead to different conclusions. This was not the case, and therefore, we report only the parametric analyses in this paper.

Internal consistency of the TWEETS and the DBCI-ES-Ex was assessed using the Cronbach alpha value of each scale at T1, T2, and T3. A value of .7 was seen as the absolute minimum, while .8-.9 was seen as good internal consistency [32,33]. Exploratory factor analysis for the TWEETS was performed using SPSS' Maximum Likelihood (ML) method with oblique rotation (SPSS' direct oblimin) for all time points separately [34]. The Scree test (by examining the scree plots) combined with an assessment of eigenvalues was used to determine the number of factors [34]. Test-retest reliability for the TWEETS and DBCI-ES-Ex were assessed by examining Pearson correlations between the TWEETS scores on T1, T2, and T3, and by examining Pearson correlations of the DCBI-E scores on T1, T2, and T3. Although no standards exist for a minimum acceptable value for a test-retest reliability estimate [35], a correlation coefficient does provide information on the stability or variation of a scale over time. Convergent validity was assessed by calculating Pearson correlations between the TWEETS, DBCI-ES-Ex, involvement, enjoyment, and use frequency at T1, T2, and T3. Moreover, differences in the TWEETS and DBCI-ES-Ex between participants that adhered and did not adhere to the app were calculated using a one-way ANOVA. Divergent validity was assessed by calculating

Pearson correlations between the TWEETS and personality (conscientiousness and intellect/imagination) and between the DBCI-ES-Ex and personality (conscientiousness and intellect/imagination). Lastly, predictive validity was assessed by separate linear regression analyses using TWEETS and separate linear regression analyses using DBCI-ES-Ex on T1 to predict perceived behavior change at T2 and T3, and TWEETS and DBCI-ES-Ex at T2 to predict perceived behavior change at T3.

## Results

### Internal Consistency and Reliability

Table 2 shows the mean values of the scores on the TWEETS and the DBCI-ES-Ex at the various time points. The Cronbach alpha values of the TWEETS were .86, .86, and 0.87 on T1, T2, and T3, respectively, indicating good internal consistency. Exploratory factor analyses of the TWEETS indicated that a 1-factor structure best fits the data and explained 41.4% of the variance, based on observed data on T1. Data from T2 and T3 also indicated a 1-factor structure with similarly explained variance at T2 (42.2%) and somewhat less explained variance at T3 (30.1%). Of the 9 items, 8 items loaded strongly ( $> 0.5$ ) on the 1 factor. The other item ("this technology is easy to use") loaded 0.35 - 0.39 on the factor on the different time points. This is still seen as acceptable to retain an item in a scale, as it is above the minimum threshold of 0.32 [30]. Furthermore, the Cronbach alpha values with this item removed showed minimal differences in internal consistency (.86, .87, and .88 on T1, T2, and T3, respectively). Pearson correlations of the TWEETS at different time points were all significant ( $P < .001$ ), with values of 0.58 (T1-T2), 0.61 (T1-T3), and 0.74 (T2-T3), showing moderate test-retest reliability. Pearson correlations of the DBCI-ES-Ex at different time points were all significant ( $P < .001$ ), with values of 0.70 (T1-T2), 0.67 (T1-T3), and 0.78 (T2-T3), showing relative stability over time.

**Table 2.** Descriptive statistics for the TWente Engagement with Ehealth Technologies Scale (TWEETS) and the experiential subscale of the Digital Behavior Change Intervention Engagement Scale (DBCI-ES-Ex) at different time points.

Scale	Time point 1, n; mean (SD)	Time point 2, n; mean (SD)	Time point 3, n; mean (SD)
TWEETS	286; 2.85 (0.63)	274; 2.65 (0.65)	259; 2.65 (0.65)
DBCI-ES-Ex	286; 4.64 (0.74)	273; 4.50 (0.86)	266; 4.49 (0.90)

### Convergent Validity

Pearson correlations between the TWEETS, DBCI-ES-Ex, involvement, enjoyment, and use frequency at the different time points are shown in Table 3. The 2 engagement scales showed moderate to strong correlations with each other. Both the TWEETS and DBCI-ES-Ex showed significant moderate to

strong correlations with involvement and enjoyment, which become stronger at later time points. The DBCI-ES-Ex scale consistently showed stronger correlations than the TWEETS with both involvement and enjoyment, with particularly strong correlations with enjoyment. Both scales show no or weak correlations with reported use frequency.

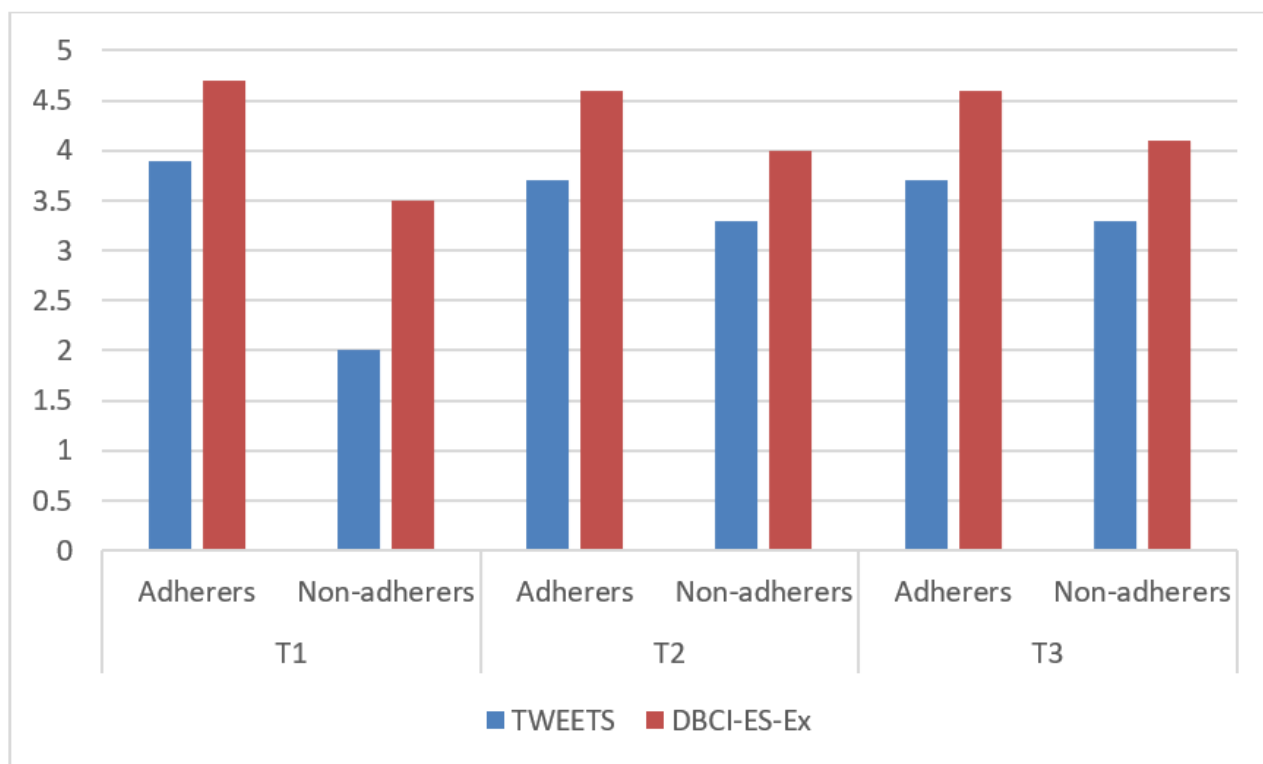
**Table 3.** Pearson correlations between the TWente Engagement with Ehealth Technologies Scale (TWEETS), the experiential subscale of the Digital Behavior Change Intervention Engagement Scale (DBCI-ES-Ex), involvement, enjoyment, and use frequency.

Scale & time point	Involvement	Enjoyment	Use frequency	DBCI-ES-Ex
<b>TWEETS</b>				
T1	0.57 <sup>a</sup>	0.60 <sup>a</sup>	0.08	0.57 <sup>a</sup>
T2	0.61 <sup>a</sup>	0.70 <sup>a</sup>	0.24 <sup>a</sup>	0.73 <sup>a</sup>
T3	0.67 <sup>a</sup>	0.77 <sup>a</sup>	0.04	0.80 <sup>a</sup>
<b>DBCI-ES-Ex</b>				
T1	0.54 <sup>a</sup>	0.69 <sup>a</sup>	0.02	N/A <sup>b</sup>
T2	0.63 <sup>a</sup>	0.81 <sup>a</sup>	0.26 <sup>a</sup>	N/A <sup>b</sup>
T3	0.70 <sup>a</sup>	0.82 <sup>a</sup>	0.03	N/A <sup>b</sup>

<sup>a</sup>Significant with  $P<.001$ .<sup>b</sup>N/A: not applicable.

Lastly, mean scores on the TWEETS between participants that adhered to the app (using it at least once a day) and those that did not adhere significantly differed at all time points ( $P<.001$ ), with adherers scoring higher than nonadherers (Figure 1). The

same was true for the DBCI-ES-Ex scale, with  $P=.028$  at T1 and  $P<.001$  at T2 and T3. However, the number of nonadherers was low, especially at T1 (T1: 2/285; T2: 36/270; T3: 46/262).

**Figure 1.** Engagement scores of adherers and nonadherers at the different time points. DBCI-ES-Ex: the experiential subscale of the Digital Behavior Change Intervention Engagement Scale; T: time point; TWEETS: TWente Engagement with Ehealth Technologies Scale.

### Divergent Validity

Pearson correlations between conscientiousness and the TWEETS at the different time points were nonsignificant. The same was true for intellect/imagination. For the DBCI-ES-Ex, the Pearson correlation with conscientiousness at T1 was weak but significant (0.17,  $P<.01$ ). Pearson correlations at the other time points and with intellect/imagination were nonsignificant.

### Predictive Validity

Results of the linear regression analyses using the TWEETS and DBCI-ES-Ex to predict perceived behavior change are shown in Table 4. These results show that both the TWEETS and DBCI-ES-Ex demonstrate predictive validity. As can be expected, the shorter the time between the measurement of engagement and the measurement of perceived behavior change, the more predictive engagement is, with engagement at T2 explaining around 24% of the variance of perceived behavior



change at T3, and engagement at T1 explaining 12-14% of the variance of perceived behavior change at T2. However, engagement measured at T1 was also predictive of perceived

behavior change at T3, with an explained variance of 16% when using the TWEETS and 9% using the DBCI-ES-Ex.

**Table 4.** Regression analyses with the TWente Engagement with Ehealth Technologies Scale (TWEETS) and the experiential subscale of the Digital Behavior Change Intervention Engagement Scale (DBCI-ES-Ex) to predict perceived behavior change.

Scale & time point	Perceived behavior change T2			Perceived behavior change T3		
	Beta	R <sup>2</sup>	F <sup>a</sup>	Beta	R <sup>2</sup>	F <sup>a</sup>
TWEETS T1	-.38	0.144	45.34	-.40	0.163	51.36
TWEETS T2	N/A <sup>b</sup>	N/A <sup>b</sup>	N/A <sup>b</sup>	-.49	0.242	84.04
DBCI-ES-Ex T1	-.35	0.120	35.84	-.30	0.087	25.70
DBCI-ES-Ex T2	N/A <sup>b</sup>	N/A <sup>b</sup>	N/A <sup>b</sup>	-.50	0.244	84.69

<sup>a</sup> $P < .001$

<sup>b</sup>N/A: not applicable.

## Discussion

### Principal Findings

This study set out to evaluate the psychometric properties of the new TWEETS. A secondary exploratory objective was to compare these psychometric properties to those of the experiential subscale of the DBCI-ES.

For the TWEETS, internal consistency was high, indicating that the different items of the scale all measure the same construct. In line with this, exploratory factor analyses showed the scale consists of 1 factor; we did not find the 3-factor structure that we expected, resulting from the behavioral, cognitive, and affective engagement components and their respective items in the scale. An explanation could be that engagement as a concept comprises only 1 component and not the proposed 3 components. However, because of the theoretical background of the concept, we would be hesitant to accept this explanation without further research. A more likely explanation might be that there is a theoretical overlap between the component; for example, enjoying seeing the progress that you make through an app (affective engagement) might also influence the extent to which you feel that the app motivates you to reach your goals (cognitive engagement). Also, behavioral engagement might be more of a consequence of cognitive and affective engagement (for example, because people feel that the technology helps them achieve their goals and is fun, they might establish a routine of using it), rather than being on the same level as these components. However, an explanation might also lie in the small number of items in the scale; 5 or more strong loading items per factor are seen as desirable [34], which indicates that we should have at least 15 items that would be retained in the scale, indicating an even higher number of a priori items. However, data showed that the 1-factor structure that was observed in the data was solid, with 8 items loading strongly and the last one above the minimum to retain an item in a factor [34]. Future research could explore whether the components can be measured as separate constructs, for example, by adding more items and examining whether this is desirable, as it would also increase the length of the scale and, therefore, its participant burden. Another option may be to retain the TWEETS as a short-form

engagement measure and develop separate measures for the components to be able to investigate more detailed patterns of engagement when in-depth analyses are needed, for example, to explore whether different people have, or different technologies induce, a different distribution of engagement over the components. What the 1-factor structure of the TWEETS does show is that the different components of engagement are indeed related and measure the same construct. This strengthens the assumption that engagement is a combination of behavior, cognition, and affect, and more than just usage behavior.

Test-retest reliability analyses showed moderate positive correlations between the first and later time points, and a strong positive correlation between the later time points. Although this may be interpreted as low test-retest reliability when looking at this first time point, this might not be surprising, as engagement after 1 day of usage might be initial engagement, which may be different from longer-term engagement. This is also reflected in different models on the process of engagement in which engagement may vary over time [8,17]. It seems reasonable to expect engagement to change after getting more acquainted with the app, and when actually trying to use the app for a longer period in daily life. As integrating technology in daily life is often one of the largest struggles in eHealth and specifically DHIs [36], being able to do so or not might be an obvious reason for a change in engagement. Taking these factors into account, it may be even more remarkable to see that participants' initial engagement, based on only 1 day of using a DHI, is still as related to engagement at a later stage. Additionally, this change in the level of engagement over time might also explain the somewhat differing findings on convergent and divergent validity of the engagement scales at different time points. As is seen in other studies, different aspects might be more important and more related to engagement in the different stages of the process of engagement [17].

Related to this, the TWEETS showed predictive validity: engagement at an earlier time point was able to predict perceived behavior change at a later time. Even the engagement measure after the first day of using a technology showed this capability, which opens up many interesting possibilities [13]; when we are able to assess early on in an intervention whether or not we



expect it is going to be effective for an individual, we may be able to direct more support to those for whom it might not be effective. This can be, for example, in the form of personalized feedback or the suggestion of using a different intervention.

Analyses of divergent validity showed that the TWEETS was not correlated with personality factors. This strengthens the assumption that engagement is a separate construct and that this is reflected in this new scale. Analyses of convergent validity showed that the TWEETS is moderately to strongly correlated with involvement (which is seen as theoretically related to cognitive engagement) and enjoyment (which is seen as theoretically related to affective engagement) [24]. The strength of the correlations indicate that the concepts are related but should be viewed as separate. For behavioral engagement, we choose to relate the TWEETS to adherence and frequency of use, as these are aspects that are often seen as related to engagement [9,14,24], or even used to assess criterion validity [19]. Correlations between the engagement measure and frequency of use were nonsignificant or small, while differences between adherers and nonadherers on engagement were significant but small. This might indicate that engagement is related to usage, but much less so than is often assumed, again strengthening the multiple component definition of engagement. However, it should be noted that our usage measure was subjective, which sheds some doubt on the robustness of this measure. This could provide an alternative explanation of the weak correlation between frequency of use and engagement. Future research could investigate whether objectively measured frequency of use should be seen as a measure of convergent validity or whether it could be more appropriate to see it as a measure to assess divergent validity.

The exploratory comparison of the psychometric properties of both engagement scales—the TWEETS and the DBCI-ES-Ex—revealed that both scales show similar properties. As could be expected, the scales are significantly correlated to each other. Interestingly, this correlation is moderate at the first time point but strong at the last time point. This might indicate that they vary on how much they take the first impression that a DHI makes into account. A few notable differences between both scales are that the DBCI-ES-Ex shows quite a strong correlation to enjoyment, a bit stronger than the TWEETS, and that DBCI-ES-Ex did not fully reflect divergent validity. This might indicate that the DBCI-ES-Ex captures engagement not as a unique concept but more like enjoyment. Additionally, the TWEETS shows more predictive validity between the first and the last time point compared to the DBCI-ES-Ex. Overall, the TWEETS seems to reflect engagement better as a multifaceted and unique concept that has predictive value for the outcomes of an intervention.

## Limitations

This study has a few limitations that need to be discussed. First, the study was done mostly with students who received credits for participating in the study and thus for using the step counter app, which might have influenced their usage and engagement. However, as we were interested in the relationships between engagement and other measures, and not the level of engagement or usage, we feel that studying this target group was an

appropriate first step. Furthermore, we needed data on a broad spectrum of engagement, including low engagement. Therefore, including a target group that might not be completely intrinsically motivated to use a DHI was seen as appropriate. Future research should focus on measuring engagement in different populations and settings, for example, with lower educated participants or within the context of regular (mental) health care. It would be interesting to see whether different populations show different styles of engagement, such as whether lower educated participants might be more often affectively engaged, and whether this impacts the factor analysis of the scale.

Second, we used self-reported usage data and our variable for predictive validation was perceived behavior change and not actual behavior change. It would have been preferable to use objective usage data and actual behavior change data, but this was deemed infeasible for this exploratory study because participants could use any step counter app for reasons of ecological validity and feasibility. Therefore, it was not possible to objectively collect the usage data of all these different apps or the number of steps participants took every day. Also, to be able to measure actual behavior change in the average number of steps per day, a larger data collection period would have been necessary, which was beyond the scope of this study. However, as this was an exploratory study that provides a first step towards establishing whether the TWEETS is a useful measure of engagement in DHIs, this approach was deemed appropriate. Based on the promising findings of this study that engagement was able to predict perceived behavior change, the next step is to validate this finding in future studies using engagement to predict actual behavior change and other intervention outcomes. Also, the direction of change was not explicitly stated in the question. However, at multiple points in the study, it was stated that the goal of the app was to increase the number of steps taken per day. Moreover, participants could indicate in what way they changed their behavior. All participants that filled out that question indicated that they took more steps per day and not less steps. Therefore, we feel that the impact of not explicitly stating the direction of change was negligible.

A final limitation is that 1 item of the scale performed a bit less well than the others, although still above the minimum threshold to retain an item in a scale: the item “*this technology is easy to use*” should be revised in a way that better reflects the behavioral component of engagement and is less related to usability, which can be seen as a predictor of engagement. A different way to state the item that better reflects the quality of engagement behavior from our definition could be “*this technology takes me little effort to use.*” Nonetheless, the scale performed quite well on most criteria, indicating the potential of this new instrument.

## Conclusion

Overall, the TWEETS seemed to perform quite well as a 1-factor engagement measure: the scale showed high internal consistency, reasonable test-retest reliability and convergent validity, good divergent validity, and reasonable predictive validity. These properties seem comparable to—and, on some aspects, somewhat superior to—those of the experiential

subscale of the DBCI-ES. Further research is needed to replicate these findings in other target groups and eHealth technologies; however, the TWEETS seems to be a valuable addition to the toolbox of eHealth researchers, allowing them, amongst others, to be able to better investigate the relationship between engagement and effectiveness.

Lastly, the psychometric quality of a scale is, of course, a reflection of how closely a scale matches the conceptualization

of the concept. Therefore, this paper can also be seen as an attempt to conceptualize and define engagement as a unique concept, different from, for example, adherence and user experience. We have argued for a definition of engagement that entails behavioral, cognitive, and affective components and, through the analyses in the paper, have provided arguments in favor of this definition. This paper provides a first step toward an acceptable standard of defining and measuring engagement.

## Acknowledgments

This work is supported by the Netherlands Organisation for Scientific Research (NWO) through the Innovational Research Incentives Scheme - VENI-project of the first author (The Personalization Approach: Using engagement as a mediator for the effectiveness of eMental Health interventions; grant number 451-17-031).

## Conflicts of Interest

None declared.

## References

1. Christensen H, Griffiths KM, Farrer L. Adherence in Internet Interventions for Anxiety and Depression. *J Med Internet Res* 2009 Apr 24;11(2):e13. [doi: [10.2196/jmir.1194](https://doi.org/10.2196/jmir.1194)]
2. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive System Design Does Matter: a Systematic Review of Adherence to Web-based Interventions. *J Med Internet Res* 2012 Nov 14;14(6):e152. [doi: [10.2196/jmir.2104](https://doi.org/10.2196/jmir.2104)]
3. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A Systematic Review of the Impact of Adherence on the Effectiveness of e-Therapies. *J Med Internet Res* 2011 Aug 05;13(3):e52. [doi: [10.2196/jmir.1772](https://doi.org/10.2196/jmir.1772)]
4. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the Dose-Response Relationship Between Usage and Outcome in an Online Intervention for Depression: Randomized Controlled Trial. *J Med Internet Res* 2013 Oct 17;15(10):e231. [doi: [10.2196/jmir.2771](https://doi.org/10.2196/jmir.2771)]
5. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. *J Med Internet Res* 2017 Dec 06;19(12):e402. [doi: [10.2196/jmir.8578](https://doi.org/10.2196/jmir.8578)]
6. Donkin L, Glozier N. Motivators and Motivations to Persist With Online Psychological Interventions: A Qualitative Study of Treatment Completers. *J Med Internet Res* 2012 Jun 22;14(3):e91. [doi: [10.2196/jmir.2100](https://doi.org/10.2196/jmir.2100)]
7. Kelders S. M. 2015 Presented at: International Conference on Persuasive Technology; 2015; Chicago. [doi: [10.1007/978-3-319-20306-5\\_1](https://doi.org/10.1007/978-3-319-20306-5_1)]
8. Kelders SM, van Zyl LE, Ludden GDS. The Concept and Components of Engagement in Different Domains Applied to eHealth: A Systematic Scoping Review. *Front. Psychol* 2020 May 27;11:926. [doi: [10.3389/fpsyg.2020.00926](https://doi.org/10.3389/fpsyg.2020.00926)]
9. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Behav. Med. Pract. Policy Res* 2016 Dec 13;7(2):254-267. [doi: [10.1007/s13142-016-0453-1](https://doi.org/10.1007/s13142-016-0453-1)]
10. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. *American Journal of Preventive Medicine* 2016 Nov;51(5):833-842. [doi: [10.1016/j.amepre.2016.06.015](https://doi.org/10.1016/j.amepre.2016.06.015)]
11. Hibbard JH, Greene J. What The Evidence Shows About Patient Activation: Better Health Outcomes And Care Experiences; Fewer Data On Costs. *Health Affairs* 2013 Feb;32(2):207-214. [doi: [10.1377/hlthaff.2012.1061](https://doi.org/10.1377/hlthaff.2012.1061)]
12. Cacciamani G, Aron M, Gill I, Desai M. Oncological outcome according to attainment of pentapecta after robot assisted radical cystectomy among bladder cancer patients using KORARC database (730 multicenter robot radical cystectomy database) *BJU Int* 2020 Jul 8. doi: 10.1111/bju.15178. *BJU Int* 2020 Sep 01:15178. [doi: [10.1111/bju.15222](https://doi.org/10.1111/bju.15222)] [Medline: [32870582](https://pubmed.ncbi.nlm.nih.gov/32870582/)]
13. Kelders S. Design for Engagement of Online Positive Psychology Interventions. In: van Zyl L, Rothmann S, editors. *Positive Psychological Intervention Design and Protocols for Multi-Cultural Contexts*. Cham: Springer; 2019:297-313.
14. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies. *J Med Internet Res* 2018 Nov 16;20(11):e292. [doi: [10.2196/jmir.9397](https://doi.org/10.2196/jmir.9397)]
15. Kelders SM, Kip H. Development and initial validation of a scale to measure engagement with eHealth technologies. 2019 Presented at: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems; 2019; Glasgow. [doi: [10.1145/3290607.3312917](https://doi.org/10.1145/3290607.3312917)]

16. Short CE, Rebar AL, Plotnikoff RC, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. *European Health Psychologist* 2015;17(1):32-38 [[FREE Full text](#)]
17. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. *J. Am. Soc. Inf. Sci* 2008 Apr;59(6):938-955. [doi: [10.1002/asi.20801](#)]
18. Lefebvre R. C., The assessment of user engagement with eHealth content: the eHealth engagement scale. *Journal of Computer-Mediated Communication* 2010;4(15):666-681. [doi: [10.1111/j.1083-6101.2009.01514.x](#)]
19. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". *Translational behavioral medicine* 2019 Mar 30;267-277. [doi: [10.1093/tbm/ibz039](#)] [Medline: [30927357](#)]
20. Perski O, Lumsden J, Garnett C, Blandford A, West R, Michie S. Assessing the Psychometric Properties of the Digital Behavior Change Intervention Engagement Scale in Users of an App for Reducing Alcohol Consumption: Evaluation Study. *J Med Internet Res* 2019 Nov 20;21(11):e16197. [doi: [10.2196/16197](#)]
21. Graffigna G. Is a Transdisciplinary Theory of Engagement in Organized Settings Possible? A Concept Analysis of the Literature on Employee Engagement, Consumer Engagement and Patient Engagement. *Front. Psychol* 2017 Jul 06;8:872. [doi: [10.3389/fpsyg.2017.00872](#)]
22. Crutzen R, van 't Riet J, Short CE. Enjoyment: A Conceptual Exploration and Overview of Experimental Evidence in the Context of Games for Health. *Games for Health Journal* 2016 Feb;5(1):15-20. [doi: [10.1089/g4h.2015.0059](#)]
23. Zaichkowsky JL. Measuring the Involvement Construct. *J CONSUM RES* 1985 Dec;12(3):341. [doi: [10.1086/208520](#)]
24. Kelders SM, Sommers-Spijkerman M, Goldberg J. Investigating the Direct Impact of a Gamified Versus Nongamified Well-Being Intervention: An Exploratory Experiment. *J Med Internet Res* 2018 Jul 26;20(7):e247. [doi: [10.2196/jmir.9923](#)]
25. Forbes L, Gutierrez D, Johnson SK. Investigating Adherence to an Online Introductory Mindfulness Program. *Mindfulness* 2017 Jul 14;9(1):271-282. [doi: [10.1007/s12671-017-0772-4](#)]
26. Donnellan MB, Oswald FL, Baird BM, Lucas RE. The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of Personality. *Psychological Assessment* 2006;18(2):192-203. [doi: [10.1037/1040-3590.18.2.192](#)]
27. Goldberg L. R., A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe* 1999;7(1):7-28 [[FREE Full text](#)]
28. McAuley E, Duncan T, Tammen VV. Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport* 1989 Mar;60(1):48-58. [doi: [10.1080/02701367.1989.10607413](#)]
29. Bishara AJ, Hittner JB. Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods* 2012 Sep;17(3):399-417. [doi: [10.1037/a0028087](#)]
30. Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *Int J Endocrinol Metab* 2012 Dec 01;10(2):486-489. [doi: [10.5812/ijem.3505](#)]
31. Schmidt AF, Finan C. Linear regression and the normality assumption. *Journal of Clinical Epidemiology* 2018 Jun;98:146-151. [doi: [10.1016/j.jclinepi.2017.12.006](#)]
32. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int. J. Medical Education* 2011 Jun 27;2:53-55. [doi: [10.5116/ijme.4dfb.8dfd](#)]
33. Hinkin TR. A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods* 2016 Jun 29;19(1):104-121. [doi: [10.1177/109442819800100106](#)]
34. Osborne JW, Costello AB, Kellow JT. Best practices in exploratory factor analysis. Louisville, KY: CreateSpace Independent Publishing Platform 2008:86-99. [doi: [10.4135/9781412995627.d8](#)]
35. Crocker L, Algina J. Introduction to classical and modern test theory. Orlando, Florida: Holt, Rinehart and Winston; 1986.
36. Sieverink F, Kelders S, Braakman-Jansen A, van Gemert-Pijnen J. Evaluating the implementation of a personal health record for chronic primary and secondary care: a mixed methods approach. *BMC Med Inform Decis Mak* 2019 Nov 27;19(1):241. [doi: [10.1186/s12911-019-0969-7](#)]

## Abbreviations

**ANOVA:** analysis of variance

**DBCI-ES:** Digital Behavior Change Intervention Engagement Scale

**DBCI-ES-Ex:** experiential subscale of the Digital Behavior Change Intervention Engagement Scale

**DHI:** digital health interventions

**TWEETS:** TWente Engagement with Ehealth Technologies Scale

*Edited by G Eysenbach; submitted 10.01.20; peer-reviewed by C Yeager, M Schulte, A Radomski; comments to author 19.03.20; revised version received 01.07.20; accepted 26.07.20; published 09.10.20.*

*Please cite as:*

*Kelders SM, Kip H, Greeff J*

*Psychometric Evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): Evaluation Study*

*J Med Internet Res 2020;22(10):e17757*

*URL: <https://www.jmir.org/2020/10/e17757>*

*doi: [10.2196/17757](https://doi.org/10.2196/17757)*

*PMID: [33021487](https://pubmed.ncbi.nlm.nih.gov/33021487/)*

©Saskia Marion Kelders, Hanneke Kip, Japie Greeff. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.09.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Developing a Process for the Analysis of User Journeys and the Prediction of Dropout in Digital Health Interventions: Machine Learning Approach

Vincent Bremer<sup>1\*</sup>, MA; Philip I Chow<sup>2\*</sup>, PhD; Burkhardt Funk<sup>1\*</sup>, PhD; Frances P Thorndike<sup>2\*</sup>, PhD; Lee M Ritterband<sup>2\*</sup>, PhD

<sup>1</sup>Institute of Information Systems, Leuphana University Lüneburg, Lüneburg, Germany

<sup>2</sup>Center for Behavioral Health & Technology, University of Virginia School of Medicine, Charlottesville, VA, United States

\* all authors contributed equally

**Corresponding Author:**

Vincent Bremer, MA

Institute of Information Systems

Leuphana University Lüneburg

C4.320

Lüneburg, 21335

Germany

Phone: 49 41316771157

Email: [vincent.bremer@leuphana.de](mailto:vincent.bremer@leuphana.de)

## Abstract

**Background:** User dropout is a widespread concern in the delivery and evaluation of digital (ie, web and mobile apps) health interventions. Researchers have yet to fully realize the potential of the large amount of data generated by these technology-based programs. Of particular interest is the ability to predict who will drop out of an intervention. This may be possible through the analysis of user journey data—self-reported as well as system-generated data—produced by the path (or journey) an individual takes to navigate through a digital health intervention.

**Objective:** The purpose of this study is to provide a step-by-step process for the analysis of user journey data and eventually to predict dropout in the context of digital health interventions. The process is applied to data from an internet-based intervention for insomnia as a way to illustrate its use. The completion of the program is contingent upon completing 7 sequential cores, which include an initial tutorial core. Dropout is defined as not completing the seventh core.

**Methods:** Steps of user journey analysis, including data transformation, feature engineering, and statistical model analysis and evaluation, are presented. Dropouts were predicted based on data from 151 participants from a fully automated web-based program (Sleep Healthy Using the Internet) that delivers cognitive behavioral therapy for insomnia. Logistic regression with L1 and L2 regularization, support vector machines, and boosted decision trees were used and evaluated based on their predictive performance. Relevant features from the data are reported that predict user dropout.

**Results:** Accuracy of predicting dropout (area under the curve [AUC] values) varied depending on the program core and the machine learning technique. After model evaluation, boosted decision trees achieved AUC values ranging between 0.6 and 0.9. Additional handcrafted features, including time to complete certain steps of the intervention, time to get out of bed, and days since the last interaction with the system, contributed to the prediction performance.

**Conclusions:** The results support the feasibility and potential of analyzing user journey data to predict dropout. Theory-driven handcrafted features increased the prediction performance. The ability to predict dropout at an individual level could be used to enhance decision making for researchers and clinicians as well as inform dynamic intervention regimens.

(*J Med Internet Res* 2020;22(10):e17738) doi:[10.2196/17738](https://doi.org/10.2196/17738)

**KEYWORDS**

dropout; digital health; machine learning



## Introduction

The efficacy of digital (ie, internet, web, and mobile) behavioral interventions to improve a range of health-related outcomes has been well documented [1-3]. However, adherence to these interventions is a significant issue [4]. Intervention dropout, defined as a participant prematurely discontinuing a program, from internet-based treatments for psychological disorders typically varies between 30% and 50% [4-6]. However, the reason for such high dropout rates is still unclear [5], whereas longer treatment duration and user engagement appear to be associated with improved treatment outcomes and greater effectiveness of the digital intervention [7-10]. Furthermore, in a research setting, high dropout rates and, consequently, low exposure to digital content might affect the reported effects of a digital intervention and the validity of the results [11,12]. Although researchers have highlighted the need for a science of user attrition [13], there have been few advances in predicting dropout through advanced quantitative approaches in eHealth interventions [14]. In particular, previous work has identified hypothetical factors influencing attrition in eHealth programs, such as ease of leaving the intervention, unrealistic expectations on behalf of users, usability and interface issues, and amount of workload required to benefit from an intervention [13]. Such factors are likely to impact how a user ultimately engages with a program and could provide indicators for predictive factors but do little to advance predictive modeling of dropout when not applied in data-driven studies. Research suggests that an increased completion of modules in digital therapeutics increases treatment outcomes [15]. Identifying those patients that are likely to drop out of treatment and addressing the related issues can, thus, improve treatment outcomes and can be the basis of the development of micro interventions that target these high-risk participants to reengage them to complete the program [16]. Thus, predicting dropout on a participant level supports the decision making of experts in the target field and consequently leads to more personalized treatment strategies. In addition, inferential results can increase insight into the causes of attrition by revealing data-driven indicators. Participant-specific factors can help to identify individuals who benefit more from digital therapies compared with individuals for whom face-to-face treatment might be a better approach. To evaluate the possibility of predicting dropout in digital interventions and to shed light on some indicators of dropout, the aim of this study is to propose a process for user journey analysis to predict dropout from a digital intervention.

A wealth of data can be collected through the use of digital interventions. They often feature content that is administered over time as users complete tasks or components of the intervention, typically over several weeks or months [17-20]. Digital interventions also track and log different types of user interactions (eg, frequency of log-ins). These data provide a nuanced understanding of the usage behavior of participants over the course of an intervention [21]. Combined with

self-reported data, passively collected user data could be captured and used to provide deeper insight into how likely users are to drop out of an intervention on an individual level and lead to increased prediction performance.

A user journey is a sequence of interactions as an individual uses a digital intervention (ie, the path an individual takes to navigate through a program). Although user journeys are well known and established in the field of web-based marketing, to the best of our knowledge, its direct application to digital health interventions has not yet been examined. Web-based marketers leverage user journeys to collect information about an individual's behavior [22], often referred to as clickstream data analysis [23,24]. This increases the understanding of users' behavior by recognizing patterns in their sequence of actions. Thus, user journey analysis can reveal insight into an individual's behavior by enabling an analysis of data (eg, Ecological Momentary Assessment [EMA] or log data) that is not frequently used in the eHealth sphere [25].

There are several possible reasons why analysis of user journeys has not achieved prominence in digital health interventions. One obstacle lies in the analysis of large amounts of raw data. Analysis of user journeys often requires transformation of raw data, feature engineering, and the application of machine learning techniques, which can be a burdensome process [26] and is not a typical skill set of eHealth behavior researchers. Although user journeys have been used to predict different psychological factors such as mood, stress levels, or treatment outcomes and costs [25,27-31], to our knowledge, no work has provided steps to be taken to analyze raw user journey data and, at the same time, predict user dropout from a digital health intervention.

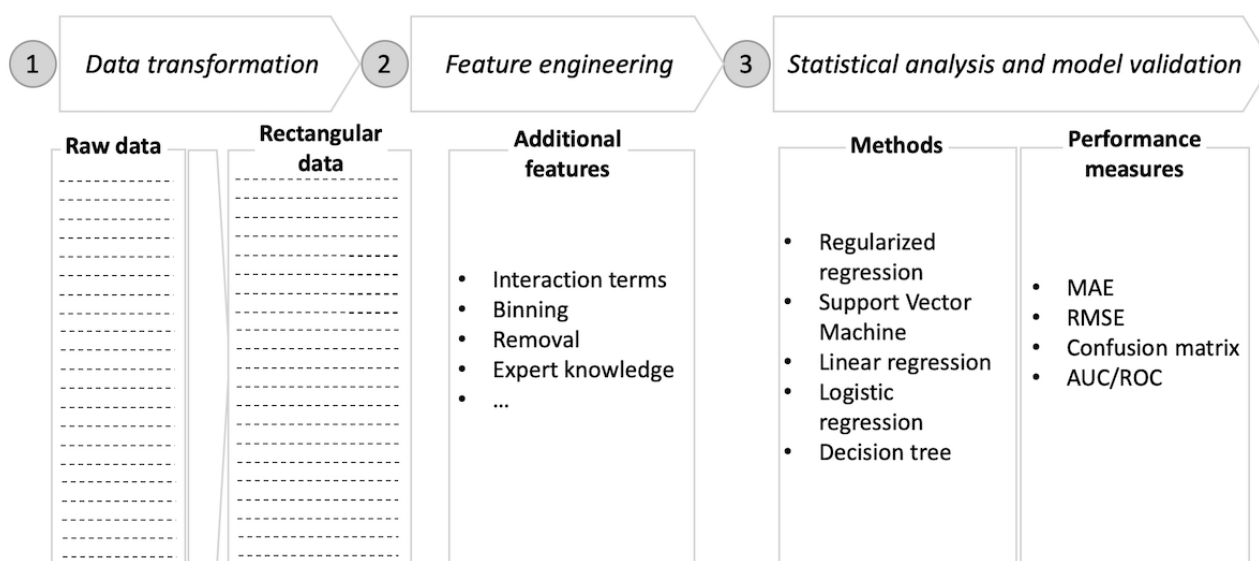
The overarching goal of this study is to establish and provide a step-by-step process that describes how to leverage user journeys to predict various behaviors (eg, dropout). This process involves several steps, including creating the basic data structure for handling user journeys, creating features that can add additional information to the existing raw data, and ultimately providing a framework for the statistical analysis. A technical implementation (R package) [32,33] of this process is provided for the research community. To demonstrate the application and potential utility of this process, we use it to predict user dropout in a randomized controlled trial of a fully automated cognitive behavior therapy intervention for insomnia (Sleep Healthy Using the Internet [SHUTi]) [34].

## Methods

### User Journey Process

The overarching steps of the user journey process are outlined in Figure 1. This process applies machine learning algorithms, specifically supervised learning, which is used when both input (eg, log-ins and mood symptoms) and output data (eg, dropout status) exist in the data set [35].

**Figure 1.** Process of analysis. AUC: area under the curve; MAE mean absolute error; ROC: receiver operating characteristics; RMSE: root mean square error.

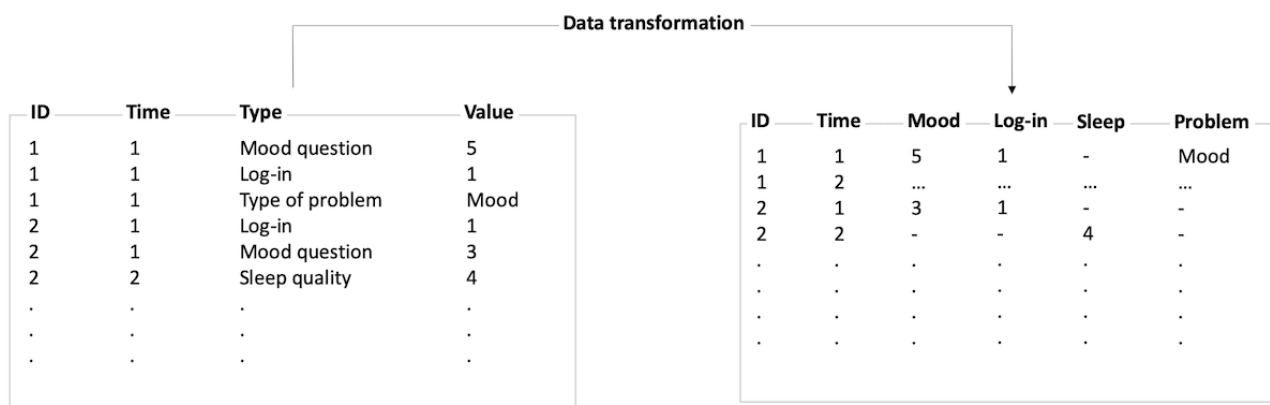


It is important for researchers to clearly define the outcome variable of interest. As dependent variables can take on different measurement scales (eg, discrete or continuous), defining the target variable has consequences for the choice of statistical models. When predicting discrete outcomes (ie, consisting of at least two discrete categories or labels), classification is often the appropriate approach. However, when predicting continuous outcome variables, the learning task is regression.

### Step One: Data Transformation

The first step in analyzing user journey data is to transform the raw data into a wide format, as can be seen in Figure 2. Thus, the transformed data are structured such that each row corresponds to a unique observation in *Time* for a particular user (*ID*).

**Figure 2.** Example of data transformation in the context of digital health interventions.



When transforming the raw data, it is important to specify the time window defining the time interval for which individual touch points are aggregated. The choice of the time window depends on the density of the observations in the raw data. For example, if a raw data set is composed of a few touch points over the course of a day, choosing a time window on a scale of days avoids sparseness of the transformed data matrix. In contrast, when predicting purchases in web-based marketing, for example, a large number of observations exist for each user on short timescales. Here, choosing a small window (eg, an hour) could be beneficial, as the resulting matrix will not be sparse and information loss is minimal. In an internet-based intervention, however, it is not unusual for self-reported data to be collected as little as once a day, with a user logging into the system only a few times a day. In this case, it would not

make sense to choose an hour-long window because the resulting matrix would be very sparse. Thus, choosing a time window on a scale of days would be a better choice.

If multiple observations of the same type occur within a time window, one must decide how to aggregate these values. For some variables, such as diary entries, taking an average may be desirable; for other variables, such as log-ins, the sum is a more appropriate aggregation. The provided technical framework supports the data transformation procedure. In addition, missing values often exist in the data. There are various procedures that can handle missing values. One might remove all rows that include missing values; however, this can lead to a reduction in observations. Other possibilities include imputation procedures such as using aggregated values of these features or

developing statistical models that predict the missing values based on other features. For more information on missing values, we refer to the study by Batista and Monard [36].

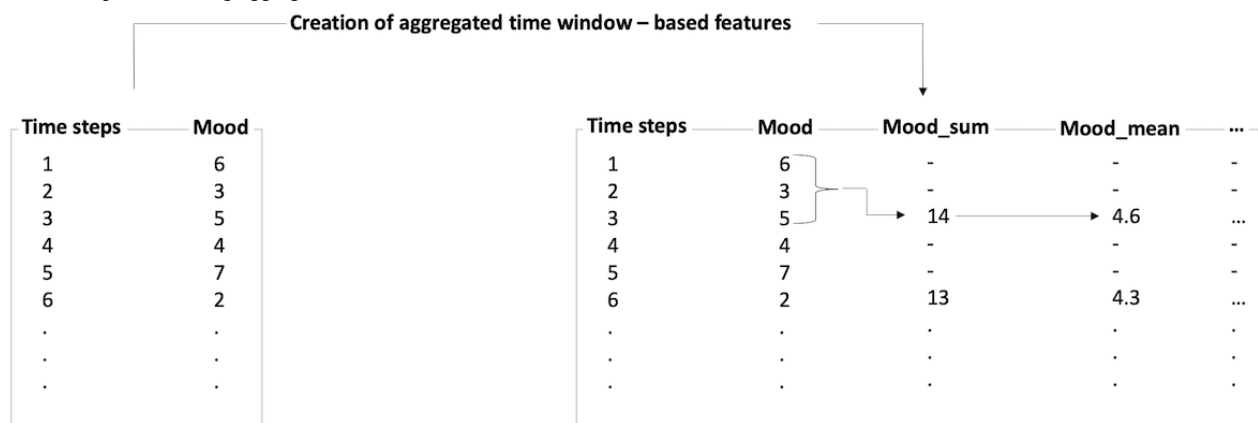
### Step Two: Feature Engineering

Feature engineering can be described as the process of including additional variables into the data with the intention of achieving increased predictive performance. As statistical learning relies heavily on the input data, this step is important for improving the accuracy of prediction [37]. There are 2 approaches to feature engineering: handcrafted or automated. Handcrafted feature engineering is a challenging task and requires human effort and domain knowledge. Therefore, it is appropriate for researchers with expertise in the domain that is represented by the data (eg, sleep) to be highly involved in the process [38-40]. A clear understanding of the problem to be solved is necessary to derive meaningful features [40]. Handcrafted feature engineering often involves a trial and error phase to experiment with different features [37]. Automated feature engineering involves the generation of candidate features that are evaluated based on their predictive performance. Tools exist for the application of automated feature engineering in different domains, such as natural language processing or machine vision [38,41,42].

Interaction terms, that is, the product of 2 original features, can lead to additional knowledge about their relationships and increased predictive accuracy. The provided technical framework supports generating them. In case of a large number of original features, however, including interaction terms results in many additional features.

In addition, time window-based aggregation methods can be beneficial in terms of predictive performance in the context of digital health interventions [31]. Here, based on a user-specified time window  $w$ , various types of aggregations are performed on the original features. Figure 3 represents the process of this task through the exemplification of self-reported EMA data. The *Mood* level is reported by an individual at different points in time (*Time steps*). For the creation of the aggregated features, a time window of  $w=3$  is specified in this example. Various statistical measures, such as the sum (*Mood\_sum*), mean (*Mood\_mean*), minimum, maximum, and SD (not shown in figure), are calculated for 3 consecutive measurements of the mood level ( $w=3$ ) and included as additional features in the data set. It should be noted that the creation of features can limit one's ability to reproduce study results if the feature engineering process is not well documented or if the data set changes over time. For the case study in this paper, we created various theory-driven features based on expert knowledge, which will be introduced in *Feature Engineering*.

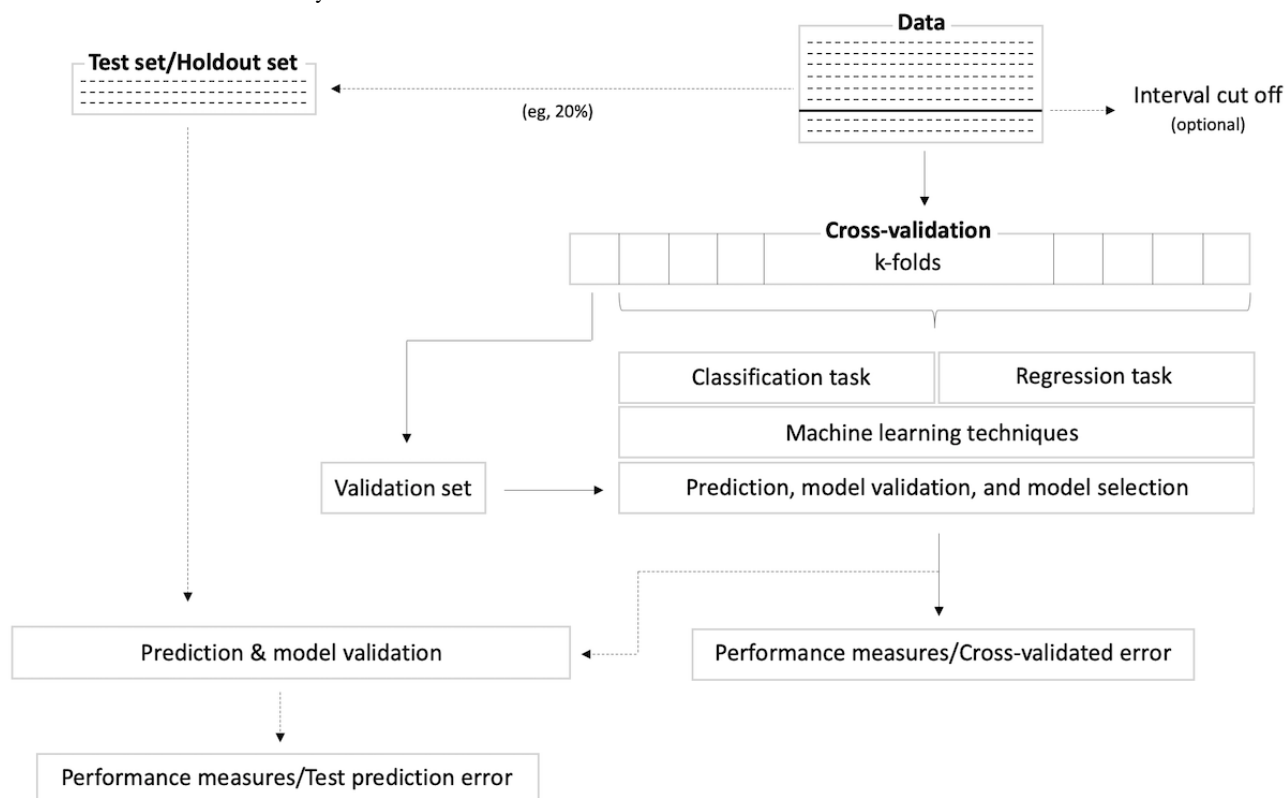
**Figure 3.** Example of creating aggregated time window-based features for  $w=3$ .



### Step Three: Statistical Analysis and Model Validation

The next step in analyzing user journey data is the application of machine learning techniques to predict the outcome variable. Figure 4 depicts this procedure. First, the data set can be split into a training set for fitting the data and learning patterns and

a test (or holdout) set. This test set is usually created if sufficient data are available. It is subsequently used to test the final model performance of the selected algorithm. It is difficult, however, to quantify *sufficient data* as it depends strongly on the field of research, applied models, and structure of the data.

**Figure 4.** Procedure of statistical analysis.

Depending on the task to be analyzed, the data can be further split based on particular points in time. If the aim of the analysis, for example, is the prediction of the outcome of an intervention, it might be useful to evaluate at what point in time the predictive accuracy is at its peak. The longer the time window, the higher the predictive accuracy can be assumed because more data are available. Thus, using time windows and basing the amount of usable data on these windows (*interval cut off*) can be useful in evaluating the feasibility of prediction.

There are a large number of machine learning techniques that can be applied to user journey data; some models can be applied to both learning tasks (classification or regression), such as support vector machines or decision trees, whereas others fit better for a specific task (ie, logistic regression for classification). Researchers may wish to compare their predictive performance to justify the model selection. Cross-validation is often applied to gauge the predictive performance of a specified model. Here, the data are divided into  $k$  chunks, where  $k-1$  chunks are used for training the machine learning techniques and the remaining data chunk is used for predicting the target variable. This procedure is repeated  $k$  times until each chunk has been used as a validation set. Ultimately, the model with the best performance is selected for the specified learning task. If a holdout set is maintained, the specified model is then trained based on all data. The target variable in the holdout set is then predicted and evaluated, which leads to the test prediction error.

Model validation checks the ability of a particular model to either fit the data or predict the outcome variable [43]. Eventually, the one with the best performance is selected. Nonvalidation can lead to inaccurate predictions and, thus, overconfidence in the developed model [44]. Model validation

should generally be executed on the validation set for each iteration of the cross-validation procedure (cross-validated prediction error) to select the best model and, subsequently, on an independent test set that was set aside earlier (test prediction error). In some cases, especially when sufficient data are not available, no independent test set is put aside and only the cross-validated error is reported, which can lead to an optimistic estimation of the error [44].

Deciding on the method of model validation also depends on the learning task. For regression, criteria such as the root mean square error or mean absolute error are often appropriate. For the classification task, confusion matrices and receiver operating characteristic (ROC) graphs are often used as performance indicators. More information about these validation procedures and their application can be found elsewhere [45].

In the provided technical framework, logistic regression, linear regression, support vector machines, boosted decision trees, and regularization techniques are implemented. As overfitting can occur when utilizing a large number of features [37] and some types of statistical procedures (eg, linear regression) cannot be applied when the number of features is greater than the number of observations, alternative techniques such as regularization and feature selection may need to be used [46]. A thorough review of these techniques is outside the scope of this paper, and readers are strongly encouraged to learn more about each of these techniques and how they pertain to their data and aims.

### Case Study

To illustrate the user journey analysis process, data were extracted from a trial of a web-based program (SHUTi) [47].



SHUTi is a fully automated web-delivered program that is tailored to individual users [47] and informed by the model for internet interventions [17]. SHUTi is based on the primary principles of face-to-face cognitive behavioral therapy for insomnia (CBT-I), including sleep restriction, stimulus control, cognitive restructuring, sleep hygiene, and relapse prevention. SHUTi contains 7 *cores* that are dispensed over time, the first core being a tutorial on how to use the program, with new cores becoming available 7 days after completion of a previous core. This format was meant to mirror traditional CBT-I delivery procedures using a weekly session format. SHUTi has been found to be more efficacious than web-based patient education in changing primary sleep outcomes (insomnia severity, sleep onset latency [SOL], and wake after sleep onset [WASO]), with the majority of SHUTi users achieving insomnia remission status 1 year later [48]. A mobile app version of SHUTi, Somryst, with equivalent content and mechanisms of action was recently cleared by Food and Drug Administration as the first prescription digital therapeutic for treating patients with chronic insomnia. Thus, the efficacy of SHUTi is well established. However, similar to other digital interventions, predicting user dropout is an important yet unaddressed issue. Thus, the primary aim of this case study is to demonstrate the feasibility of predicting user dropout from data generated by a digital health intervention.

The sample for this study was drawn from a trial consisting of 303 participants (218/303, 71.9% female) aged between 21 and 65 years (mean 43.3 years, SD 11.6). They were 83.8% (254/303) White, 6.9% (21/303) Black, 4.0% (12/303) Asian, and 5.3% (16/303) *other*. Participants were randomly assigned (using a random number generator) to receive SHUTi or web-based patient education (control condition). The study was approved by the local university's institutional review board, and the project was registered on clinicaltrials.gov

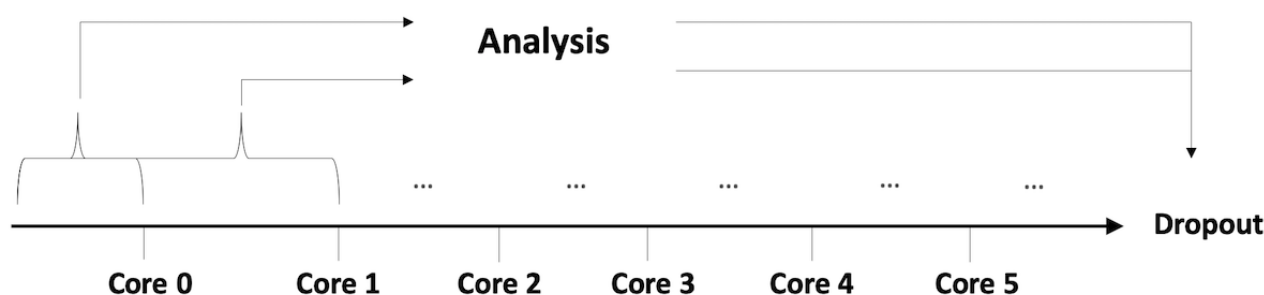
(NCT01438697). Inclusionary and exclusionary criteria as well as outcomes are reported in detail elsewhere [48].

Data from 151 participants who were assigned to SHUTi were used in this study. Both self-reported and system-generated types of data are available. Participants completed a battery of self-report measures at baseline and post intervention. A list and detailed description of the measures have been published previously [48]. Sleep diaries were also collected throughout the intervention period, along with information about bedtime, length of sleep onset, number and duration of awakenings, perceived sleep quality, and rising time. Data were collected prospectively for 10 days (during a 2-week period) at each of the 4 assessment periods (pre- and postintervention and 6- and 12-month follow-ups). Sleep diary questions mirrored those from the consensus sleep diary [49]. Values for SOL and WASO were averaged across the 10 days of diary collection at each assessment period. The system-generated data included individual log-ins and automated emails sent by the system as well as trigger events logged in the system. All data were used to predict user dropout, defined as not completing all 7 SHUTi cores (core 0 through core 6). Thus, users were classified as having dropped out or not. As noted elsewhere [48], 60.3% (91/151) participants completed all 7 cores in the SHUTi program.

## Results

The primary aim was to predict whether users prematurely dropped out of SHUTi (dropped out by core 6/completed core 6). Therefore, the learning problem is a binary classification (drop out/did not drop out). To verify the point at which the machine learning techniques were capable of predicting dropout, separate analyses were executed after the completion of each core (Figure 5) and only included data up to the core in question. The number of participants included in each analysis was 146, 141, 133, 116, 102, and 101 for cores 0 to 5, respectively.

**Figure 5.** Setup of analysis for dropout prediction.



## Data Transformation

As a first step, the raw data were transformed into a rectangular data matrix (wide format), which led to 981 basic features. Basic features are those features that were already included in the raw data. As an example, see column *Type* in Figure 2. In addition, 25 handcrafted and theory-driven features that were derived from the raw data were implemented. These features are introduced in the next section *Feature Engineering*. In total, 1006 features were used for the analyses. Whenever the same question (ie, in the case of diary data) was administered multiple

times a day, the mean of the reported values was chosen for numeric data and the mode for categorical data. To reduce the sparseness of the resulting data matrix, reported values for questionnaires such as the Insomnia Severity Index were repeated for each participant until the next occurrence of the questionnaire (this questionnaire was administered before each core). To address the issue of missing data, features were deleted based on the quantity of missing data. To evaluate how the deletion affects the predictive performance of the models, features were deleted that contained more than 5%, 10%, 15%, and 20% of missing values. This procedure reduced the number



of features tremendously. In addition, categorical variables that had only one level or category were removed. Less data are available for the analysis at time point core 0 compared with time point core 5. Thus, the number of features for each level of missing data was 83, 263, 299, and 401 features.

As the aim of this study was to predict dropout at core 6, each participant only had exactly one outcome value—they could either complete core 6 or not. Users that dropped out between cores 1 to 5 would be classified as having dropped out at core 6. Therefore, the user journey data must be aggregated for each

user. For most of the variables, the mean and mode were used as the aggregation method. However, for some variables, such as log-in information or number of days since the last contact, the sum is more appropriate. Table 1 illustrates the different aggregation procedures and the corresponding features. Features that are not listed were aggregated by mean and mode. The rest of the missing data were imputed using the median for numeric variables and mode for categorical features. In addition, an imputation based on the k-nearest neighbor (KNN) algorithm was applied (k=5). Both approaches were used to reveal which of them led to a better prediction performance.

**Table 1.** Aggregation of theory-determined features.

Feature aggregation method	Handcrafted features	Existing clinically important features
Sum: The sum of all observations of a specific feature for an individual	<ul style="list-style-type: none"> <li>Days since the last contact (any interaction)</li> <li>If sleeping duration is decreasing from core to core</li> <li>If sleep window duration is 5 or 8 hours</li> </ul>	<ul style="list-style-type: none"> <li>If the participant had an alcoholic drink that day</li> <li>If the participant took a nap</li> <li>If the system recorded a triggered event that day</li> <li>If the participant logged in that day</li> <li>If the system sent an email that day</li> </ul>
Last: The last observation of a specific feature for an individual	<ul style="list-style-type: none"> <li>Difference between preferred arising time in core 2 and core 3</li> <li>If preferred arising time is greater than 8 AM in core 2</li> <li>Average time in days to complete a core among all cores that have been available</li> <li>Time needed in days to complete a core in days (6 features for core 0-5)</li> </ul>	<ul style="list-style-type: none"> <li>If the participant finished homework in core 2</li> <li>Number of days where no diaries have been completed in the period of analysis</li> <li>Precipitating factor includes <i>major life event</i> or <i>health/psychological</i></li> </ul>
Mean: Mean of the observations of a specific feature for an individual	<ul style="list-style-type: none"> <li>Difference between awake and arise time</li> <li>Difference between preferred arise time and actual arise time (AM/PM)</li> <li>Difference between preferred arise time and actual arise time (minutes)</li> <li>Difference between preferred bedtime and actual bedtime</li> </ul>	<ul style="list-style-type: none"> <li>Naptime in minutes</li> </ul>

## Feature Engineering

A total of 25 theory-driven features were implemented for this case study. Some of these features, shown in Table 1, were handcrafted and some were already existing in the data set. Specifically, the handcrafted features were computed from the raw data and were deemed useful for model prediction. Few of these features are study-specific (eg, *if the participant finished homework in core 2*), whereas others could be used in any type of digital intervention (eg, *if the participant logged in*). As the number of features generated from the study data was already large, none of the generic feature generation methods were used. These 25 features were not deleted based on the missing value ratio (mentioned above) because there was a clinical or theory-driven rationale that they would influence prediction performance.

## Statistical Analysis and Model Validation

For the learning task, a set of machine learning techniques was used to select the model with the best prediction performance.

Specifically, support vector machines, boosted decision trees, and logistic regression with L1 and L2 regularization were applied. The optimal parameters were determined using a grid-based search and cross-validation. In addition, stratified 10-fold cross-validation was used for each analysis. To choose an appropriate statistical model, a heat map was created to illustrate the average area under the curve (AUC) across all core analyses for each model, imputation procedure, and threshold for percentage of missing values (Figure 6). As can be seen, the method of imputing the missing values did not have a strong influence on the performance of the applied statistical model. Increasing the percentage threshold negatively influenced L1 regularization and the support vector machine, whereas L2 regularization and boosted decision trees seemed not to be influenced tremendously. The best average AUC value (0.719) was achieved by applying boosted decision trees, deleting each feature that contained more than 15% of missing values, and imputing the rest of the missing values by KNN.

**Figure 6.** Heat map of average area under the curve values across core analyses for each model, imputation procedure, and threshold for percentage of missing values. AUC: area under the curve; KNN: k-nearest neighbor; LASSO: least absolute shrinkage and selection operator; SVM: support vector machine.

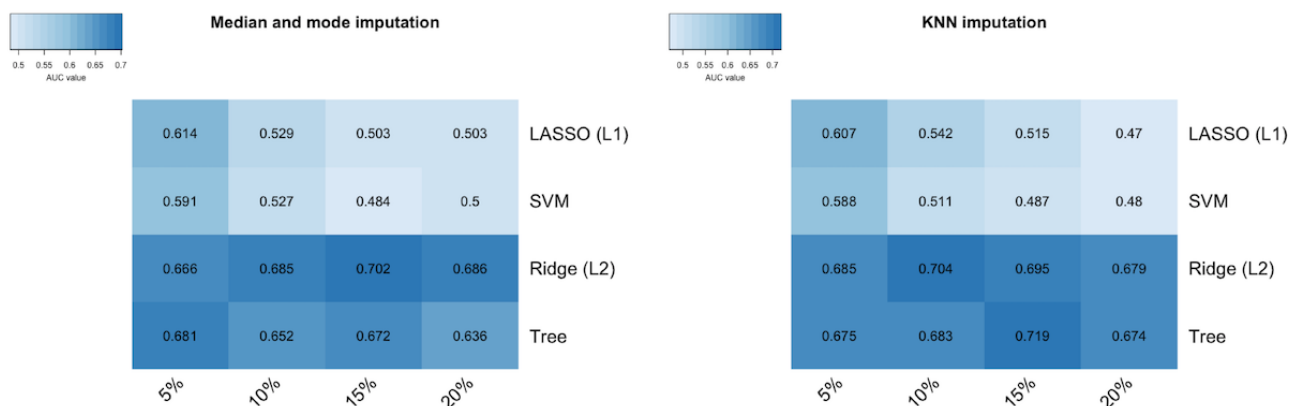
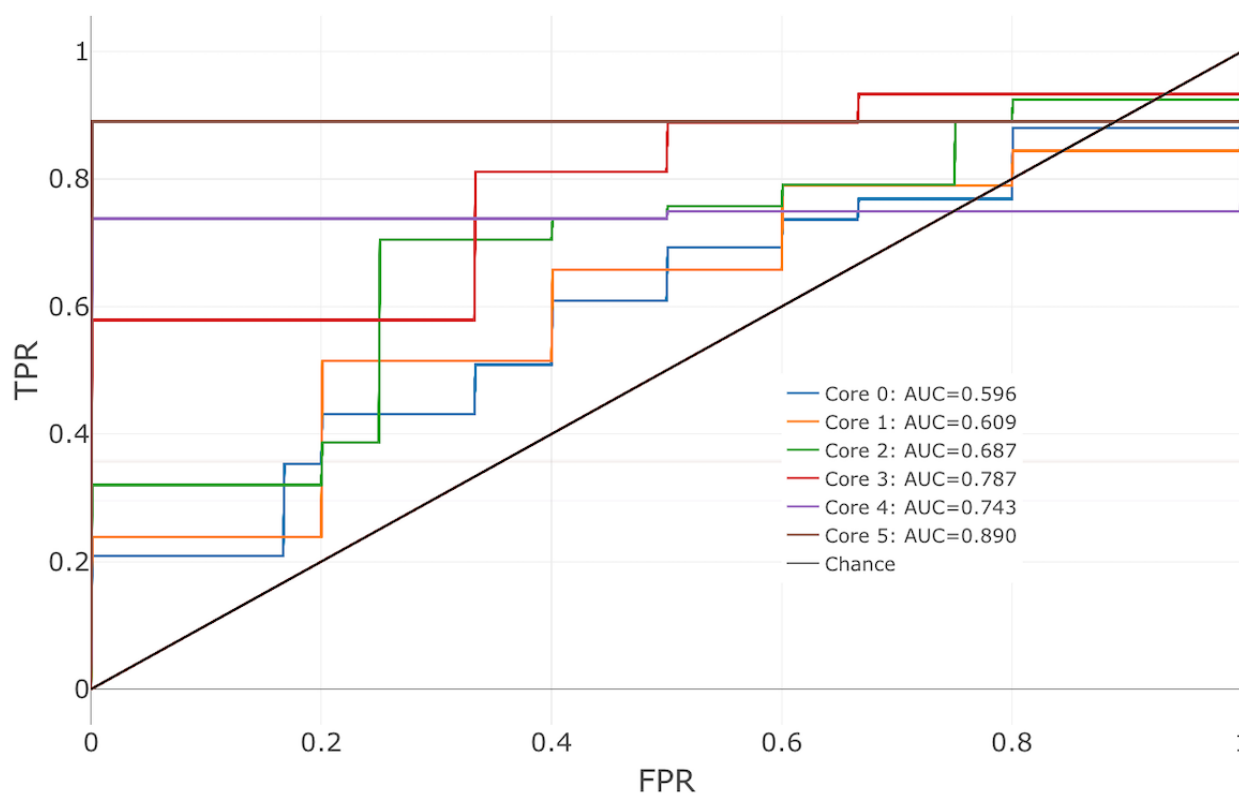


Figure 7 illustrates the ROC curves for each core analysis using the specified parameters. With the exception of core 4, the AUC values increased with each analysis. For each core, the predictions were better than random, indicated by AUC values above 0.5. Generally, the AUC values ranged between 0.6 and 0.9. Importantly, the prediction of dropout appears feasible early

in the intervention period (ie, core 1 and core 2). In addition, the area under the precision-recall curve (PRAUC) was computed. Across all core analyses, a PRAUC of 0.48 was observed, whereas chance had an average of 0.24. Thus, the model performs better than chance.

**Figure 7.** Receiver operating characteristic for each core analysis based on boosted decision trees (15% missing value deletion, k-nearest neighbor imputation). AUC: area under the curve; FPR: false-positive rate; TPR: true-positive rate.



Boosted decision trees were used to identify important features. Here, SHapley Additive exPlanation (SHAP) values were used [50]. SHAP values are a relatively new concept in the field of machine learning and essentially represent the importance of each feature and their contribution to the prediction by comparing the prediction of the model with and without a specified feature value depending on the order of their

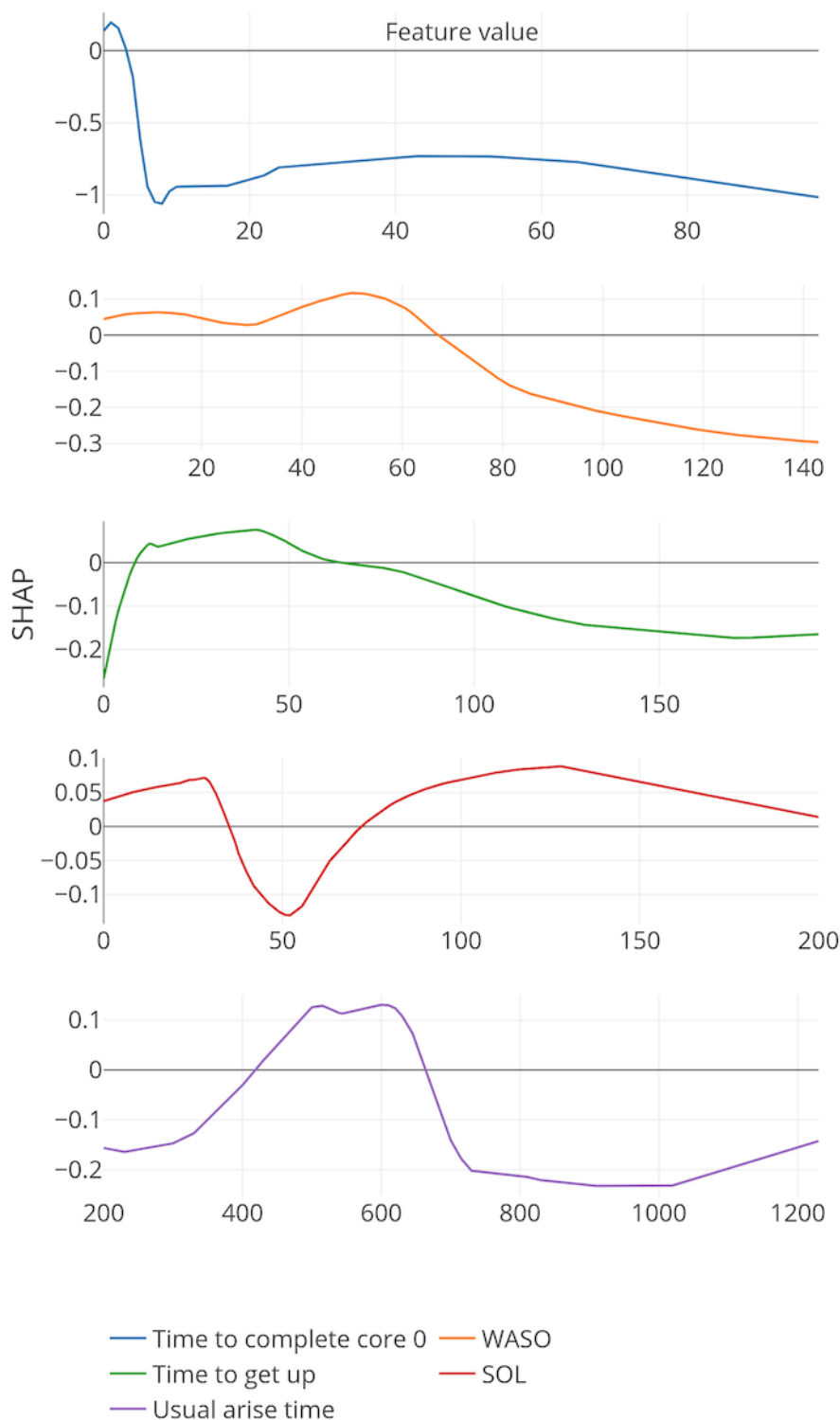
introduction to the model. In addition to the importance of each feature, SHAP values quantify how features contribute to the prediction of the model.

Figures 8-13 include the 5 most important features according to the boosted decision trees for each core analysis. In each graph, the x-axis represents the values for each feature and the y-axis represents the SHAP values (ie, the effect each feature

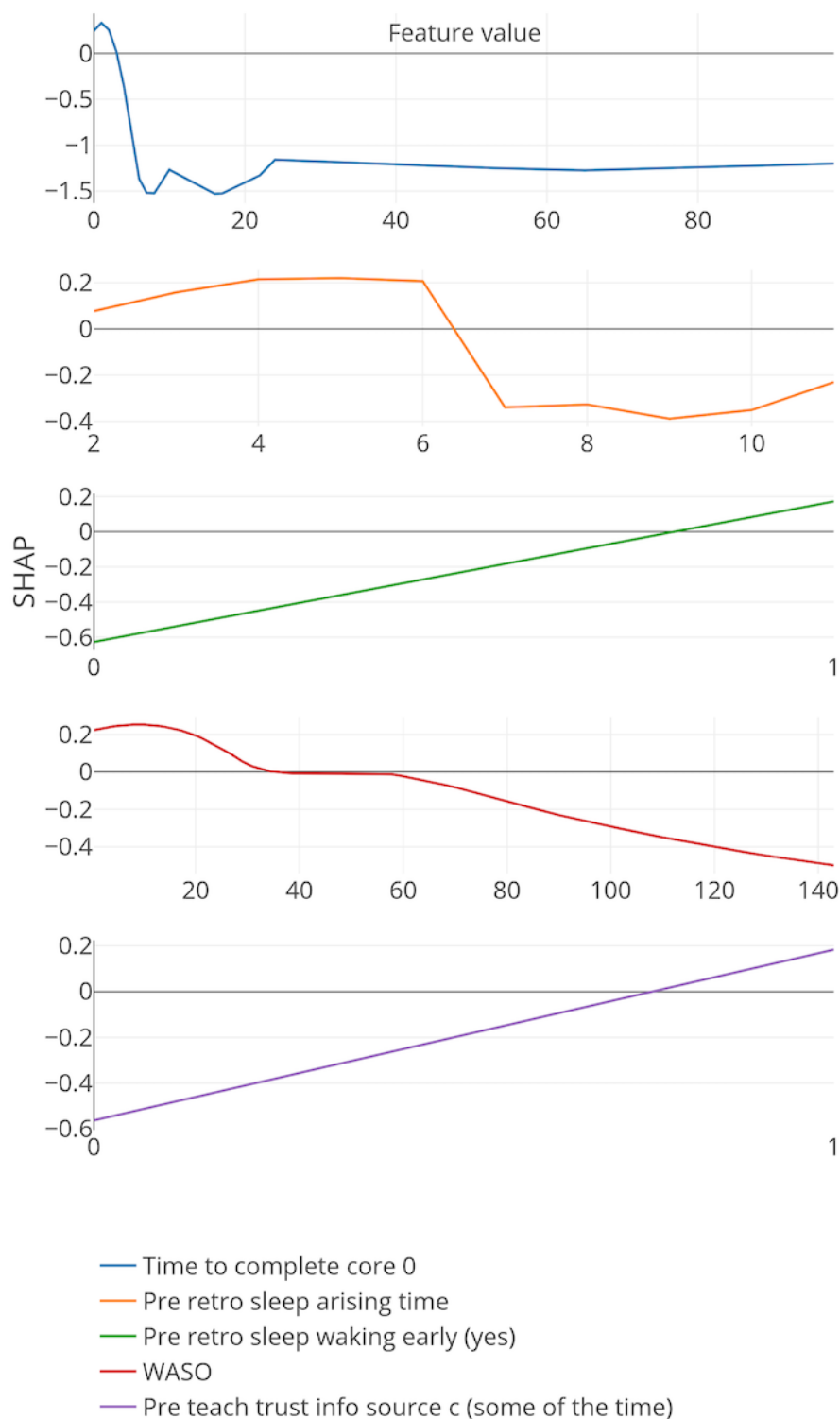
has on predicting the completion of core 6 of the intervention). In the core 0 analysis, for example, finishing core 0 within 3 days (x-axis) has a positive influence on dropout, as can be seen on the y-axis above zero. However, taking more time to

complete core 0 (where x-axis is greater than 3) influences dropout prediction negatively as the graph approaches values under zero.

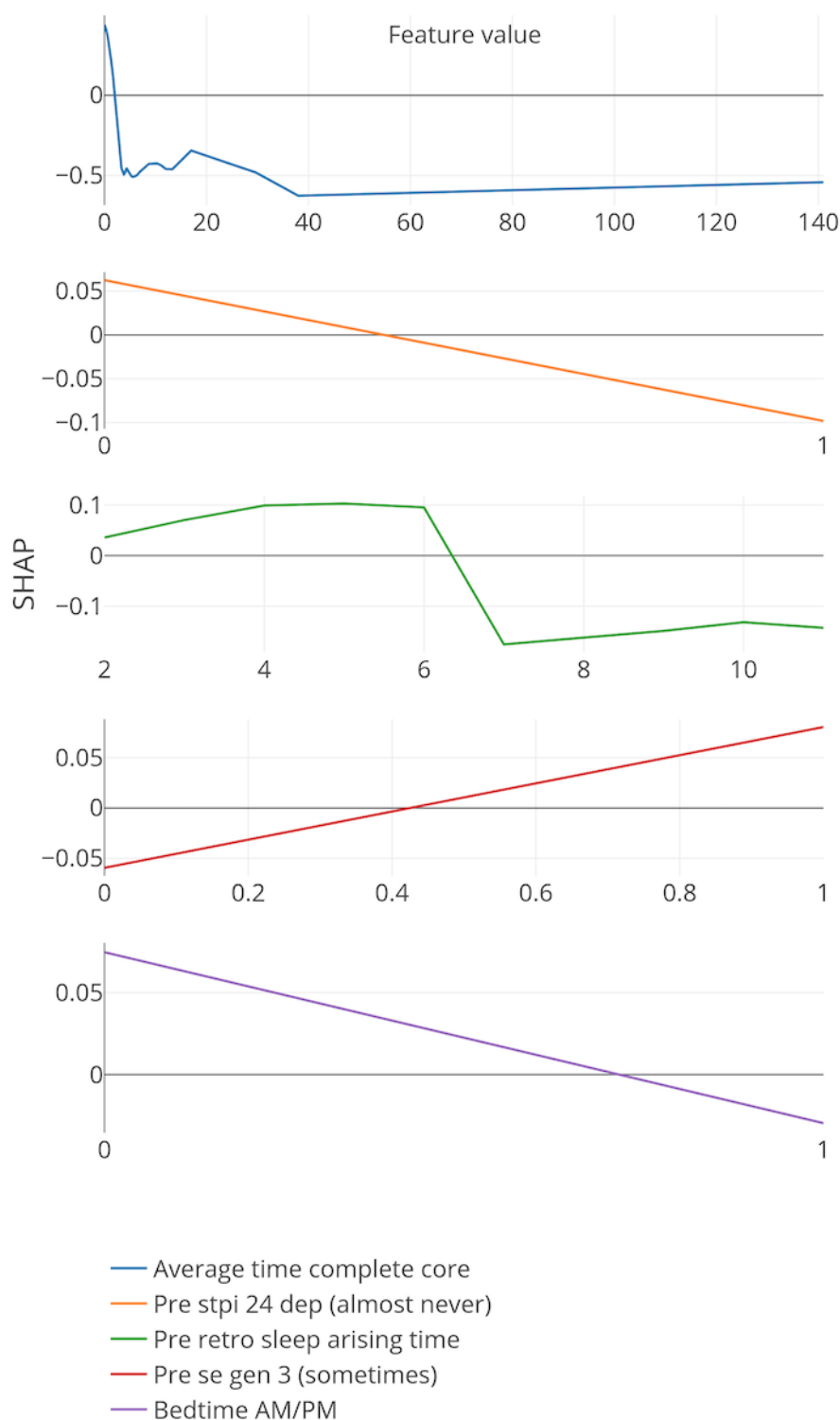
**Figure 8.** Five most important features for each core analysis according to boosted decision trees (15% deletion of missing values, and k-nearest neighbor imputation). The x-axis represents the values for each feature, and the y-axis represents the SHAP values. SHAP: SHapley Additive exPlanation; SOL: sleep onset latency; WASO: wake after sleep onset.



**Figure 9.** Five most important features for each core analysis according to boosted decision trees (15% deletion of missing values, KNN imputation, and Core 1 analysis). SHAP: SHapley Additive exPlanation; WASO: wake after sleep onset.

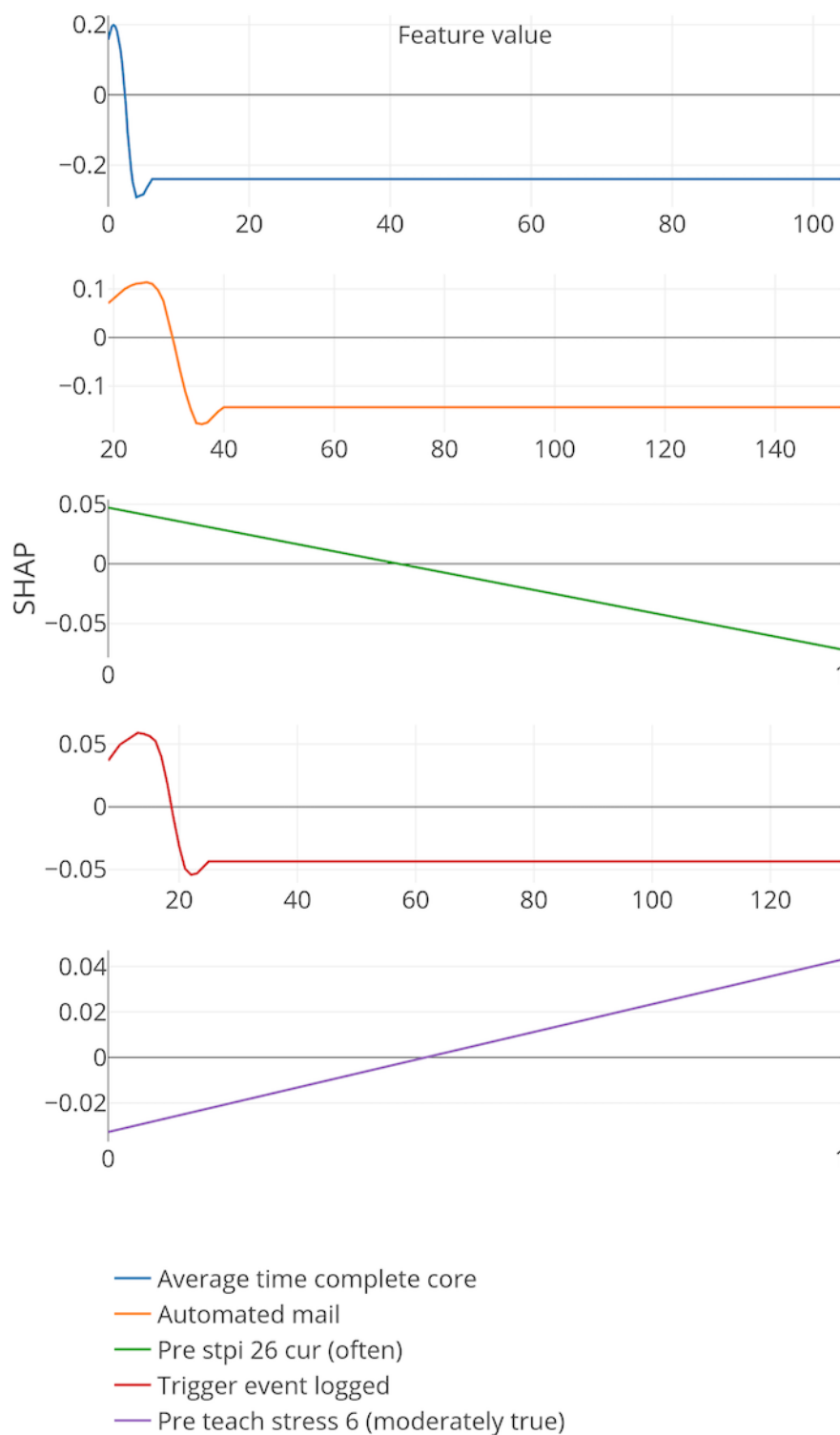


**Figure 10.** Five most important features for each core analysis according to boosted decision trees (15% deletion of missing values, KNN imputation, and Core 2 analysis). SHAP: SHapley Additive exPlanation.

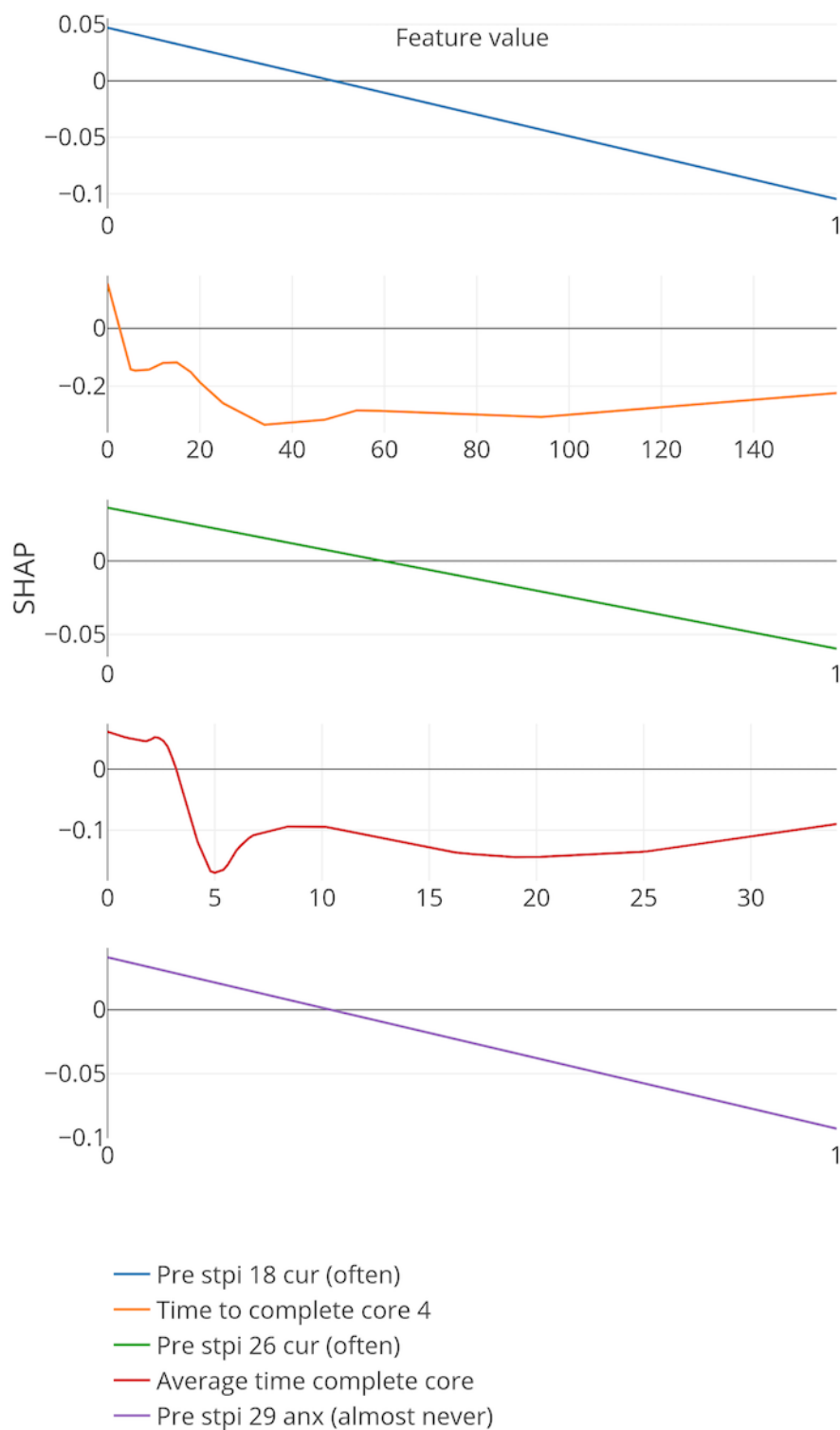




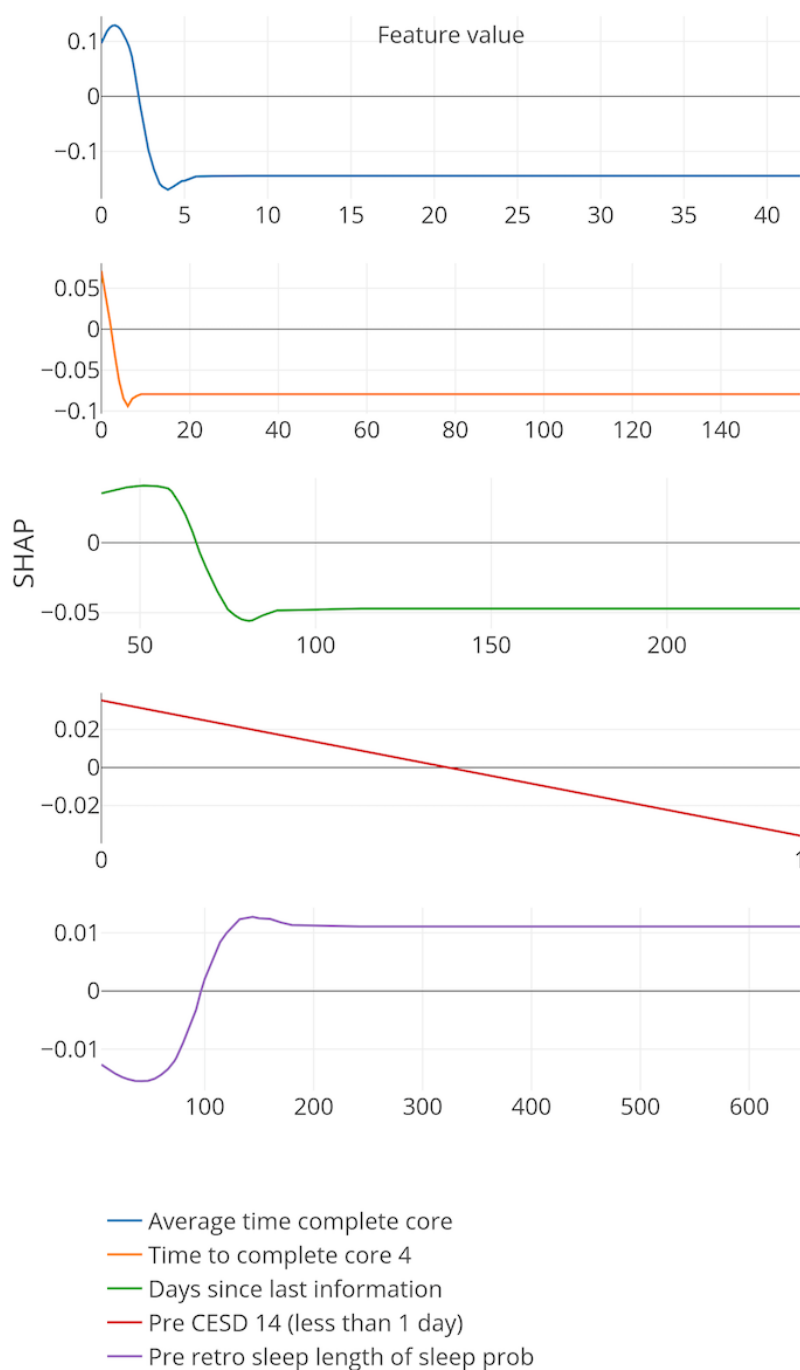
**Figure 11.** Five most important features for each core analysis according to boosted decision trees (15% deletion of missing values, KNN imputation, and Core 3 analysis). SHAP: SHapley Additive exPlanation.



**Figure 12.** Five most important features for each core analysis according to boosted decision trees (15% deletion of missing values, KNN imputation, and Core 4 analysis). SHAP: SHapley Additive exPlanation.



**Figure 13.** Five most important features for each core analysis according to boosted decision trees (15% deletion of missing values, KNN imputation, and Core 5 analysis). SHAP: SHapley Additive exPlanation.



In general, 7 out of the strongest 22 features were handcrafted and theory driven. Table 2 summarizes all the features. Taking more time to complete the cores appeared to influence dropout. The time to complete core 0 predicted whether a participant eventually dropped out (core 0 and core 1 analysis). In addition, usual arise time and the time needed to get out of bed (from awake to arise) affected the prediction of dropout early on in the intervention. Participants who got up earlier than 4:30 AM and later than 6:45 AM, and participants who needed less than 9 min or more than 66 min to get up, negatively influenced the

prediction of completing core 6 of the intervention (x-axis of the feature usual arise time and time to get up for core 0). Furthermore, a greater WASO also appeared to influence the prediction of dropout status. These variables could, therefore, be an early indicator of dropout in this particular intervention.

In addition, if triggers were logged on for more than 18 days or participants received emails for more than 30 days, dropping out was more likely (core 3 analysis). Furthermore, if there was no interaction between the system and the participants for more than 67 days, the individuals were more likely to drop out.

**Table 2.** Summary of the unique top 5 most important features across analyses.

Predictors		Analysis at each point in time					
Feature	Description	Core 0	Core 1	Core 2	Core 3	Core 4	Core 5
Core 0 completion date—intervention start date <sup>a</sup>	Time to complete core 0 in days	+ <sup>b</sup>	+	N/A <sup>c</sup>	N/A	N/A	N/A
Arise time—awake time <sup>a</sup>	Difference between time of awakening and getting out of bed in minutes (time to get up)	+	N/A	N/A	N/A	N/A	N/A
Usual arise time	Retrospective report specified from baseline data	+	N/A	N/A	N/A	N/A	N/A
Wake after sleep onset	Minutes awake in the middle of the night from sleep diaries	+	+	N/A	N/A	N/A	N/A
Sleep onset latency	Minutes to fall asleep from sleep diaries	+	N/A	N/A	N/A	N/A	N/A
Baseline arise time (pre retro sleep arising time)	Time the user specified that they got out of bed from baseline data	N/A	+	+	N/A	N/A	N/A
Pre retro sleep waking early	User indicates having problems waking up too early in the morning	N/A	+	N/A	N/A	N/A	N/A
Pre teach trust info source c	How much the user trusts health information	N/A	+	N/A	N/A	N/A	N/A
Average time to complete core <sup>a</sup>	Average time to complete a core among all cores that have been available up to the point of the analysis	N/A	N/A	+	+	+	+
Pre stpi 24 dep <sup>d,e</sup>	How low the user feels at baseline	N/A	N/A	+	N/A	N/A	N/A
Pre se gen 3 <sup>f</sup>	How well the user feels things have been going	N/A	N/A	+	N/A	N/A	N/A
Bedtime	If a participant went to bed in the AM or PM (before or after 12 AM)	N/A	N/A	+	N/A	N/A	N/A
Email sent <sup>a</sup>	If the system sent an email that day	N/A	N/A	N/A	+	N/A	N/A
Pre stpi 26 cur <sup>g</sup>	How stimulated the user feels at baseline	N/A	N/A	N/A	+	+	N/A
Trigger event logged <sup>a</sup>	If the system logged a trigger event that day	N/A	N/A	N/A	+	N/A	N/A
Pre teach stress 6	User feels he or she can solve most problems if necessary effort is put in	N/A	N/A	N/A	+	N/A	N/A
Pre stpi 18 cur <sup>h</sup>	How eager the user feels at baseline	N/A	N/A	N/A	N/A	+	N/A
Core 4 completion date—core 4 start date <sup>a</sup>	Time to complete core 4 in days	N/A	N/A	N/A	N/A	+	+
Pre stpi 29 anx <sup>i</sup>	How much self-confidence the user feels at baseline	N/A	N/A	N/A	N/A	+	N/A
Days since the last information <sup>a</sup>	Days since the last contact (any interaction)	N/A	N/A	N/A	N/A	N/A	+
Pre CESD <sup>j</sup> 14 <sup>k</sup>	How lonely the user feels at baseline	N/A	N/A	N/A	N/A	N/A	+
Pre retro sleep length of sleep prob	Number of months the user reports having had sleep difficulties at baseline.	N/A	N/A	N/A	N/A	N/A	+

<sup>a</sup>Handcrafted/theory-driven features.<sup>b</sup>+ indicates appearance of feature in corresponding core analysis.<sup>c</sup>N/A: not applicable.<sup>d</sup>STPI: state-trait personality inventory.<sup>e</sup>Pre stpi 24 dep: baseline STPI measure item #24 depression subscale.<sup>f</sup>Pre se gen 3: baseline Perceived Stress Scale item #5.<sup>g</sup>Pre stpi 26 cur: baseline STPI measure item #26 curiosity subscale.<sup>h</sup>Pre stpi 18 cur: baseline STPI measure item #18 curiosity subscale.<sup>i</sup>Pre stpi 29 anx: baseline STPI measure item #29 anxiety subscale.<sup>j</sup>Center for Epidemiologic Studies Depression Scale.<sup>k</sup>Pre CESD 14: baseline CESD measure item #14.

## Discussion

### Principal Findings

Considering the increasing use of digital health interventions and the tremendous amount of data gathered in such interventions, a variety of methods can be used for the analysis of various data types and structures. In this study, a process for the analysis of user journey data in this context was proposed, and a step-by-step guide and technical framework for the analysis as an R package was provided. Challenges of data analysis based on user journeys, such as data transformation, feature engineering, and statistical model application and evaluation, were discussed. The analysis of user journeys can be a powerful tool for the prediction of various factors on an individual participant level. Here, it has been applied to real-world data to predict dropout from an internet-based intervention.

The application of the proposed process and evaluation of statistical models indicated the feasibility of dropout prediction by using this process. AUC values ranged between 0.6 and 0.9 for the selected machine learning algorithm (boosted decision trees). Most importantly, it was shown that the prediction of user dropout was possible early in the intervention, which could be helpful to clinicians and policy makers as treatment decisions are made and adjusted. In addition, this study indicated the importance of expert knowledge and subsequent implementation of handcrafted features. Not all existing statistical models necessarily require handcrafted features because automated feature engineering can already provide crucial insight; however, handcrafted features can increase prediction performance and lead to increased interpretability. In this study, handcrafted features appeared to be among the most important features according to the boosted decision trees, perhaps given the more nuanced understanding necessary for treating insomnia. It is important to keep in mind, though, that the analysis presented here was meant as a demonstration of the power of this approach. A much larger data set is needed to draw more firm and generalizable conclusions.

With this caveat, a number of interesting results emerged related to features and impact on dropout prediction. For example, as participants took longer to complete earlier steps of the intervention, they were less likely to complete the final step of the intervention. Thus, a discussion about how users can be motivated to complete early steps in the intervention may be very beneficial. In addition, the findings suggest that the time participants get out of bed in the morning and how much time they actually needed to get up might be an important factor for completing the sleep intervention. Participants who get out of bed between 4:30 AM and 6:45 AM and do not need more than 66 min to get out of bed were more likely to complete the final step of the intervention. In addition, trigger events might only have a positive effect in the short term, as the appearance of triggers more often than 18 days appeared to increase the likelihood of dropping out. However, it could be possible that this finding only accounts for participants who would not have completed the final step of the intervention. Assuming this, these participants were, therefore, not influenced by trigger

events. It is also important to emphasize that these results are based on a bottom-up, data-driven learning approach. Therefore, it is up to researchers to interpret the results and cross-validate them in other samples. Predictions in this context based on user journey data and the resulting knowledge about factors that influence these predictions, especially on an individual level, could lead to the implementation of strategies that seek to improve the utilization and efficacy of digital health interventions.

### Limitations

There are a number of limitations of this study that should be considered when interpreting the results. One limitation is the relatively limited number of participants included in the analysis and the large feature space. The predictive performance of the applied models is satisfactory, especially early on in the intervention. The process and models described in this study are technically feasible, although the reliability of the ensuing results may be impacted by limitations to sample size. Owing to the limited number of participants, the results of this study should be replicated in a larger sample. Furthermore, the amount of missing values impacts the analyses and can lead to bias. Obtaining more complete data can further increase the interpretability and predictive accuracy of the models. In addition to time window-based features and time-dependent variables, the demonstrated steps and this study in general do not include time-dependent feature engineering, such as the relation between features and observations across time. Researchers should examine the data set they are planning to analyze to determine whether time-dynamic features could be used in their projects. Another limitation is the fact that the data are heterogeneous at an individual participant level; thus, the application of models that consider heterogeneous parameters might provide deeper and more individualized information about the participants. However, considering the number of participants in the data, heterogeneous models have not yet been investigated. The results are, nevertheless, promising and can lead to increased knowledge about users and how dropout from digital health interventions is affected by various factors. Studies using larger data sets are necessary to improve model performance and confirm findings.

### Conclusions

This study proposes a step-by-step process for the analysis of user journey data in the context of digital health interventions and provides a technical framework. Furthermore, the proposed framework was applied to data from an internet-based intervention for insomnia to predict dropout of participants. These participants needed to complete 7 cores to finish the program. Importantly, our process was able to predict user dropout at each core better than chance. The predictive performance also varied by core; although the AUC was approximately 0.6 for cores 0 and 1, it was noticeably higher for the latter cores. This indicates that the user journey process can be used to predict dropout early in the intervention and prediction accuracy increases over the course of the intervention. This may allow researchers to preemptively address dropout before it occurs by providing support to users that may be struggling to engage. Among the machine learning techniques



we evaluated, boosted decision trees provided the greatest accuracy while deleting features that contained more than 15% missing values. In addition, a varying set of features was revealed that contributed to the prediction performance of dropout in this context. Replicating the results of this study in

a larger sample is needed to further validate the process outlined in this paper. Researchers may also wish to develop methods that predict the likelihood of user dropout over the duration of an intervention, which could enable researchers to devote resources to those at the highest risk of dropping out.

## Acknowledgments

This study was supported by grant R01 MH86758 from the National Institute of Mental Health. The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The authors thank Christina Frederick, BS, for her help with the study administration tasks. The authors specially thank Gabe D Heath, BA, and Steve P Johnson, BA, developers of the SHUTi intervention, for extracting and making all the data readily available for analysis.

## Conflicts of Interest

FT and LR report having a financial and/or business interest in BeHealth Solutions and Pear Therapeutics, 2 companies that develop and disseminate digital therapeutics, including by licensing the therapeutic developed, based in part, on early versions of the software utilized in research reported in the enclosed paper. These companies had no role in preparing this manuscript. LR is also a consultant to Mahana Therapeutics, a separate digital therapeutic company not affiliated with this research. Some of the research in this paper was conducted while FT was a faculty member at the University of Virginia. At that time for FT, and ongoing for LR, the terms of these arrangements have been reviewed and approved by the University of Virginia in accordance with its policies.

## References

1. Saddichha S, Al-Desouki M, Lamia A, Linden IA, Krausz M. Online interventions for depression and anxiety - a systematic review. *Health Psychol Behav Med* 2014 Jan 1;2(1):841-881 [FREE Full text] [doi: [10.1080/21642850.2014.945934](https://doi.org/10.1080/21642850.2014.945934)] [Medline: [25750823](https://pubmed.ncbi.nlm.nih.gov/25750823/)]
2. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cogn Behav Ther* 2018 Jan;47(1):1-18. [doi: [10.1080/16506073.2017.1401115](https://doi.org/10.1080/16506073.2017.1401115)] [Medline: [29215315](https://pubmed.ncbi.nlm.nih.gov/29215315/)]
3. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. *J Med Internet Res* 2017 Sep 15;19(9):e306 [FREE Full text] [doi: [10.2196/jmir.6588](https://doi.org/10.2196/jmir.6588)] [Medline: [28916506](https://pubmed.ncbi.nlm.nih.gov/28916506/)]
4. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. *Br J Clin Psychol* 2010 Nov;49(Pt 4):455-471. [doi: [10.1348/014466509X472138](https://doi.org/10.1348/014466509X472138)] [Medline: [19799804](https://pubmed.ncbi.nlm.nih.gov/19799804/)]
5. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. *J Affect Disord* 2020 Feb 15;263:413-419. [doi: [10.1016/j.jad.2019.11.167](https://doi.org/10.1016/j.jad.2019.11.167)] [Medline: [31969272](https://pubmed.ncbi.nlm.nih.gov/31969272/)]
6. Horsch C, Lancee J, Beun RJ, Neerincx MA, Brinkman W. Adherence to technology-mediated insomnia treatment: a meta-analysis, interviews, and focus groups. *J Med Internet Res* 2015 Sep 4;17(9):e214 [FREE Full text] [doi: [10.2196/jmir.4115](https://doi.org/10.2196/jmir.4115)] [Medline: [26341671](https://pubmed.ncbi.nlm.nih.gov/26341671/)]
7. Wickwire EM. The value of digital insomnia therapeutics: what we know and what we need to know. *J Clin Sleep Med* 2019 Jan 15;15(1):11-13 [FREE Full text] [doi: [10.5664/jcsn.7558](https://doi.org/10.5664/jcsn.7558)] [Medline: [30621849](https://pubmed.ncbi.nlm.nih.gov/30621849/)]
8. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. *Am J Prev Med* 2007 Jul;33(1):54-64. [doi: [10.1016/j.amepre.2007.02.041](https://doi.org/10.1016/j.amepre.2007.02.041)] [Medline: [17572313](https://pubmed.ncbi.nlm.nih.gov/17572313/)]
9. Funk KL, Stevens VJ, Appel LJ, Bauck A, Brantley PJ, Champagne CM, et al. Associations of internet website use with weight change in a long-term weight loss maintenance program. *J Med Internet Res* 2010 Jul 27;12(3):e29 [FREE Full text] [doi: [10.2196/jmir.1504](https://doi.org/10.2196/jmir.1504)] [Medline: [20663751](https://pubmed.ncbi.nlm.nih.gov/20663751/)]
10. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of technology-based strategies to promote engagement with digital interventions: a systematic review protocol. *J Med Internet Res Protoc* 2015 Apr 28;4(2):e47 [FREE Full text] [doi: [10.2196/resprot.3990](https://doi.org/10.2196/resprot.3990)] [Medline: [25921274](https://pubmed.ncbi.nlm.nih.gov/25921274/)]
11. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions: a systematic review. *J Med Internet Res* 2011 Jan 6;13(1):e2 [FREE Full text] [doi: [10.2196/jmir.1639](https://doi.org/10.2196/jmir.1639)] [Medline: [21212045](https://pubmed.ncbi.nlm.nih.gov/21212045/)]
12. Geraghty AW, Wood AM, Hyland ME. Attrition from self-directed interventions: investigating the relationship between psychological predictors, intervention content and dropout from a body dissatisfaction intervention. *Soc Sci Med* 2010 Jul;71(1):30-37. [doi: [10.1016/j.socscimed.2010.03.007](https://doi.org/10.1016/j.socscimed.2010.03.007)] [Medline: [20400220](https://pubmed.ncbi.nlm.nih.gov/20400220/)]

13. Eysenbach G. The law of attrition. *J Med Internet Res* 2005 Mar 31;7(1):e11 [[FREE Full text](#)] [doi: [10.2196/jmir.7.1.e11](#)] [Medline: [15829473](#)]
14. Pedersen DH, Mansourvar M, Sortsø C, Schmidt T. The law of attrition predicting dropouts from an electronic health platform for lifestyle interventions: analysis of methods and predictors. *J Med Internet Res* 2019 Sep 4;21(9):e13617 [[FREE Full text](#)] [doi: [10.2196/13617](#)] [Medline: [31486409](#)]
15. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. *J Med Internet Res* 2011 Aug 5;13(3):e52 [[FREE Full text](#)] [doi: [10.2196/jmir.1772](#)] [Medline: [21821503](#)]
16. Fernández-Álvarez J, Díaz-García A, González-Robles A, Baños R, García-Palacios A, Botella C. Dropping out of a transdiagnostic online intervention: a qualitative analysis of client's experiences. *Internet Interv* 2017 Dec;10:29-38 [[FREE Full text](#)] [doi: [10.1016/j.invent.2017.09.001](#)] [Medline: [30135750](#)]
17. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. *Ann Behav Med* 2009 Aug;38(1):18-27 [[FREE Full text](#)] [doi: [10.1007/s12160-009-9133-4](#)] [Medline: [19802647](#)]
18. Christensen H, Batterham PJ, Gosling JA, Ritterband LM, Griffiths KM, Thorndike FP, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. *Lancet Psychiatry* 2016 Apr;3(4):333-341. [doi: [10.1016/S2215-0366\(15\)00536-2](#)] [Medline: [26827250](#)]
19. Ritterband LM, Thorndike FP, Gonder-Frederick LA, Magee JC, Bailey ET, Saylor DK, et al. Efficacy of an internet-based behavioral intervention for adults with insomnia. *Arch Gen Psychiatry* 2009 Jul;66(7):692-698 [[FREE Full text](#)] [doi: [10.1001/archgenpsychiatry.2009.66](#)] [Medline: [19581560](#)]
20. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. *Am J Prev Med* 2016 Nov;51(5):843-851 [[FREE Full text](#)] [doi: [10.1016/j.amepre.2016.06.008](#)] [Medline: [27745684](#)]
21. Iida M, Shrout P, Laurenceau J, Bolger N. *Using Diary Methods in Psychological Research*. Washington, DC: American Psychological Association; 2012.
22. Nottorf F, Mastel A, Funk B. The user-journey in online search - an empirical study of the generic-to-branded spillover effect based on user-level data. In: DCNET, ICE-B and OPTICS. 2012 Presented at: DIO'12; July 24-27, 2012; Rome, Italy p. 145-154. [doi: [10.5220/0004052101450154](#)]
23. Chatterjee P, Hoffman DL, Novak TP. Modeling the Clickstream: Implications for Web-Based Advertising Efforts. *Mark Sci* 2003;22(4):520-541 [[FREE Full text](#)]
24. Stange M, Funk B. How Much Tracking Is Necessary - The Learning Curve in Bayesian User Journey Analysis. In: European Conference on Information Systems. 2015 Presented at: ECIS'15; November 29, 2015; Münster, Germany. [doi: [10.18151/7217484](#)]
25. van Breda W, Pastor J, Hoogendoorn M, Ruwaard J, Asselbergs J, Riper H. Exploring and Comparing Machine Learning Approaches for Predicting Mood Over Time. In: KES Conference on Innovation in Medicine and Healthcare. 2016 Presented at: IMH'16; June, 2016; Tenerife, Spain p. 37. [doi: [10.1007/978-3-319-39687-3\\_4](#)]
26. Sen A, Dacin P, Pattichis C. Current Trends in Web Data Analysis. *ACM Digital Library*. 2006. URL: <http://dl.acm.org/citation.cfm?id=1167842>
27. Jaques N, Rudovic O, Taylor S, Sano A, Picard R. Predicting Tomorrow's Mood, Health, and Stress Level using Personalized Multitask Learning and Domain Adaptation. *Proceedings of Machine Learning Research* 2017;66:17-33 [[FREE Full text](#)]
28. Becker D, Bremer V, Funk B, Asselbergs J, Riper H, Ruwaard J. How to Predict Mood: Delving into Features of Smartphone-Based Data. In: European Conference on Information Systems. 2016 Presented at: ECIS'16; September 1, 2016; San Diego, USA.
29. Bremer V, Becker D, Kolovos S, Funk B, van Breda W, Hoogendoorn M, et al. Predicting therapy success and costs for personalized treatment recommendations using baseline characteristics: data-driven analysis. *J Med Internet Res* 2018 Aug 21;20(8):e10275. [doi: [10.2196/10275](#)] [Medline: [30131318](#)]
30. van Breda W, Bremer V, Becker D, Hoogendoorn M, Funk B, Ruwaard J, et al. Predicting therapy success for treatment as usual and blended treatment in the domain of depression. *Internet Interv* 2018;12:100-104 [[FREE Full text](#)] [doi: [10.1016/j.invent.2017.08.003](#)] [Medline: [29862165](#)]
31. van Breda W, Hoogendoorn M, Eiben A, Andersson G, Riper H, Ruwaard J, et al. A Feature Representation Learning Method for Temporal Datasets. In: 2016 IEEE Symposium Series on Computational Intelligence. 2016 Presented at: SSCI'16; December 6-9, 2016; Athens, Greece p. 1-8. [doi: [10.1109/ssci.2016.7849890](#)]
32. A Language and Environment for Statistical Computing. R Core Team. 2018. URL: <https://www.r-project.org/> [accessed 2020-01-01]
33. Bremer V. UJ-Analysis. Github Repos. URL: <https://github.com/VBremer/UJ-Analysis> [accessed 2020-01-01]
34. Gosling JA, Glozier N, Griffiths K, Ritterband L, Thorndike F, Mackinnon A, et al. The GoodNight study--online CBT for insomnia for the indicated prevention of depression: study protocol for a randomised controlled trial. *Trials* 2014 Feb 13;15:56 [[FREE Full text](#)] [doi: [10.1186/1745-6215-15-56](#)]
35. Kotsiantis SB. Supervised Machine Learning: A Review of Classification Techniques. *Informatica* 2007;31(3):249-268.

36. Batista GEAPA, Monard MC. An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence* 2003 May;17(5-6):519-533. [doi: [10.1080/713827181](https://doi.org/10.1080/713827181)]
37. Domingos P. A few useful things to know about machine learning. *Commun ACM* 2012;55(10):78. [doi: [10.1145/2347736.2347755](https://doi.org/10.1145/2347736.2347755)] [Medline: [1000183096](https://pubmed.ncbi.nlm.nih.gov/1000183096/)]
38. Kanter JM, Veeramachaneni K. Deep Feature Synthesis: Towards Automating Data Science Endeavors. In: *IEEE International Conference on Data Science and Advanced Analytics*. 2015 Presented at: DSAA'15; October 19-21, 2015; Paris, France p. 1-10.
39. Khurana U, Nargesian F, Samulowitz H, Khalil E, Turaga D. Automating Feature Engineering. In: *NIPS workshop*. 2016 Presented at: NIPS'16; 5-10 December, 2016; Barcelona, Spain.
40. Lam H, Thiebaut JM, Sinn M, Chen B, Mai T, Alkan O. One button machine for automating feature engineering in relational databases. *arxiv*. 2017. URL: <https://arxiv.org/abs/1706.00327> [accessed 2018-06-10]
41. Cheng W, Kasneci G, Graepel T, Stern D, Herbrich R. Automated Feature Generation From Structured Knowledge. In: *Conference on Information and Knowledge Management*. 2011 Presented at: CIKM'11; October 11, 2011; Glasgow, Scotland, UK p. 1395-1404. [doi: [10.1145/2063576.2063779](https://doi.org/10.1145/2063576.2063779)]
42. Lu X, Lin Z, Jin H, Yang J, Wang J. RAPID: Rating Pictorial Aesthetics using Deep Learning. In: *Proceedings of the ACM International Conference on Multimedia*. 2014 Presented at: ACM'14; November, 2014; Orlando, Florida, USA p. 457-466. [doi: [10.1145/2647868.2654927](https://doi.org/10.1145/2647868.2654927)]
43. Marcus AH, Elias RW. Some useful statistical methods for model validation. *Environ Health Perspect* 1998;106:1541-1550. [doi: [10.1289/ehp.98106s61541](https://doi.org/10.1289/ehp.98106s61541)]
44. Arboretti R, Salmaso L. Model performance analysis and model validation in logistic regression. *Statistica* 2003;63(2):375-396. [doi: [10.6092/issn.1973-2201/358](https://doi.org/10.6092/issn.1973-2201/358)]
45. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters* 2006 Jun;27(8):861-874. [doi: [10.1016/j.patrec.2005.10.010](https://doi.org/10.1016/j.patrec.2005.10.010)]
46. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 2018 Dec 05;58(1):267-288. [doi: [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x)]
47. Thorndike F, Saylor D, Bailey E, Gonder-Frederick L, Morin C, Ritterband L. Development and Perceived Utility and Impact of an Internet Intervention for Insomnia. *EJAP* 2008 Dec 23;4(2):32-42. [doi: [10.7790/ejap.v4i2.133](https://doi.org/10.7790/ejap.v4i2.133)]
48. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a Web-Based Cognitive Behavior Therapy for Insomnia Intervention With 1-Year Follow-up: A Randomized Clinical Trial. *JAMA Psychiatry* 2017 Jan 01;74(1):68-75. [doi: [10.1001/jamapsychiatry.2016.3249](https://doi.org/10.1001/jamapsychiatry.2016.3249)]
49. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep* 2012 Feb 01;35(2):287-302 [FREE Full text] [doi: [10.5665/sleep.1642](https://doi.org/10.5665/sleep.1642)] [Medline: [22294820](https://pubmed.ncbi.nlm.nih.gov/22294820/)]
50. Lundberg S, Lee S. A Unified Approach to Interpreting Model Predictions. In: *Neural Information Processing Systems*. 2017 Presented at: NIPS'17; December 4-9, 2017; Long Beach, USA p. 4765-4774.

## Abbreviations

**AUC:** area under the curve  
**CBT-I:** cognitive behavioral therapy for insomnia  
**EMA:** Ecological Momentary Assessment  
**KNN:** k-nearest neighbor  
**PRAUC:** area under the precision-recall curve  
**ROC:** receiver operating characteristic  
**SHAP:** SHapley Additive exPlanation  
**SHUTi:** Sleep Healthy Using the Internet  
**SOL:** sleep onset latency  
**WASO:** wake after sleep onset

*Edited by G Eysenbach; submitted 09.01.20; peer-reviewed by J Wolff, N Jacobson, ZSY Wong, J Oldenburg; comments to author 11.07.20; revised version received 03.09.20; accepted 20.09.20; published 28.10.20.*

*Please cite as:*

Bremer V, Chow PI, Funk B, Thorndike FP, Ritterband LM

*Developing a Process for the Analysis of User Journeys and the Prediction of Dropout in Digital Health Interventions: Machine Learning Approach*

*J Med Internet Res* 2020;22(10):e17738

URL: <http://www.jmir.org/2020/10/e17738/>

doi: [10.2196/17738](https://doi.org/10.2196/17738)

PMID: [33112241](https://pubmed.ncbi.nlm.nih.gov/33112241/)

©Vincent Bremer, Philip I Chow, Burkhardt Funk, Frances P Thorndike, Lee M Ritterband. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Effectiveness of a Participatory and Interactive Virtual Reality Intervention in Patients With Social Anxiety Disorder: Longitudinal Questionnaire Study

Hyun-Jin Kim<sup>1,2\*</sup>, MD; Seulki Lee<sup>2\*</sup>, MA; Dooyoung Jung<sup>3</sup>, MD, PhD; Ji-Won Hur<sup>4</sup>, PhD; Heon-Jeong Lee<sup>5</sup>, MD, PhD; Sungkil Lee<sup>6</sup>, PhD; Gerard J Kim<sup>7</sup>, PhD; Chung-Yean Cho<sup>8</sup>, MFA; Seungmoon Choi<sup>9</sup>, PhD; Seung-Moo Lee<sup>8</sup>, MFA; Chul-Hyun Cho<sup>1,2</sup>, MD, PhD

<sup>1</sup>Department of Psychiatry, College of Medicine, Chungnam National University, Daejeon, Republic of Korea

<sup>2</sup>Department of Psychiatry, Chungnam National University Sejong Hospital, Sejong, Republic of Korea

<sup>3</sup>Department of Human Factors Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

<sup>4</sup>Department of Psychology, Korea University, Seoul, Republic of Korea

<sup>5</sup>Department of Psychiatry, College of Medicine, Korea University, Seoul, Republic of Korea

<sup>6</sup>Department of Software, Sungkyunkwan University, Suwon, Republic of Korea

<sup>7</sup>Digital Experience Laboratory, Department of Computer Science and Engineering, Korea University, Seoul, Republic of Korea

<sup>8</sup>Department of Film & Multimedia, Korea National University of Arts, Seoul, Republic of Korea

<sup>9</sup>Department of Computer Science and Engineering, Pohang University of Science and Technology, Pohang, Republic of Korea

\*these authors contributed equally

**Corresponding Author:**

Chul-Hyun Cho, MD, PhD

Department of Psychiatry

College of Medicine

Chungnam National University

82 Munhwa-ro, Jung-gu

Daejeon, 35015

Republic of Korea

Phone: 82 449954775

Email: [david0203@gmail.com](mailto:david0203@gmail.com)

## Abstract

**Background:** Social anxiety disorder (SAD) is characterized by excessive fear of negative evaluation and humiliation in social interactions and situations. Virtual reality (VR) treatment is a promising intervention option for SAD.

**Objective:** The purpose of this study was to create a participatory and interactive VR intervention for SAD. Treatment progress, including the severity of symptoms and the cognitive and emotional aspects of SAD, was analyzed to evaluate the effectiveness of the intervention.

**Methods:** In total, 32 individuals with SAD and 34 healthy control participants were enrolled in the study through advertisements for online bulletin boards at universities. A VR intervention was designed consisting of three stages (introduction, core, and finishing) and three difficulty levels (easy, medium, and hard) that could be selected by the participants. The core stage was the exposure intervention in which participants engaged in social situations. The effectiveness of treatment was assessed through Beck Anxiety inventory (BAI), State - Trait Anxiety Inventory (STAI), Internalized Shame Scale (ISS), Post-Event Rumination Scale (PERS), Social Phobia Scale (SPS), Social Interaction Anxiety Scale (SIAS), Brief-Fear of Negative Evaluation Scale (BFNE), and Liebowitz Social Anxiety Scale (LSAS).

**Results:** In the SAD group, scores on the BAI ( $F=4.616$ ,  $P=.009$ ), STAI-Trait ( $F=4.670$ ,  $P=.004$ ), ISS ( $F=6.924$ ,  $P=.001$ ), PERS-negative ( $F=1.008$ ,  $P<.001$ ), SPS ( $F=8.456$ ,  $P<.001$ ), BFNE ( $F=6.117$ ,  $P=.004$ ), KSAD ( $F=13.259$ ,  $P<.001$ ), and LSAS ( $F=4.103$ ,  $P=.009$ ) significantly improved over the treatment process. Compared with the healthy control group before treatment, the SAD group showed significantly higher scores on all scales ( $P<.001$ ), and these significant differences persisted even after treatment ( $P<.001$ ). In the comparison between the VR treatment responder and nonresponder subgroups, there was no significant difference across the course of the VR session.



**Conclusions:** These findings indicated that a participatory and interactive VR intervention had a significant effect on alleviation of the clinical symptoms of SAD, confirming the usefulness of VR for the treatment of SAD. VR treatment is expected to be one of various beneficial therapeutic approaches in the future.

**Trial Registration:** Clinical Research Information Service (CRIS) KCT0003854; [https://cris.nih.go.kr/cris/search/search\\_result\\_st01.jsp?seq=13508](https://cris.nih.go.kr/cris/search/search_result_st01.jsp?seq=13508)

(*J Med Internet Res* 2020;22(10):e23024) doi:[10.2196/23024](https://doi.org/10.2196/23024)

## KEYWORDS

anxiety; social anxiety disorder; virtual reality; intervention; effectiveness; questionnaires

## Introduction

Social anxiety disorder (SAD) is characterized by excessive fear of negative evaluation and humiliation in social interactions (eg, meeting unfamiliar people) and situations (eg, being observed while eating or drinking, and performing in front of others) [1]. Individuals with SAD avoid expressing opinions, talking to people, and forming friendships with colleagues. People with high social anxiety tend to negatively interpret ambiguous social information [2]. Therefore, individuals with SAD show negative interpersonal behavior, including conflict and emotional avoidance. This weakens interpersonal relationships and ultimately leads to social isolation [3]. In particular, SAD can impair work, research, and social life [4,5], as well as reduce well-being [6].

Contemporary theories of SAD emphasize the role of cognitive processes in maintenance of the disorder [7]. Individuals with SAD feel fear and anxiety about being embarrassed in social situations. Cognitive aspects of social anxiety include recalling past experiences of failure and having postevent negative ruminations, both of which exacerbate anxiety by negatively predicting future social events [8]. In addition, individuals with SAD are very sensitive to negative evaluation and social rejection [1]. People with internalized shame are likely to experience anxiety in interpersonal relationships, as this is based on awareness of negative evaluations from others [9].

Cognitive behavior therapy (CBT) is an effective treatment that targets the characteristics of SAD. CBT supposes that when people who experience social anxiety are exposed to socially threatening situations, negative thoughts are automatically evoked, triggering unstable behaviors, emotions, and physical reactions [10,11]. CBT helps identify the unhealthy core beliefs and rigid personal rules that contribute to social anxiety, and then provides various skills and strategies to test and weaken unhealthy attitudes, and to develop and strengthen alternatives. However, people with social anxiety symptoms have high barriers to seeking help from experts due to fear of the stigma surrounding mental health [12]. Additionally, patients with SAD are often not treated as the cost of treatment is high, and it is difficult to access information about professional treatments and services available to individuals. Moreover, the waiting time for treatment is long and the treatment barrier is high due to limited access to specialized services [13]. Therefore, the accessibility of CBT via mobile and desktop computers has been increasing [14], and the scope of treatment through virtual reality (VR) has been expanding recently [15].

VR creates a virtual environment that is blocked from the outside, providing the feeling of being in a new life-sized, computer-generated environment in which one can be immersed. Stereo audio, hand controllers, and eye trackers can be used to create a much more immersive experience. Safety, cost effectiveness, and convenience are advantages to using VR in psychiatric settings. VR can provide exposure treatment to individuals in a safer way than actual exposure, allowing for a quick response and change of stimulus factors if patients experience difficulty. It also reduces the time and costs involved in real exposure treatment [15].

Research has demonstrated the effectiveness of VR for several psychiatric conditions such as anxiety disorder, eating disorder, posttraumatic stress disorder, fear of misconduct, and arachnophobia [16]. VR therapy has an advantage of alleviating the burden of exposure treatment as it is difficult to configure specific, appropriate in vivo exposure conditions for people with SAD [17]. The effectiveness of the therapy can be enhanced through participatory VR because the patients can participate directly with the controllers.

In recent VR treatment research, an environment such as a virtual street, bus, or cafe was created. The results of the research demonstrated less anxiety and paranoia about social encounters in everyday life. In addition, patients reported social interaction anxiety, reduced depressive symptoms, and improved quality of life after VR treatment [18]. In a VR treatment study that presented scenarios such as speaking in front of an audience in a conference room, interviewing, self-introduction, and talking with relatives in an apartment, this intervention was effective in alleviating symptoms of social anxiety [19]. Interactive and participatory VR therapies have the advantage of providing high-immersion situations for participants. The incorporation of interactive virtual scenarios into VR therapy might more adequately target the idiosyncratic fears of participants with SAD.

The purpose of this study was to create a participatory and interactive VR intervention for the treatment of SAD and to evaluate the effectiveness of the intervention as the treatment progressed. Most VR treatments offer a variety of scenarios to individuals with SAD, but each scenario has limitations in that they are not tailored to the patient. In this study, the VR intervention consisted of scenarios in which the participant makes a presentation and the response of others varied according to the level of difficulty. This scenario allows for the provision of patient-specific VR treatment, and is configured to be more immersive and participatory. The VR scenarios were created

considering patients with SAD as the target audience. The intervention was assessed in young adults with SAD, specifically comparing the characteristic features of SAD from baseline to after several intervention sessions.

## Methods

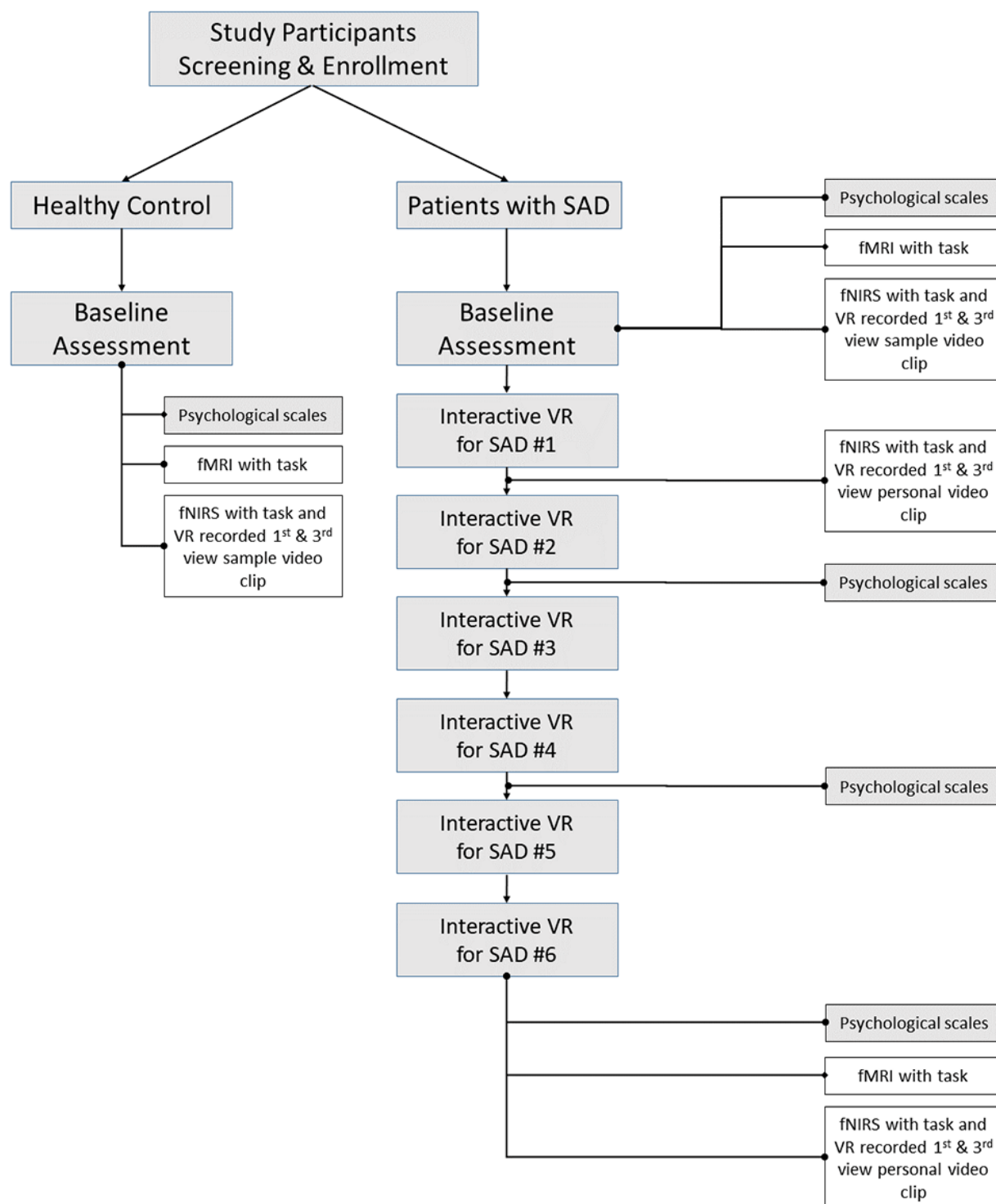
### Participants

We recruited individuals for the SAD and healthy control groups through advertisements for online bulletin boards at universities. The inclusion criteria for the SAD group were as follows: (1) Korean-speaking men or women between the ages of 19 and 31 years, (2) met the Diagnostic and Statistical Manual of Mental Disorders-IV criteria for SAD assessed by Mini-International Neuropsychiatric Interview [20], (3) individuals who were psychotropic medication-naïve without a psychiatric comorbidity (excluding depressive disorder and panic disorder), (4) not currently receiving psychotherapy, (5) no current medical or neurological diagnoses, (6) no history of psychotic symptoms vulnerable to a VR experience, and (7) not vulnerable to visual stimuli such as epilepsy. Exclusion criteria were as follows: (1) previous history of intellectual disability or organic brain damage, (2) experienced psychotic symptoms vulnerable to VR experiences, (3) vulnerable to visual stimuli such as epilepsy, and (4) unsuitable for participation in magnetic resonance imaging research. Healthy control participants had

no other neurological or psychiatric diagnoses. To determine whether the intervention had a clear, positive effect on individuals with SAD, only participants who had a score of 82 or higher on the Korean Social Avoidance and Distress Scale (KSAD) were enrolled [21]. The KSAD is a measure of the degree of experiencing anxiety in social situations and the tendency to avoid, and consists of 28 items assessed on a 5-point scale. Potential participants completed the KSAD online before enrollment. After the enrollment of participants who had a score of 82 or higher on the KSAD, we were able to collect data from a more homogeneous group of individuals with SAD who clearly had social anxiety.

A total of 40 patients with SAD and 34 healthy participants were enrolled in the study. Of these, 8 patients with SAD and 1 healthy control participant dropped out for personal reasons (eg, time constraints). Thus, a total of 32 patients with SAD and 33 healthy control participants completed this study. This study was part of a larger project that was conducted to evaluate the effects of interactive and participatory VR solutions using psychological scales, functional near-infrared spectroscopy, functional magnetic resonance imaging, and several physiological signals. An overview of the entire study is presented in Figure 1, and among these, only the psychological scale results were analyzed for this study to reduce distraction and clarify the subject. Other results will be presented in subsequent articles.

**Figure 1.** Overview of research flow chart for the entire project of participatory and interactive virtual reality (VR) treatment in patients with social anxiety disorder (SAD). This study reports the results of the analyzed data from the psychological scales. The content in the unshaded boxes represents different studies that will be presented in other papers. fMRI: functional magnetic resonance imaging; fNIRS: functional near-infrared spectroscopy.



The study was approved by the Korea University Anam Hospital Institutional Review Board and was conducted in accordance with the Declaration of Helsinki. All participants provided informed written consent after explanation of the study procedures.

## Participatory and Interactive VR Intervention for SAD

### Composition and Content of the VR Intervention

The participatory and interactive VR intervention for the control of social anxiety symptoms consisted of three stages (introduction, core, and finishing stages) and was divided into three levels (easy, medium, and hard) according to the difficulty of the core stage content. The VIVE (HTC Corporation, Taiwan)

VR headset was used for the intervention, and the heart rate, skin tension, and eye movement of the participant during the VR experience were measured.

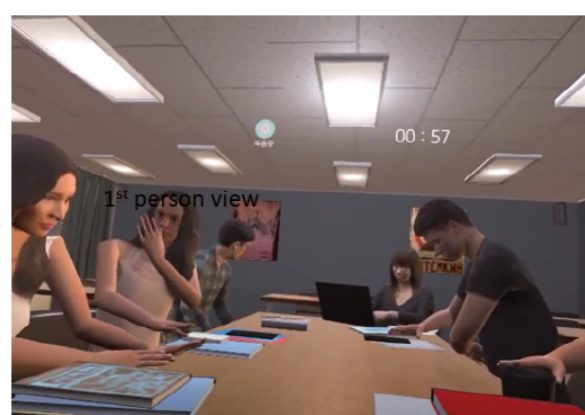
In the introduction stage, participants were invited to select their own avatar and learn how to use the VR interface. To help participants adapt to VR and calm their mind, they partook in a meditation-based warm-up session. The core stage was the exposure intervention in which participants engaged in social situations. The situation was to enter a room where they were going to meet several other college students who were to introduce themselves to each other. In the virtual situation, 7 to 8 nonplaying characters appeared and introduced themselves (Figure 2). Once they finished their introductions, the participant

with SAD pressed the record button on the screen to introduce themselves. The nonplaying characters in the scenario listened to the participant's introduction, and the level of difficulty (easy, medium, and hard) was determined according to the nonplaying characters' attitude and degree of reaction to the participant. As the level of difficulty increased, the attitudes of the nonplaying characters who listened to the participant's introduction changed, in that they became more distracted and made small talk among themselves. At the hard level, one of the nonplaying characters challenged the participant when they were introducing himself/herself by saying, "Please introduce yourself properly." The finishing stage presented general cognitive and behavioral safety guidelines for SAD in both voice and text form in VR [22].

**Figure 2.** Screenshots of participatory and interactive virtual reality (VR) treatment in patients with social anxiety disorder, shown from the first- and third-person views [22].



First-person view



Third-person view

### Number of VR Sessions and Procedure

The VR intervention was designed for participants to perform a total of 6 sessions. Participants were allowed to perform 2 sessions in a row in a single visit, and the first session was started at an easy level. During the second session, participants could select and proceed to their desired level. It was explained that participants could stop at any time during the VR experience. Researchers were present throughout the VR experience to deal with any unexpected situations.

Participants completed a battery of assessments to evaluate the psychological state before and after the therapeutic VR sessions. Participants answered the self-reported psychological scales four times: at baseline (before the VR experience), after the second VR session, after the fourth session, and after termination (ie, after the sixth session).

### Assessments

#### Beck Anxiety Inventory

The Beck Anxiety Inventory (BAI) [23] measures the occurrence and severity of anxiety symptoms. The BAI consists of 21 items and each answer is scored on a scale of 0 (not at all) to 3 (severely). Higher total scores indicate more severe anxiety symptoms. We used the Korean version of the BAI (K-BAI) [24].

### State-Trait Anxiety Inventory

The State-Trait Anxiety Inventory (STAI) [25] was designed to measure state and trait anxiety in research and clinical practice. State anxiety is defined as a transient, momentary emotional status that results from situational stress, whereas trait anxiety represents a predisposition to react with anxiety in stressful situations [25]. The STAI consists of two subscales for measuring the intensity of anxiety as an emotional state (20 items) and individual differences in anxiety proneness as a personality trait (20 items). Each answer is scored on a scale of 1 (not at all) to 4 (almost always/very much so). We used a Korean version of the STAI (K-STAI) [26].

### Social Phobia Scale

The Social Phobia Scale (SPS) [27] was designed to measure the level of anxiety and fear in various social performance situations. The SPS consists of 20 items and each answer is scored on a scale of 0 (not at all) to 4 (extremely). We used the Korean version of the SPS (K-SPS) [28].

### Social Interaction Anxiety Scale

The Social Interaction Anxiety Scale (SIAS) [27] was designed to measure fear of social interaction situations such as meeting and talking with other people. Each question is presented in the form of a self-describing statement that describes cognitive, emotional, and behavioral responses in various social interaction situations. We used the Korean-Social Interaction Anxiety Scale



(K-SIAS) [28,29]. The K-SIAS consists of 20 items and each answer is scored on a scale of 0 (not at all) to 4 (extremely).

### **Brief-Fear of Negative Evaluation Scale**

The Brief-Fear of Negative Evaluation Scale (BFNE) measures the fear of being negatively evaluated by other people. The original, longer scale was developed by Watson and Friend [30], while the short-form version selected only items that had a correlation of 0.50 or more with the overall score. The Korean version (K-BFNE) developed by Lee and Choi [21] was used in this study. The K-BFNE consists of 12 items and each answer is scored on a scale of 1 (strongly disagree) to 5 (strongly agree).

### **Internalized Shame Scale**

The Internalized Shame Scale (ISS) [31] assesses shame proneness and internalized shame. The ISS has 30 items that consist of two basic scales for shame (24 items) and self-esteem (6 items). Shame-related questions evaluate the extent of shame that becomes magnified and internalized. We used the Korean version of the ISS (K-ISS) [32]. A factor analysis of the K-ISS identified a 4-factor structure: inadequacy (10 items), emptiness (5 items), self-punishment (5 items), and fear of mistakes (4 items).

### **Post-Event Rumination Scale**

The Post-Event Rumination Scale (PERS) [33,34] was designed to measure the frequency of postevent ruminations in social situations. The PERS comprises two scales including negative rumination (15 items) and positive rumination (9 items). Each answer is scored on a scale of 0 (never) to 4 (very often); higher scores indicate more frequent rumination. We used the Korean version of the PERS (K-PERS) [35].

### **Liebowitz Social Anxiety Scale**

The Liebowitz Social Anxiety Scale (LSAS) [36] assesses the degree of anxiety and avoidance in several typical social and performance situations. The LSAS consists of 24 items, which are answered for the degree of fear or anxiety (0, none to 3, severe) and avoidance (0, never to 3, usually). Higher total

scores indicate more severe social anxiety symptoms. We used the Korean version of the LSAS (K-LSAS) [37].

### **Statistical Analysis**

To examine the difference in categorical and continuous variables between the 2 groups (SAD and healthy control groups), we used the  $\chi^2$  test and the independent  $t$  test, respectively. To compare differences in the scores on the psychological scales between the SAD group and healthy control group at baseline and after termination,  $t$  tests were conducted. Repeated-measures analysis of variance (ANOVA) was conducted to investigate changes in the scores of the psychological scales in participants with SAD over the course of the VR intervention. All statistical analyses were performed using SPSS software version 24.0 (IBM Corp, Armonk, NY, USA). The significance level was set at  $P<.05$  (two-tailed).

## **Results**

Demographic information showed that the mean age of the SAD ( $n=32$ ) and healthy control ( $n=33$ ) participants was 23.12 (SD 3.12) and 23.55 (SD 3.38) years, respectively, with no significant difference ( $P=.60$ ). There was also no significant intergroup difference in gender distribution ( $P=.16$ ), with 11 males (34%) and 21 females (66%) in the SAD group, and 17 males (52%) and 16 females (49%) in the healthy control group. The mean education duration of the SAD group was 14.59 years (SD 1.37) and that of the healthy control group was 14.94 years (SD 1.77), which was not significantly different ( $P=.38$ ).

The psychological scales related to SAD were classified into several categories: general anxiety symptoms (BAI, STAI), SAD symptoms (SPS, SIAS, BFNE, KSAD, and LSAS), shame (ISS), and rumination (PERS). The comparison of scores on the psychological scales between the healthy control and SAD groups at baseline is shown in Table 1. For each scale, scores were significantly higher in the SAD group compared with those of the healthy control group (all  $P<.001$ ).



**Table 1.** Comparison of psychological states between the social anxiety disorder (SAD) and healthy control (HC) groups at baseline.

Measures	SAD group (n=32), mean (SD)	HC group (n=33), mean (SD)	<i>t</i>	df	<i>P</i> value
BAI <sup>a</sup>	15.03 (10.46)	3.39 (3.89)	6.16	63	<.001
STAI-S <sup>b</sup>	48.72 (9.78)	35.42 (7.50)	8.45	63	<.001
STAI-T <sup>c</sup>	52.28 (10.50)	33.49 (7.16)	9.82	47.1	<.001
SPS <sup>d</sup>	30.63 (13.69)	5.76 (5.50)	9.71	48.25	<.001
SIAS <sup>e</sup>	42.81 (12.51)	18.30 (6.98)	7.76	63	<.001
BFNE <sup>f</sup>	43.63 (8.99)	27.94 (7.25)	14.71	61.85	<.001
KSAD <sup>g</sup>	107.34 (14.60)	49.03 (17.29)	8.73	63	<.001
LSAS <sup>h</sup>	72.34 (23.54)	21.85 (15.58)	9.96	53.56	<.001
<b>ISS<sup>i</sup></b>					
Total	50.09 (17.30)	16.15 (9.26)	7.84	41.82	<.001
Inadequacy	16.34 (7.98)	4.33 (3.43)	7.32	63	<.001
Emptiness	10.88 (4.81)	3.30 (3.34)	8.27	38.78	<.001
Self-punishment	10.09 (5.10)	2.18 (1.84)	9.99	63	<.001
Fear of mistake	12.78 (2.24)	6.33 (2.91)	4.69	63	<.001
<b>PERS<sup>j</sup></b>					
Negative rumination	31.38 (10.88)	9.52 (7.37)	−4.87	53.23	<.001
Positive rumination	15.06 (6.28)	25.33 (10.29)	9.55	40.50	<.001

<sup>a</sup>BAI: Beck Anxiety Inventory.<sup>b</sup>STAI-S: State-Trait Anxiety Inventory–State.<sup>c</sup>STAI-T: State-Trait Anxiety Inventory–Trait.<sup>d</sup>SPS: Social Phobia Scale.<sup>e</sup>SIAS: Social Interaction Anxiety Scale.<sup>f</sup>BFNE: Brief-Fear of Negative Evaluation Scale.<sup>g</sup>KSAD: Korean Social Avoidance and Distress Scale.<sup>h</sup>LSAS: Liebowitz Social Anxiety Scale.<sup>i</sup>ISS: Internalized Shame Scale.<sup>j</sup>PERS: Post-Event Rumination Scale.

## Changes in Social Anxiety

Repeated-measures ANOVA was conducted to assess changes in scores from baseline on the psychological scales in the participants with SAD after 2, 4, and 6 VR sessions. As shown in [Table 2](#), general anxiety symptoms as measured by the BAI and STAI-T significantly improved after treatment, whereas the STAI-S did not improve significantly. SPS, SIAS, KSAD,

BFNE, and LSAS, which are measures to evaluate the symptoms of SAD, all showed significant improvement after VR. The ISS showed significant improvement on the overall scale, and on the emptiness, self-punishment, and fear of mistakes subscales. There was no significant difference in inadequacy. Negative rumination, which is a subscale of the PERS, showed a significant improvement after the VR treatment, but positive rumination did not ([Multimedia Appendix 1, Table 2](#)).

**Table 2.** Changes in psychological states of participants with social anxiety disorder.

Measures	Baseline, mean (SD)	Session 2, mean (SD)	Session 4, mean (SD)	Session 6, mean (SD)	F statistic (df)	P value
BAI <sup>a</sup>	15.03 (10.46)	11.97 (9.33)	12.94 (11.29)	9.94 (9.58)	4.616 (2.348)	.009
STAI-S <sup>b</sup>	48.72 (9.78)	48.25 (10.14)	49.44 (11.97)	46.25 (9.96)	1.264 (2.684)	.29
STAI-T <sup>c</sup>	52.28 (10.50)	49.88 (11.02)	51.81 (10.97)	47.56 (11.57)	4.670 (2.556)	.004
SPS <sup>d</sup>	30.63 (13.69)	27.16 (13.61)	25.22 (13.70)	23.00 (13.03)	8.456 (2.400)	<.001
SIAS <sup>e</sup>	42.81 (12.51)	41.06 (11.86)	37.44 (12.34)	35.47 (12.83)	13.155 (2.734)	<.001
BFNE <sup>f</sup>	43.63 (8.99)	39.31 (11.00)	38.75 (7.72)	37.59 (7.29)	6.117 (2.027)	.004
KSAD <sup>g</sup>	107.34 (14.60)	104.16 (14.28)	99.66 (13.37)	97.69 (12.55)	13.259 (2.369)	<.001
LSAS <sup>h</sup>	71.34 (4.16)	74.03 (4.77)	68.28 (4.13)	64.00 (4.11)	4.103 (2.503)	.009
<b>ISS<sup>i</sup></b>						
Overall	50.09 (17.30)	43.91 (17.40)	43.13 (18.34)	38.78 (16.85)	6.924 (2.372)	.001
Inadequacy	16.34 (7.98)	14.75 (8.38)	14.72 (8.22)	13.44 (7.81)	2.604 (2.719)	.06
Emptiness	10.88 (4.81)	8.81 (5.52)	8.81 (5.03)	7.69 (4.47)	5.152 (2.769)	.002
Self-punishment	10.09 (5.09)	9.16 (4.68)	8.41 (4.76)	7.50 (4.33)	5.528 (2.194)	.005
Fear of mistake	12.78 (2.24)	11.19 (2.42)	11.25 (3.20)	10.16 (3.05)	10.891 (2.662)	<.001
<b>PERS<sup>j</sup></b>						
Negative rumination	31.38 (10.88)	27.03 (10.97)	25.81 (10.78)	23.41 (11.10)	6.974 (2.730)	<.001
Positive rumination	15.06 (6.29)	13.13 (5.84)	14.47 (6.73)	13.56 (6.83)	1.008 (2.812)	.39

<sup>a</sup>BAI: Beck Anxiety Inventory.<sup>b</sup>STAI-S: State-Trait Anxiety Inventory–State.<sup>c</sup>STAI-T: State-Trait Anxiety Inventory–Trait.<sup>d</sup>SPS: Social Phobia Scale.<sup>e</sup>SIAS: Social Interaction Anxiety Scale.<sup>f</sup>BFNE: Brief-Fear of Negative Evaluation Scale.<sup>g</sup>KSAD: Korean Social Avoidance and Distress Scale.<sup>h</sup>LSAS: Liebowitz Social Anxiety Scale.<sup>i</sup>ISS: Internalized Shame Scale.<sup>j</sup>PERS: Post-Event Rumination Scale.

Posthoc analyses for pairwise comparisons between baseline to sessions 2, 4, and 6 were performed by the Bonferroni method. The scales of BAI ( $P=.003$ ), and overall ( $P=.01$ ), emptiness ( $P=.001$ ), and self-punishment ( $P=.001$ ) of the ISS showed significant differences between baseline and session 6. The scales of PERS ( $P=.01$ ), SPS ( $P=.005$ ), SIAS ( $P<.001$ ), BFNE ( $P=.01$ ), and KSAD ( $P<.001$ ) showed significant differences between baseline and session 4. The fear mistake subscale of the ISS ( $P=.005$ ) showed a significant difference between baseline and session 2 ([Multimedia Appendix 1](#)).

### Changes in Psychological States Between Responder and Nonresponder Subgroups After VR Treatment

In general, the response to treatment is defined as a reduction of 50% or more from the baseline psychological scale. We

conducted a repeated-measures ANOVA for comparative analysis by dividing the participants into subgroups: those with a 50% or more reduction of the BAI scale value compared to the baseline were defined as the treatment responder group, and those with a less than 50% reduction were classified as the nonresponder group.

As presented in [Table 3](#), all psychological scale values except for STAI-S, positive subscale of the PERS, and inappropriate subscale of the ISS showed significant changes across the VR treatment session in both subgroups. A significant session-by-group interaction was found for the BAI value; however, no session-by-group interaction was observed for the other psychological scales.

**Table 3.** Changes in psychological states between responder and nonresponder subgroups according to virtual reality treatment.<sup>a</sup>

Measures	Responders (n=14), mean (SD)	Nonresponders (n=18), mean (SD)	Main effect of session		Session-by-group interaction	
			<i>F</i> (df)	<i>P</i> value	<i>F</i> (df)	<i>P</i> value
<b>BAI<sup>b</sup></b>			6.021 (2.446)	.002	4.684 (2.446)	.008
Baseline	12.93 (10.28)	16.68 (10.60)				
Session 2	8.71 (7.79)	14.50 (9.84)				
Session 4	6.57 (6.05)	17.89 (12.04)				
Session 6	3.21 (4.04)	15.17 (9.41)				
<b>STAI-S<sup>c</sup></b>			1.178 (2.737)	.32	1.087 (2.737)	.36
Baseline	45.64 (8.39)	51.11 (10.32)				
Session 2	44.79 (8.53)	50.94 (10.70)				
Session 4	43.14 (6.21)	54.33 (13.18)				
Session 6	41.93 (6.22)	49.61 (11.14)				
<b>STAI-T<sup>d</sup></b>			4.875 (2.515)	.006	1.675 (2.515)	.19
Baseline	47.00 (10.06)	56.39 (9.10)				
Session 2	46.00 (10.68)	52.89 (10.58)				
Session 4	57.06 (9.04)	57.06 (9.50)				
Session 6	40.57 (8.93)	53.00 (10.56)				
<b>SPS<sup>e</sup></b>			8.849 (2.407)	<.001	0.983 (2.407)	.39
Baseline	26.00 (15.65)	34.22 (11.11)				
Session 2	23.14 (15.75)	30.28 (11.16)				
Session 4	18.64 (13.18)	30.33 (12.18)				
Session 6	16.71 (12.39)	27.89 (11.58)				
<b>SIAS<sup>f</sup></b>			13.819 (2.659)	<.001	1.418 (2.659)	.25
Baseline	37.21 (12.21)	47.17 (11.21)				
Session 2	36.43 (11.79)	44.67 (10.91)				
Session 4	29.86 (11.55)	43.33 (9.56)				
Session 6	29.21 (10.37)	40.33 (12.67)				
<b>BFNE<sup>g</sup></b>			5.976 (2.052)	.004	0.672 (2.052)	.52
Baseline	41.93 (9.56)	44.94 (8.56)				
Session 2	39.36 (9.96)	39.28 (12.04)				
Session 4	36.64 (7.89)	40.39 (7.39)				
Session 6	35.64 (6.71)	39.11 (7.55)				
<b>KSAD<sup>h</sup></b>			12.854 (2.359)	<.001	0.168 (2.359)	.88
Baseline	102.36 (13.68)	111.22 (14.46)				
Session 2	97.93 (8.57)	109.00 (16.09)				
Session 4	94.29 (12.30)	103.83 (12.95)				
Session 6	91.71 (5.89)	102.33 (14.44)				
<b>LSAS<sup>i</sup></b>			3.024 (2.471)	.04	0.727 (2.471)	.51
Baseline	33.50 (12.73)	42.67 (11.51)				
Session 2	36.29 (15.51)	43.11 (10.42)				

Measures	Responders (n=14), mean (SD)	Nonresponders (n=18), mean (SD)	Main effect of session		Session-by-group interaction	
			<i>F</i> (df)	<i>P</i> value	<i>F</i> (df)	<i>P</i> value
Session 4	31.07 (11.10)	42.72 (10.38)				
Session 6	29.29 (9.43)	39.94 (11.02)				
<b>ISS<sup>j</sup> Overall</b>			6.835 (2.345)	.001	0.309 (2.345)	.77
Baseline	45.79 (18.97)	53.44 (15.60)				
Session 2	40.86 (18.22)	46.28 (16.87)				
Session 4	39.07 (18.12)	46.28 (18.38)				
Session 6	33.00 (15.49)	43.28 (16.89)				
<b>ISS Inadequacy</b>			2.574 (2.698)	.07	0.343 (2.698)	.77
Baseline	14.36 (8.42)	17.89 (7.50)				
Session 2	13.57 (8.65)	15.67 (8.29)				
Session 4	12.86 (8.03)	16.17 (8.30)				
Session 6	11.07 (6.96)	15.28 (8.12)				
<b>ISS Emptiness</b>			5.090 (2.761)	.004	0.350 (2.761)	.77
Baseline	9.86 (5.11)	11.67 (4.55)				
Session 2	8.14 (4.96)	9.33 (6.01)				
Session 4	8.07 (5.03)	9.39 (5.09)				
Session 6	6.14 (3.44)	8.89 (4.89)				
<b>ISS - Self-punishment</b>			5.486 (2.198)	.005	0.143 (2.198)	.88
Baseline	9.43 (5.12)	10.61 (5.16)				
Session 2	8.21 (4.71)	9.89 (4.65)				
Session 4	7.29 (4.43)	9.28 (4.96)				
Session 6	6.43 (4.43)	8.33 (4.17)				
<b>ISS Fear of mistake</b>			10.598 (2.645)	<.001	0.417 (2.645)	.72
Baseline	12.14 (2.45)	13.28 (1.99)				
Session 2	10.93 (2.76)	11.39 (2.17)				
Session 4	10.86 (2.93)	11.56 (3.45)				
Session 6	9.36 (3.32)	10.78 (2.76)				
<b>PERS<sup>k</sup> Negative rumination</b>			6.931 (2.720)	.001	0.354 (2.720)	.77
Baseline	28.14 (10.31)	33.89 (10.93)				
Session 2	24.71 (12.64)	28.83 (9.45)				
Session 4	22.29 (12.00)	28.56 (9.14)				
Session 6	19.00 (8.69)	26.83 (11.77)				
<b>PERS Positive rumination</b>			0.855 (2.822)	.46	0.375 (2.822)	.76
Baseline	15.29 (5.01)	14.89 (7.27)				
Session 2	14.43 (5.69)	12.11 (5.90)				
Session 4	15.64 (6.46)	13.56 (6.97)				

Measures	Responders (n=14), mean (SD)	Nonresponders (n=18), mean (SD)	Main effect of session		Session-by-group interaction	
			<i>F</i> (df)	<i>P</i> value	<i>F</i> (df)	<i>P</i> value
Session 6	15.21 (6.80)	12.28 (6.76)				

<sup>a</sup>Treatment responder and nonresponder subgroups were divided according to the reduction of the BAI scale value by 50% or more compared to the baseline value.

<sup>b</sup>BAI: Beck Anxiety Inventory.

<sup>c</sup>STAI-S: State-Trait Anxiety Inventory–State.

<sup>d</sup>STAI-T: State-Trait Anxiety Inventory–Trait.

<sup>e</sup>SPS: Social Phobia Scale.

<sup>f</sup>SIAS: Social Interaction Anxiety Scale.

<sup>g</sup>BFNE: Brief-Fear of Negative Evaluation Scale.

<sup>h</sup>KSAD: Korean Social Avoidance and Distress Scale.

<sup>i</sup>LSAS: Liebowitz Social Anxiety Scale.

<sup>j</sup>ISS: Internalized Shame Scale.

<sup>k</sup>PERS: Post-Event Rumination Scale.

### Comparison of SAD and Control Groups After VR Treatment

We analyzed whether the scores on the psychological scales differed in the SAD group compared with those of the healthy

control group after 6 VR treatment sessions. Results from independent *t* tests showed that even after completing the VR sessions, the SAD group continued to have significantly higher scores than the healthy control group on all psychological scales (Table 4).



**Table 4.** Comparison of psychological states between the social anxiety disorder (SAD) and healthy control (HC) groups after virtual reality treatment.

Measures	SAD group (n=32), mean (SD)	HC group <sup>a</sup> (n=33), mean (SD)	<i>t</i>	df	<i>P</i> value
BAI <sup>b</sup>	9.94 (9.58)	3.39 (3.89)	3.59	40.70	.001
STAI-S <sup>c</sup>	46.25 (46.25)	35.42 (35.42)	4.96	63	<.001
STAI-T <sup>d</sup>	47.56 (11.57)	33.49 (7.16)	5.92	63	<.001
SPS <sup>e</sup>	23.00 (13.03)	5.76 (5.50)	6.91	41.45	<.001
SIAS <sup>f</sup>	35.47 (12.83)	18.30 (6.98)	6.67	47.54	<.001
BFNE <sup>g</sup>	37.59 (7.29)	27.94 (7.25)	5.35	63	<.001
KSAD <sup>h</sup>	97.69 (12.55)	49.03 (17.29)	13.01	58.43	<.001
LSAS <sup>i</sup>	64.00 (23.22)	21.85 (15.58)	8.62	63	<.001
<b>ISS<sup>j</sup></b>					
Overall	38.78 (16.85)	16.15 (9.26)	6.69	47.85	<.001
Inadequacy	13.43 (7.81)	4.33 (3.43)	6.05	42.28	<.001
Emptiness	7.69 (4.47)	3.30 (3.43)	4.45	63	<.001
Self-punishment	7.50 (4.33)	2.18 (1.84)	6.41	41.65	<.001
Fear of mistake	10.16 (3.05)	6.33 (2.91)	5.17	63	<.001
<b>PERS<sup>k</sup></b>					
Negative rumination	23.41 (11.10)	9.52 (7.37)	5.92	53.66	<.001
Positive rumination	13.56 (6.82)	25.33 (10.29)	-5.45	55.76	<.001

<sup>a</sup>Since the control group did not receive the virtual reality treatment, psychological states of the healthy control group were measured only at baseline.

<sup>b</sup>BAI: Beck Anxiety Inventory.

<sup>c</sup>STAI-S: State-Trait Anxiety Inventory–State.

<sup>d</sup>STAI-T: State-Trait Anxiety Inventory–Trait.

<sup>e</sup>SPS: Social Phobia Scale.

<sup>f</sup>SIAS: Social Interaction Anxiety Scale.

<sup>g</sup>BFNE: Brief-Fear of Negative Evaluation Scale.

<sup>h</sup>KSAD: Korean Social Avoidance and Distress Scale.

<sup>i</sup>LSAS: Liebowitz Social Anxiety Scale.

<sup>j</sup>ISS: Internalized Shame Scale.

<sup>k</sup>PERS: Post-Event Rumination Scale.

## Discussion

### Principal Findings

The aim of this study was to analyze the effectiveness of a newly developed participatory and interactive VR intervention for individuals with SAD. In particular, using psychological tests related to SAD, we evaluated whether the VR intervention was effective in improving various symptoms.

General anxiety symptoms, measured by the BAI and STAI-T, were significantly improved compared with those measured before treatment, whereas state anxiety (ie, STAI-S) was not. State anxiety increases the threat value assigned to a stimulus or situation, and trait anxiety gives rise to a tendency to constantly direct attention toward the source of threat [38]. Any trait that does not change easily over a lifetime is a distinguishing feature of a person's character, and

anxiety-related traits are among the most important risk factors of anxiety disorder [39].

The scales that directly assess the symptoms of SAD showed significant improvement compared with those measured before treatment. SAD can be divided into symptoms that appear in specific situations and general social interactions [27], which were assessed with the SPS and SIAS, both of which showed significant improvement after VR. Scores on the LSAS and KSAD were significantly reduced after treatment. However, the LSAS score increased after session 2, and then decreased in the subsequent two time points (session 4 and session 6). BAI scores decreased in session 2, and then increased in session 4 before decreasing again in session 6. Similarly, the STAI (both state and trait scales) showed a decrease from baseline to session 2, followed by an increase after session 4 and then a decrease after session 6. Although this is only a pattern, it could be speculated that the tendency of anxiety to increase during a VR session is probably due to exposure to a VR session at a stage

in the middle of the intervention process before finally showing a significant therapeutic effect. This is an important consideration, because increased anxiety during a VR session may lead to a problem of dropping out of treatment due to poor treatment adherence after the initial treatment session. Treatment compliance is important in CBT because it requires a certain period of treatment, and exposure techniques can exacerbate symptoms in some cases [40]. In general, VR treatment is considered to have higher compliance than conventional treatment [41], but it is necessary to pay attention to changes in compliance during the treatment sessions.

Cognitive and emotional components are known to be involved in the development of SAD [42]. Scores on the ISS, which is used to evaluate internalized shame, improved significantly as VR treatment progressed, including a significant decrease on almost every ISS subscale (ie, emptiness, self-punishment, and fear of mistakes). Shame is an emotional component related to social anxiety that is based on a negative evaluation of the awareness of others, suggesting a positive relationship between social anxiety and shame [43-45]. A key part of cognitive models of social anxiety is that individuals perceive a negative evaluation by others [7]. In this study, negative rumination significantly improved after treatment, whereas positive rumination was not significantly different. Past studies of rumination have shown a relationship between negative rumination and social anxiety, but the relationship between social anxiety and positive rumination has been inconsistent or was not significant [33,46]. The improvement of negative rumination suggests that this can affect the cognitive structure of SAD.

Posthoc analysis to compare baseline scores to each session time point showed that VR had a significant effect on the fear of mistake subscale of the ISS at the initial point of treatment (after session 2). In comparison, BAI and the overall, emptiness, and self-punishment subscales of the ISS changed significantly only at the end of treatment (after session 6), and PERS, SPS, SIAS, BFNE, and KSAD scales showed significant changes around the middle point of VR treatment (after session 4). We could speculate that the improvement patterns of psychological symptoms may differ across the process of the VR treatment sessions. Since prolonging or shortening the number of VR treatment sessions might affect the therapeutic outcome on each psychological state, future studies are needed to determine how and when psychological symptoms change according to the VR treatment process and to establish the optimal number for VR treatment sessions.

Although we found significant improvements in SAD symptoms after treatment, these symptoms were still significantly different compared with those of the control group. This result shows that the VR intervention had a significant effect in the SAD group, but that the participants with SAD still exhibited prominent symptoms after VR treatment sessions. This suggests that the treatment effect of VR alone might not be sufficient to draw out the treatment response or remission. A meta-analysis of VR treatment focused on anxiety disorders found a small effect size [47]. Therefore, it is necessary to consider changes in treatment methods such as the combination of various treatment techniques, including conventional psychiatric

treatment or more repeated VR trials, to achieve a sufficient therapeutic effect in the real world.

We did not find any significant session-by-group interaction in the majority of the psychological scales except for BAI between the two subgroups of responders and nonresponders (according to changes in the BAI scale). This result can be presumed to be due to the limitation of the methodology in which the SAD group was divided into responder and nonresponder groups according to a single psychological scale. It is challenging to determine a clear cutoff as to what scale or criteria to use for distinguishing responders. Moreover, since this study attempted to collect data from a more homogeneous sample of individuals with social anxiety symptoms ( $KSAD \geq 82$ ), subgroup differences might not be expected. We hope to be able to derive meaningful results through further studies by comparison between subgroups according to the degree of severity based on the scores of different psychological scales, or by a comparative study according to whether or not the individual is also under medication for anxiety.

The results of this study indicated improvements in almost all elements of social anxiety measured, including general anxiety, social anxiety symptoms, and cognitive and emotional aspects of social anxiety, in the SAD group after treatment. Previous studies related to VR treatment for social anxiety have also shown an improvement in symptoms [47-49]; however, this study demonstrated an effect not only on social anxiety symptoms but also on a wide range of related factors, including general anxiety symptoms and cognitive and emotional characteristics.

## Limitations and Strengths

Our results should be interpreted within the context of the study's limitations. First, this study did not have a sham or waitlist control group, which limits interpretation of the results. Second, this study employed self-rated scales that could be confounded by bias (eg, participant motivation). Nonetheless, a comprehensive evaluation related to social anxiety disorder was attempted in analyzing the effectiveness of treatment, including changes during treatment sessions. This study enrolled participants in the patient group who met strict diagnostic criteria and had not received any other treatments, including psychiatric drugs. Given these strengths, the therapeutic effect of a VR intervention could be evaluated more accurately.

## Conclusion

This study investigated the effectiveness of a newly developed participatory and interactive VR intervention in patients with SAD. The results show that a VR intervention can be an effective treatment for various dimensions of SAD. Technology-based treatment in psychiatry is more cost-effective, easier to handle, and more manageable for both the therapist and client. VR treatment is a promising tool in the field of psychiatry, which can simulate situations for patients with anxiety in a safe, controllable, and reproducible way [15,50,51]. Future research should focus on ensuring that the effectiveness of these immersive VR treatments persist after treatment, and that treatments are made more effective through different treatment combinations or changes in techniques.

## Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Government of Korea (MSIT) (NRF-2020R1C1C1007463), and the ICT & Future Planning for Convergent Research in the Development Program for R&D Convergence over Science and Technology Liberal Arts (NRF-2017M3C1B6070978).

## Authors' Contributions

DJ, JWH, and CHC conceived and designed the study. HJK, SL, and CHC performed the statistical analyses. HJK, SL, DJ, JWH, and CHC wrote the first draft of the manuscript. DJ, JWH, HJK, SL, GK, CYC, SC, SML, and CHC participated in data collection. All authors edited all versions of the manuscript. All the authors were involved in the interpretation of the results, and read, commented, and approved the final version of the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Changes in psychological state due to participatory and interactive virtual reality (VR) treatment in patients with social anxiety disorder.

[PNG File , 513 KB - [jmir\\_v22i10e23024\\_app1.png](#)]

## References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5). Washington, DC: American Psychiatric Association Publishing; 2013.
2. Huppert JD, Pasupuleti RV, Foa EB, Mathews A. Interpretation biases in social anxiety: response generation, response selection, and self-appraisals. *Behav Res Ther* 2007 Jul;45(7):1505-1515. [doi: [10.1016/j.brat.2007.01.006](#)] [Medline: [17336922](#)]
3. Davila J, Beck JG. Is social anxiety associated with impairment in close relationships? A preliminary investigation. *Behav Ther* 2002;33(3):427-446. [doi: [10.1016/s0005-7894\(02\)80037-5](#)]
4. Alden LE, Taylor CT. Interpersonal processes in social phobia. *Clin Psychol Rev* 2004 Nov;24(7):857-882. [doi: [10.1016/j.cpr.2004.07.006](#)] [Medline: [15501559](#)]
5. Wittchen H, Fuetsch M, Sonntag H, Müller N, Liebowitz M. Disability and quality of life in pure and comorbid social phobia. Findings from a controlled study. *Eur Psychiatry* 2000 Mar;15(1):46-58. [doi: [10.1016/s0924-9338\(00\)00211-x](#)] [Medline: [10713802](#)]
6. Jazaieri H, Goldin PR, Werner K, Ziv M, Gross JJ. A randomized trial of MBSR versus aerobic exercise for social anxiety disorder. *J Clin Psychol* 2012 Jul 23;68(7):715-731 [FREE Full text] [doi: [10.1002/jclp.21863](#)] [Medline: [22623316](#)]
7. Clark D, Wells A. A cognitive model of social phobia. In: *Social phobia: Diagnosis, assessment, and treatment*. New York: Guilford Press; 1995.
8. Hofmann SG. Cognitive factors that maintain social anxiety disorder: a comprehensive model and its treatment implications. *Cogn Behav Ther* 2007 Dec;36(4):193-209 [FREE Full text] [doi: [10.1080/16506070701421313](#)] [Medline: [18049945](#)]
9. Căndea DM, Szentágotai-Tătar A. Shame as a predictor of post-event rumination in social anxiety. *Cogn Emot* 2017 Dec 15;31(8):1684-1691. [doi: [10.1080/02699931.2016.1243518](#)] [Medline: [27744788](#)]
10. Beck A. *Cognitive therapy of depression*. New York: Guilford Press; 1979.
11. Beck J, Beck A. *Cognitive behavior therapy: Basics and beyond*. New York: Guilford Press; 2011.
12. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychol Med* 2015 Jan;45(1):11-27. [doi: [10.1017/S0033291714000129](#)] [Medline: [24569086](#)]
13. Chartier-Otis M, Perreault M, Bélanger C. Determinants of barriers to treatment for anxiety disorders. *Psychiatr Q* 2010 Jun 4;81(2):127-138. [doi: [10.1007/s11126-010-9123-5](#)] [Medline: [20131095](#)]
14. Stolz T, Schulz A, Krieger T, Vincent A, Urech A, Moser C, et al. A mobile app for social anxiety disorder: A three-arm randomized controlled trial comparing mobile and PC-based guided self-help interventions. *J Consult Clin Psychol* 2018 Jun;86(6):493-504. [doi: [10.1037/ccp0000301](#)] [Medline: [29781648](#)]
15. Maples-Keller JL, Bunnell BE, Kim S, Rothbaum BO. The Use of Virtual Reality Technology in the Treatment of Anxiety and Other Psychiatric Disorders. *Harv Rev Psychiatry* 2017;25(3):103-113 [FREE Full text] [doi: [10.1097/HRP.0000000000000138](#)] [Medline: [28475502](#)]
16. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence. *Psychiatry Res* 2016 Mar 28;236:189-195. [doi: [10.1016/j.psychres.2016.01.015](#)] [Medline: [26795129](#)]

17. Powers MB, Emmelkamp PM. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *J Anxiety Disord* 2008 Apr;22(3):561-569. [doi: [10.1016/j.janxdis.2007.04.006](https://doi.org/10.1016/j.janxdis.2007.04.006)] [Medline: [17544252](https://pubmed.ncbi.nlm.nih.gov/17544252/)]
18. Geraets CN, Veling W, Witlox M, Staring AB, Matthijssen SJ, Cath D. Virtual reality-based cognitive behavioural therapy for patients with generalized social anxiety disorder: a pilot study. *Behav Cogn Psychother* 2019 Nov;47(6):745-750. [doi: [10.1017/S1352465819000225](https://doi.org/10.1017/S1352465819000225)] [Medline: [30915939](https://pubmed.ncbi.nlm.nih.gov/30915939/)]
19. Bouchard S, Dumoulin S, Robillard G, Guitard T, Klinger E, Forget H, et al. Virtual reality compared with exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. *Br J Psychiatry* 2017 Apr 02;210(4):276-283. [doi: [10.1192/bjp.bp.116.184234](https://doi.org/10.1192/bjp.bp.116.184234)] [Medline: [27979818](https://pubmed.ncbi.nlm.nih.gov/27979818/)]
20. Yoo S, Kim Y, Noh J, Oh K, Kim C, NamKoong K. Validity of Korean version of the mini-international neuropsychiatric interview. *Anxiety Mood* 2006;2(1):50-55.
21. Lee J, Choi CH. A study of the reliability and the validity of the Korean versions of Social Phobia Scales (K-SAD, K-FNE). *Korean J Clin Psychol* 1997;16(2):251-264.
22. Cho C. Social Anxiety Disorder VR treatment. 2019. URL: [https://www.youtube.com/channel/UCCpzKK\\_qEqz8LRfxl\\_TfGXw?view\\_as=subscriber](https://www.youtube.com/channel/UCCpzKK_qEqz8LRfxl_TfGXw?view_as=subscriber) [accessed 2019-07-05]
23. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol* 1988 Dec;56(6):893-897. [doi: [10.1037//0022-006x.56.6.893](https://doi.org/10.1037//0022-006x.56.6.893)] [Medline: [3204199](https://pubmed.ncbi.nlm.nih.gov/3204199/)]
24. Yook S, Kim Z. A clinical study on the Korean version of Beck Anxiety Inventory: comparative study of patient and non-patient. *Korean J Clin Psychol* 1997;16(1):185-197.
25. Spielberger C. State-trait anxiety inventory for adults: Manual and Sample: Manual, Instrument and Scoring Guide. Palo Alto: Consulting Psychologists Press; 1983.
26. Kim J, Shin D. A study on the standardization of STAI for Korea. *New Med J* 1978;21(11):69-75.
27. Mattick RP, Clarke J. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav Res Ther* 1998 Apr;36(4):455-470. [doi: [10.1016/s0005-7967\(97\)10031-6](https://doi.org/10.1016/s0005-7967(97)10031-6)]
28. Kim H. Memory bias in subtypes of social phobia, Masters thesis. Seoul, Korea: Seoul National University; 2001.
29. Heimberg RG, Becker RE. Cognitive-Behavioral Group Therapy for Social Phobia: Basic Mechanisms and Clinical Strategies. New York: Guilford Press; Jul 03, 2002.
30. Watson D, Friend R. Measurement of social-evaluative anxiety. *J Consult Clin Psychol* 1969 Aug;33(4):448-457. [doi: [10.1037/h0027806](https://doi.org/10.1037/h0027806)] [Medline: [5810590](https://pubmed.ncbi.nlm.nih.gov/5810590/)]
31. Cook DR. Measuring shame: The internalized shame scale. *Alcohol Treat Quart* 1988 Mar 07;4(2):197-215. [doi: [10.1300/j020v04n02\\_12](https://doi.org/10.1300/j020v04n02_12)]
32. Lee I, Choi H. Assessment of shame and its relationship with maternal attachment, hypersensitive narcissism and loneliness. *Korean J Couns Psychother* 2005;17(3):651-670.
33. Abbott MJ, Rapee RM. Post-event rumination and negative self-appraisal in social phobia before and after treatment. *J Abnorm Psychol* 2004 Mar;113(1):136-144. [doi: [10.1037/0021-843X.113.1.136](https://doi.org/10.1037/0021-843X.113.1.136)] [Medline: [14992666](https://pubmed.ncbi.nlm.nih.gov/14992666/)]
34. Edwards S, Rapee R, Franklin J. Postevent rumination and recall bias for a social performance event in high and low socially anxious individuals. *Cogn Ther Res* 2003;27(6):603-617.
35. Im S, Kwon S, Choi H. *Korean J Clin Psychol* 2007 Feb;26(1):39-56.
36. Liebowitz M. Social phobia. *Mod Probl Pharmacopsychiatry* 1987;22:141-173. [doi: [10.1159/000414022](https://doi.org/10.1159/000414022)] [Medline: [2885745](https://pubmed.ncbi.nlm.nih.gov/2885745/)]
37. Yu E, Ahn CY, Park KH. Factor Structure and Diagnostic Efficiency of A Korean Version of the Liebowitz Social Anxiety Scale. *K J Clin Psychol* 2007 Feb;26(1):251-270. [doi: [10.15842/kjcp.2007.26.1.015](https://doi.org/10.15842/kjcp.2007.26.1.015)]
38. Williams J, Watts F, MacLeod C, Mathews A. Cognitive psychology and emotional disorders. New Jersey: John Wiley & Sons; 1988.
39. Blankstein KR. Relationships between Spielberger trait anxiety and Lykken social and physical trait anxiety. *J Clin Psychol* 1976 Oct;32(4):781-782. [Medline: [977740](https://pubmed.ncbi.nlm.nih.gov/977740/)]
40. Wheaton MG, Galfalvy H, Steinman SA, Wall MM, Foa EB, Simpson HB. Patient adherence and treatment outcome with exposure and response prevention for OCD: Which components of adherence matter and who becomes well? *Behav Res Ther* 2016 Oct;85:6-12 [FREE Full text] [doi: [10.1016/j.brat.2016.07.010](https://doi.org/10.1016/j.brat.2016.07.010)] [Medline: [27497840](https://pubmed.ncbi.nlm.nih.gov/27497840/)]
41. Välimäki M, Hätönen HM, Lahti ME, Kurki M, Hottinen A, Metsäranta K, et al. Virtual reality for treatment compliance for people with serious mental illness. *Cochrane Database Syst Rev* 2014 Oct 08(10):CD009928. [doi: [10.1002/14651858.CD009928.pub2](https://doi.org/10.1002/14651858.CD009928.pub2)] [Medline: [25300174](https://pubmed.ncbi.nlm.nih.gov/25300174/)]
42. Morrison AS, Heimberg RG. Social anxiety and social anxiety disorder. *Annu Rev Clin Psychol* 2013 Mar 28;9(1):249-274. [doi: [10.1146/annurev-clinpsy-050212-185631](https://doi.org/10.1146/annurev-clinpsy-050212-185631)] [Medline: [23537485](https://pubmed.ncbi.nlm.nih.gov/23537485/)]
43. Gilbert P. The relationship of shame, social anxiety and depression: the role of the evaluation of social rank. *Clin Psychol Psychother* 2000 Jul;7(3):174-189. [doi: [10.1002/1099-0879\(200007\)7:3<174::aid-cpp236>3.0.co;2-u](https://doi.org/10.1002/1099-0879(200007)7:3<174::aid-cpp236>3.0.co;2-u)]
44. Zimmerman J, Morrison AS, Heimberg RG. Social anxiety, submissiveness, and shame in men and women: a moderated mediation analysis. *Br J Clin Psychol* 2015 Mar 28;54(1):1-15. [doi: [10.1111/bjc.12057](https://doi.org/10.1111/bjc.12057)] [Medline: [24866818](https://pubmed.ncbi.nlm.nih.gov/24866818/)]
45. Weeks JW, Heimberg RG, Heuer R. Exploring the Role of Behavioral Submissiveness in Social Anxiety. *J Soc Clin Psychol* 2011 Mar;30(3):217-249. [doi: [10.1521/jscp.2011.30.3.217](https://doi.org/10.1521/jscp.2011.30.3.217)]



46. Zou JB, Abbott MJ. Self-perception and rumination in social anxiety. *Behav Res Ther* 2012 Apr;50(4):250-257. [doi: [10.1016/j.brat.2012.01.007](https://doi.org/10.1016/j.brat.2012.01.007)] [Medline: [22398151](https://pubmed.ncbi.nlm.nih.gov/22398151/)]
47. Kampmann IL, Emmelkamp PM, Morina N. Meta-analysis of technology-assisted interventions for social anxiety disorder. *J Anxiety Disord* 2016 Aug;42:71-84. [doi: [10.1016/j.janxdis.2016.06.007](https://doi.org/10.1016/j.janxdis.2016.06.007)] [Medline: [27376634](https://pubmed.ncbi.nlm.nih.gov/27376634/)]
48. Anderson PL, Price M, Edwards SM, Obasaju MA, Schmertz SK, Zimand E, et al. Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial. *J Consult Clin Psychol* 2013 Oct;81(5):751-760. [doi: [10.1037/a0033559](https://doi.org/10.1037/a0033559)] [Medline: [23796315](https://pubmed.ncbi.nlm.nih.gov/23796315/)]
49. Bouchard S, Dumoulin S, Robillard G, Guitard T, Klinger E, Forget H. A randomized control trial for the use of in virtual reality exposure in the treatment of social phobia: final results. *J Cyberther Rehab* 2011;4(2):197-200.
50. Wiederhold B, Wiederhold M. Virtual reality therapy for anxiety disorders: Advances in evaluation and treatment. Washington, DC: American Psychological Association; 2005.
51. Shibata Y, Brütting J, Pauli P, Mühlberger A. Fear reactivation prior to exposure therapy: does it facilitate the effects of VR exposure in a randomized clinical sample? *J Behav Ther Exp Psychiatry* 2015 Mar;46:133-140. [doi: [10.1016/j.jbtep.2014.09.009](https://doi.org/10.1016/j.jbtep.2014.09.009)] [Medline: [25460259](https://pubmed.ncbi.nlm.nih.gov/25460259/)]

## Abbreviations

**ANOVA:** analysis of variance  
**BAI:** Beck Anxiety Inventory  
**BFNE:** Brief-Fear of Negative Evaluation Scale  
**CBT:** cognitive behavioral therapy  
**ISS:** Internalized Shame Scale  
**KSAD:** Korean Social Avoidance and Distress Scale  
**LSAS:** Liebowitz Social Anxiety Scale  
**PERS:** Post-Event Rumination Scale  
**SAD:** social anxiety disorder  
**SIAS:** Social Interaction Anxiety Scale  
**SPS:** Social Phobia Scale  
**STAI:** State-Trait Anxiety Inventory  
**VR:** virtual reality

*Edited by G Eysenbach; submitted 31.07.20; peer-reviewed by HJ Jeon; comments to author 24.08.20; revised version received 28.08.20; accepted 15.09.20; published 06.10.20.*

### *Please cite as:*

Kim HJ, Lee S, Jung D, Hur JW, Lee HJ, Lee S, Kim GJ, Cho CY, Choi S, Lee SM, Cho CH

Effectiveness of a Participatory and Interactive Virtual Reality Intervention in Patients With Social Anxiety Disorder: Longitudinal Questionnaire Study

*J Med Internet Res* 2020;22(10):e23024

URL: <http://www.jmir.org/2020/10/e23024/>

doi: [10.2196/23024](https://doi.org/10.2196/23024)

PMID: [33021481](https://pubmed.ncbi.nlm.nih.gov/33021481/)

©Hyun-Jin Kim, Seulki Lee, Dooyoung Jung, Ji-Won Hur, Heon-Jeong Lee, Sungkil Lee, Gerard J Kim, Chung-Yean Cho, Seungmoon Choi, Seung-Moo Lee, Chul-Hyun Cho. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Original Paper

# Effects of Interactivity on Recall of Health Information: Experimental Study

Emília Margit Pajor<sup>1</sup>, MSc, PhD; Sander Matthijs Eggers<sup>1,2</sup>, MSc, PhD; Hein de Vries<sup>1</sup>, PhD; Anke Oenema<sup>1</sup>, PhD

<sup>1</sup>Department of Health Promotion, Care and Public Health Research Institute, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

<sup>2</sup>International Trade, Statistics Netherlands, Heerlen, Netherlands

**Corresponding Author:**

Anke Oenema, PhD

Department of Health Promotion, Care and Public Health Research Institute

Faculty of Health Medicine and Life Sciences

Maastricht University

Debijeplein 1

Maastricht

Netherlands

Phone: 31 43 3882131

Email: [a.oenema@maastrichtuniversity.nl](mailto:a.oenema@maastrichtuniversity.nl)

## Abstract

**Background:** Information provided in an interactive way is believed to be engaging because users can actively explore the information. Yet empirical findings often contradict this assumption. Consequently, there is still little known about whether and how interactivity affects communication outcomes such as recall.

**Objective:** The aim of this study was to investigate mechanisms through which interactivity affects recall of online health information. We tested whether and how cognitive involvement, perceived active control, and cognitive load mediate the effects of interactivity on recall. In addition, we examined need for cognition and health literacy as potential moderators of the mediation effects. Given the increasing popularity of dietary supplement use, our health website focused on this topic.

**Methods:** In an online between-subjects experiment (n=983), participants were randomly assigned to control condition (no interactive features), moderate interactivity (dropdown menus), and high interactivity (dropdown menus and responsive infographics). Two weeks before the experiment, background characteristics and moderating variables were measured. During website visit, data on users' online behavior were collected. Recall was measured postexposure.

**Results:** Participants recalled significantly less information in the moderate (mean 3.48 [SD 2.71]) and high (mean 3.52 [SD 2.64]) interactivity conditions compared with the control condition (mean 5.63 [SD 2.18]). In the mediation analysis, we found direct, negative effects of moderate ( $b=-2.25$ , 95% CI  $-2.59$  to  $-1.90$ ) and high ( $b=-2.16$ , 95% CI  $-2.51$  to  $-1.81$ ) levels of interactivity on recall as well. In the relationship between interactivity and recall, cognitive involvement had a partial negative mediation effect (moderate interactivity:  $b=-.20$ , 95% CI  $-0.31$  to  $-0.10$ ; high interactivity:  $b=-.21$ , 95% CI  $-0.33$  to  $-0.10$ ) and perceived active control had a partial positive mediation effect (moderate interactivity:  $b=.28$ , 95% CI  $0.18$  to  $0.40$ ; high interactivity:  $b=.27$ , 95% CI  $0.16$  to  $0.40$ ).

**Conclusions:** Interactivity decreased recall. In addition, through interactivity participants were less involved with the content of the information, yet they felt they had more control over the information. These effects were stronger in the high need for cognition and high health literate groups compared with their counterparts.

(*J Med Internet Res* 2020;22(10):e14783) doi:[10.2196/14783](https://doi.org/10.2196/14783)

**KEYWORDS**

Interactivity; cognitive involvement; active control; cognitive load; recall; need for cognition; health literacy; online health information; information processing; dietary supplements

## Introduction

### Relevance and Aim

One of the unique features of internet-delivered health information is that it can be provided in an interactive way. Interactive features are thought to create more engagement and involvement with the information, as visitors can actively interact with the information [1]. However, clear guidelines are lacking on how interactivity could be used in health communication [2]. Moreover, previous research has resulted in inconclusive findings regarding how interactive features affect communication outcomes. Interactive features may enhance user enjoyment, positive attitudes, and desirable behavioral intentions but they do not necessarily improve cognitive elaboration or information recall [3]. The aim of this study is to investigate whether and how interactivity can be used for improving recall of online health information. Given the increasing popularity of dietary supplement use and the complexity of the behavior [4,5], our study focuses on this topic.

### Theoretical Background

#### *Conceptualization of Interactivity*

In interactivity research, three approaches are distinguished: structural, experiential, and message exchange. In the structural approach, interactivity is conceptualized in terms of the technical attributes of the medium [6]. Such technical attributes include on-screen interactive features such as menus that allow user-to-system or user-to-user interactions [7]. According to the experiential approach, interactivity is the user's subjective perception of the level of the medium's interactivity [8]. The message exchange approach regards interactivity as an ongoing communication process in which (semantic) meanings between two or more communicators are exchanged [9]. We conceptualize interactivity according to the structural approach, since in terms of causality a media stimulus that is manipulated precedes users' responses to that stimulus [6]. Yet the structural approach often has been criticized for focusing only on direct relations between a media stimulus and the dependent measures omitting the possibility of third variable effects [6]. Therefore, in the conceptual model of this study, four types of variables were included: interactivity (manipulated independent variable), possible mediators, possible moderators, and a dependent variable.

#### *Moderated Mediation Model of Interactivity Effects*

In our conceptual model, we will test whether cognitive involvement, perceived active control, and cognitive load mediate the relationship between interactivity and recall. In addition, we will look at whether the need for cognition and health literacy moderate the proposed mediation effects.

Dual process models, such as the elaboration likelihood model (ELM), propose that information elaborated via the central route is likely to produce greater and more permanent changes in communication outcomes [10]. Individuals tend to get (more) motivated to process the information systematically, which involves effortful thinking, if the information is perceived as personally relevant to them [10]. It is proposed that individuals generate (more) cognitive responses (ie, the number of

content-related thoughts generated during exposure to the stimuli) to the information content if they process the information systematically [11]. Interactive environments may stimulate cognitive responses since they enable nonlinear, cognitively flexible information use (cognitive flexibility theory) [12]. For instance, navigation through hypertext might be beneficial for knowledge since hypertexts mimic the associative network of human memory [13]. If individuals are afforded the possibility to adapt their information use to their own preferences and cognitive needs, they may get more involved and more motivated to process the information more deeply and elaborate better on the content [14,15]. Therefore, we propose that interactivity improves recall through higher levels of cognitive involvement.

Research has shown that individuals' perceptions about interactivity are closely related to communicational outcomes such as attitudes [16,17]. From the different dimensions of perceived interactivity distinguished by Liu and Shrum [18], the control dimension, which refers to the autonomy individuals have in controlling the information flow, has been most often associated with cognitive elements of information processing [19]. Active control is characterized by voluntary and instrumental actions through which users are able to customize the information flow [18]. This entails navigational choices based on the user's own goals and wills [18]. We assume that if users are afforded the possibility to make conscious choices about the information flow based on their needs, they might be more motivated to process the information. Therefore, we propose that interactivity may improve recall through higher levels of perceived active control.

Interactivity may challenge individuals' information processing capacities by putting an extra burden on users. Tremayne and Dunwoody [20] found evidence that when users visited a more (vs less) interactive website, much cognitive effort was spent in navigation and orientation which had a detrimental effect on recall. In interactive environments, individuals must complete different tasks performed concurrently (eg, reading, navigating). Every task generates a cognitive cost on the working memory in terms of cognitive load. Moreover, tasks may interfere with each other since they compete for the same limited cognitive resources [21,22]. Consequently, there are fewer capacities left for information processing (ie, encoding, storage, retrieval). Indeed, research has shown that individuals retain less information when performing more than one task at the same time because multitasking inhibits the transfer of information into the short- and long-term memory [23,24]. Therefore, we assume that cognitive load increases with higher levels of interactivity, which may lead to decreased levels of recall.

#### *Moderators: Need for Cognition and Health Literacy*

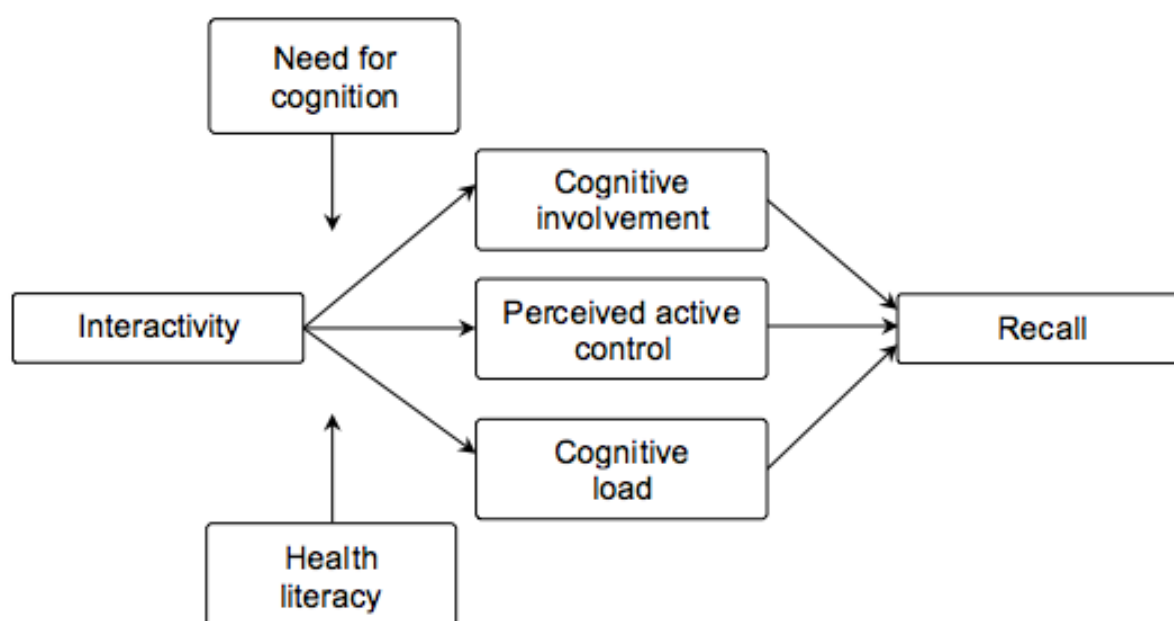
In addition to the mediation effects described above, we aim to explore whether individual difference variables moderate the proposed mechanisms. Information processing is influenced by individuals' ability and/or motivation to process information [10,22]. Need for cognition reflects the tendency to engage in and enjoy effortful cognitive endeavors [25,26]. It is considered a stable trait that may be influenced by situational factors such as interactivity [27]. Evidence suggests that interactivity

improves information processing, especially among low need for cognition individuals [28]. This may be related to the fact that individuals with low need for cognition prefer interactive websites more than their high need for cognition counterparts [29]. In situations with low personal relevance, individuals with low need for cognition are more attracted by peripheral cues which are attributes that are not inherent to the strength of arguments in the message [30]. In contrast, individuals with high need for cognition concentrate more on the real attributes of the information, such as the strength of arguments. Therefore, they rely less on the way information is presented [28]. However, high need for cognition individuals are more cognitively immersed when engaging in interactive websites than low need for cognition individuals [28,31]. According to the findings of Sicilia et al [28], in both low and high need for cognition individuals, the flow experience increases when they visit an interactive (vs noninteractive) website, but this increase is higher among individuals with high need for cognition.

Individuals may also differ in their ability and skills to understand and use health information. Health literacy entails

“the motivation, knowledge, and competencies to access, understand, appraise and apply health information in order to make judgments and take decisions...concerning health care, disease prevention, etc” [32]. In general, low health literate individuals engage less in health information seeking and have greater difficulties with reading and searching for health information on the internet [33,34]. Moreover, interactivity may challenge users with limited literacy skills since they have difficulties with recognizing graphic links (ie, pictures that function as hyperlinks), using navigational tools, and understanding graphics that respond to mouse movements [35]. In addition, sufficient levels of metacognitive skills are needed to make mindful navigational selections and build meaningful sequences of information, for instance, in a hypermedia environment [36]. However, it should be noted that research has shown that interactive features designed specifically for individuals with low health literacy (eg, low text difficulty) are beneficial for online health information processing [37]. We summarized our conceptual model in Figure 1.

**Figure 1.** Conceptual model of moderated mediation effects of interactivity.



In sum, we aim to answer the following research questions: through which mechanisms does interactivity affect recall of health information? Do these mechanisms differ according to individuals' level of need for cognition and health literacy?

## Methods

### Design

A between-subjects experiment with three levels of interactivity (no interactivity, moderate interactivity, high interactivity) was conducted to investigate the effects of interactivity on cognitive involvement, perceived active control, cognitive load, and recall. Two weeks prior to the experiment, background characteristics

and moderator variables were measured. During the experiment, pre- and postexposure measures were performed.

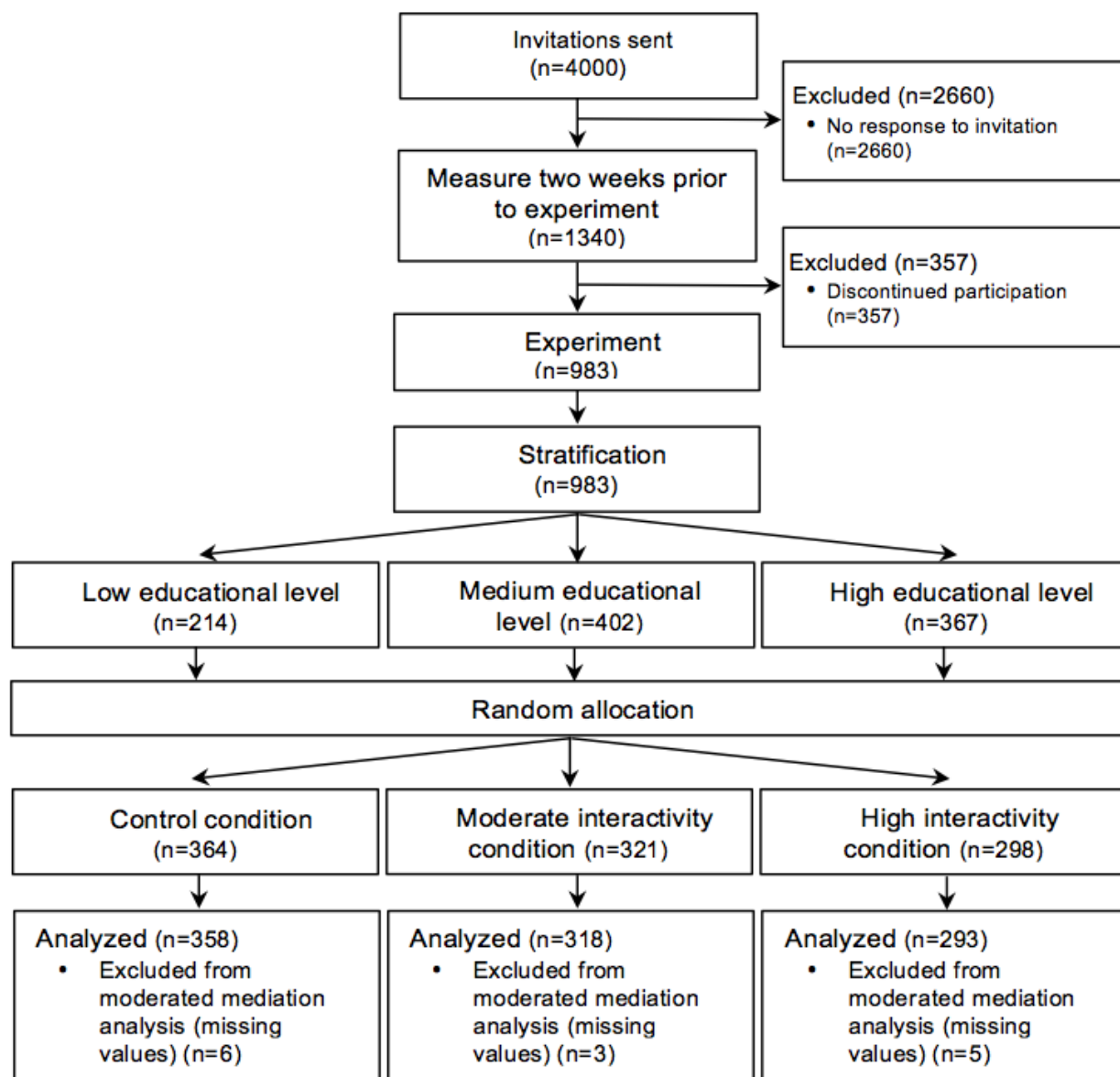
### Participants

A priori power analysis with G\*Power indicated that at least 776 participants were needed to detect small effects (ie,  $f^2=.02$ ) with an alpha level of .05, .90 statistical power, and 4 predictors. Participants were recruited from 26,000 active panel members of I&O Research, an ISO 26362–certified research bureau for access panels in market, opinion, and social research [38]. I&O Research recruits its panel members offline (eg, from municipal registers) and self-registration is not allowed in order to prevent selection biases such as the overrepresentation of frequent internet users [38]. A random sample consisting of 4000

individuals was drawn from the panel, of which 33.50% (1340/4000) responded to the invitation to participate in the first part of the study (see Figure 2). This response rate was comparable to the average response rate of this panel (ie, 35%) [38]. The final sample consisted of 983 (73.4%) individuals who participated in both parts of the study and completed the

pre- and postexposure questionnaires. Among these participants, 50 €10 gift cards for an internet warehouse were raffled. Due to technical issues, data of 15 respondents were lost in the preexposure measurement. Therefore, these participants were excluded from the moderated mediation analyses.

**Figure 2.** Flowchart of the stratification and random allocation procedure.



## Procedure

An invitation with the subject line “Consumer information on food and dietary supplements” was sent by email to participants on March 21, 2017. The email contained a short description of the study, an explanation of the procedure, and contact details of the researcher. Upon agreeing to participate in the study and giving informed consent, individuals were asked to complete a short questionnaire about health literacy, need for cognition, and educational level. Based on individuals’ educational level, the sample was divided into three strata: low, medium, and high educational level. On April 25, 2017, participants were invited to take part in the actual experiment. Prior to visiting the website, individuals completed a preexposure measurement

about their knowledge of the research topic and dietary supplement use. Then, within each stratum, participants were randomly assigned to one of the three versions of a website about vitamin B6 and dietary supplements. The random allocation was programmed by the website developer, Done Digital Kft. In order to make the browsing task similar to a real-life online health information search, participants had the freedom to decide what information and in which order they wanted to explore and no specific instructions or time limits were given. Participants were allowed to view the website only once. In order to prevent any preexposure to the stimulus material, the website was not publicly accessible. Once participants were finished browsing on the website, they were directed to the postmeasurement questionnaire.

## Ethical Approval

According to the decision of the Research Ethics Commission of Maastricht University Medical Centre and Maastricht University (decision number: METC 16-4-268), the Medical Research Involving Human Subjects Act does not apply to this study. At the time of the study, further ethical clearance was not required.

## Stimulus Material

In all three conditions, the information presented was identical and aimed to provide complete information about vitamin B6, its physiological effects, how it relates to food, and the risks and benefits of supplementation with vitamin B6. The goal was to offer information that is well balanced in terms of describing advantages and disadvantages of vitamin B6 and improves individuals' understanding of whether and to whom supplementation with vitamin B6 may be reasonable. The recommendations of the Ottawa Decision Support Tutorial were used [39].

In addition to the textual information, the website included four educative infographics that contained information about the Recommended Dietary Allowance of vitamin B6 by age category and gender, the vitamin B6 content of different food products, scientific evidence of possible health effects of vitamin B6, and safe and unsafe doses of vitamin B6 dietary

supplements. Further details about the content and structure of the website are presented in [Multimedia Appendix 1](#).

## Experimental Stimuli

Three versions of a website were developed sharing the same content, layout, and pictures but differing in terms of levels of interactivity. In our study, interactivity refers to the technical attributes of the medium [6]. In line with previous research that falls into the structural approach of interactivity [1,40], we manipulated levels of interactivity in terms of the amount of interactive tools available on the website interface (eg, menus) for accessing and interacting with the content of the website. Accordingly, in the control condition no interactive features were presented. Participants navigated by scrolling up or down on the webpage and the infographics did not respond to users' mouse movements. In the moderate interactivity condition, a dropdown menu consisting of nine submenus was the only interactive tool presented on the website. The infographics shown in the moderate interactivity condition were static. In the high interactivity condition, two types of interactive features were presented: a dropdown menu and responsive infographics. Three rounds of pilot tests were conducted before the websites were finalized. In [Figure 3](#), the differences in the navigation (scrolling vs dropdown menu) between the conditions are presented. In [Figure 4](#), the differences in the infographics (static vs responsive) between the conditions are presented.

**Figure 3.** Navigation in the control condition (left) and in the moderate and high interactivity condition (right).





**Figure 4.** Example of an infographic in the control and moderate interactivity condition (left) and the high interactivity condition (right).

## Measures

### Dependent Variable: Recall

The measure of recall was based on the construction integration model [41] that distinguishes between two levels of comprehension: the text base level (ie, literal recall of information) and the situation model level (ie, to make inferences to situations based on the information) comprehension. During two cycles of pretesting, 4 questions were developed to measure text base level comprehension and 2 questions to measure situation level comprehension; 3 of the 6 questions were in multiple-choice format in which only one answer option was correct. Three other questions were in multiple response format in which multiple answer options were correct. In both types of response options, the “I don’t know” option was also presented. Participants received 1 point for each correct answer, up to 9 points in total.

According to the principal component analysis, the 6 items loaded on one single underlying construct (eigen value: 2.866, 47.765% of variance explained). In addition, factor loadings were .582 or higher. All 6 items correlated significantly with each other, ranging from  $P=.25$  to  $P=.51$ .

### Mediating Variables

Participants’ cognitive involvement with the website content was measured with a thought listing task [11]. Individuals were asked to list thoughts about vitamin B6 that came to their mind during their website visit. Participants could list up to 12 thoughts in empty text boxes, each beginning with the statement: “Vitamin B6 is:...” Each field of text that was completed counted as a thought. The cognitive involvement measure was placed second in the postexposure questionnaire, after the 1-item manipulation check of interactivity. Compared with thought listing measures during exposure, obtaining responses after the stimulus can be accomplished without the interruption or

distraction from the stimulus [11]. In addition, loss or retention in poststimulus measures of cognitive involvement is negligible [11].

Perceived active control refers to users’ voluntary and instrumental actions that directly influence their website experience [8,18]. Liu’s [8] 4-item measurement of perceived active control was used.

Cognitive load was conceptualized as the perception of “the cognitive capacity that is actually allocated to accommodate the demands imposed by the task” [42]. It was measured with 4 items derived from the Subjective Workload Assessment Technique (mental effort subdimension) [43,44] and from the NASA Task Load Index (effort subdimension) [43,45].

### Moderating Variables

Need for cognition was measured with 7 items derived from the Dutch version of the 18-item Need for Cognition Scale (NCS) [46]. A limited number of items may be sufficient to measure need for cognition since evidence suggests that the full NCS measures a single underlying construct [26,47,48]. Therefore, the items with the highest factor loadings were chosen from the study of Hevey and colleagues [49].

Health literacy was measured using the Newest Vital Sign in Dutch (NVS-D), which was developed to measure an individual’s capacity to assess, understand, and use textual and numerical health information [50]. Consequently, this 6-item tool tests three types of skills that are important in finding and interpreting health information: math, locate-the-information (by reading and comprehending), and abstract reasoning (ie, making inferences from the information to specific situations) [51]. For each correct answer, respondents received 1 point, up to 6 points in total. Details of the measurement scales (number of items, example items, answer categories, mean scores, and Cronbach  $\alpha$ ) are presented in Table 1.

**Table 1.** Overview of number and examples of questions, answering categories, mean scores, and Cronbach

Variable	Number of questions	Example of questions/items, answer options	Mean (SD)	Median <sup>a</sup> (IQR)	Cronbach alpha
Recall <sup>b</sup>	6	“Some groups of individuals are at risk of developing vitamin B6 deficiency. Which groups are these?” “Someone is using high-dose vitamin B6 dietary supplements (100 milligram) for a long period of time. What kind of influence can this have on the health condition of this person?” Multiple choice or multiple response answer options	4.30 (2.71)	— <sup>c</sup>	.77
Cognitive involvement <sup>d</sup>	12	“During your website visit, certain thoughts about vitamin B6 may have come to your mind. Please write down your thoughts about vitamin B6 in the text boxes below.” Open-ended questions	—	3 (3)	—
Perceived active control	3	“While I was on the website, I could choose freely what I wanted to see.” Totally disagree (1) to totally agree (7)	3.98 (1.29)	—	.80
Cognitive load	4	“I had to think hard in order to understand the information on the website” Totally disagree (1) to totally agree (7)	4.05 (1.42)	—	.92
Health literacy <sup>e</sup>	6	“If you are allowed 60 grams of carbohydrates as a snack, how much ice cream could you have?” Open-ended questions	—	6 (1)	—
Need for cognition	7	“I really enjoy a task that involves coming up with new solutions to problems” Totally disagree (1) to totally agree (7)	4.94 (1.02)	—	.83

<sup>a</sup>For variables with non-normal distribution, the median and IQR are presented.

<sup>b</sup>Recall scores could range from 0 (no correct answers) to 9 (all answers are correct).

<sup>c</sup>Not applicable.

<sup>d</sup>Cognitive involvement scores could range from 0 (no thoughts) to 12 (12 thoughts).

<sup>e</sup>Health literacy scores could range from 0 (no correct answers) to 6 (all answers are correct).

## Manipulation Check Measure

The effectiveness of the manipulation was measured by asking respondents: “To what extent do you agree with the following statement: This website is interactive” with answer options ranging from 1 (totally disagree) to 7 (totally agree) [52].

## User Activity

During participants’ website visit, the following user activity indicators were measured: duration of website visit in seconds (mean 11,765.69 [SD 92,169.31], range 5-1,302,883), the extent of scrolling down on the website (control condition: mean 0.98 [SD 0.32]), total amount of clicks on the nine menus (moderate interactivity condition: mean 6.11 [SD 3.45]; high interactivity condition: mean 6.13 [SD 2.64]), and total amount of clicks on the four infographics (high interactivity condition: mean 11.05 [SD 9.09]). Due to personal browser settings such as a disabled JavaScript [53], user activity data of 524 participants out of the 983 participants were collected.

## Control Variables

Several variables were measured to control for their potential influence in the statistical analyses. The following demographic background characteristics were measured: gender, age, and highest educational level. The latter was measured with 7 responses: 1=primary education or less, 2=preparatory secondary vocational education (level 1) or equivalent, 3=secondary vocational education, 4=senior secondary vocational education

(level 2-4) or equivalent, 5=senior general secondary education, university preparation education, 6=bachelor’s level or equivalent, and 7=master’s level or above. The strata of low (1-3), middle (4 and 5), and high educational level (6 and 7) were based on these responses. In addition, meat consumption, diet, mode of life (eg, anthroposophic nutrition), dietary supplement use, and involvement with the topic vitamin B6 were measured. Involvement was measured with 3 items from Zaichkowsky’s [54] personal involvement inventory scale (eg, “The topic vitamin B6 is important to me,” 1=totally disagree to 7=totally agree).

## Statistical Analyses

Descriptive statistics were run to investigate sample characteristics. In order to check normal distribution of the variables, we looked at skewness and kurtosis and visually inspected the histogram and boxplot of each variable. We detected nonnormal distribution in three variables: health literacy (skewness: -1.67, kurtosis: 2.28), cognitive involvement (skewness: 1.48, kurtosis: 3.68), and time spent on website (skewness: 9.42, kurtosis: 101.41). To investigate the proposed mediations and examine whether differences exist between subgroups, the PROCESS macro for SPSS Statistics version 2.16.3 (IBM Corporation) was used [55]. PROCESS applies bootstrapping to estimate 95% bias-corrected confidence intervals for total and indirect effects. Bootstrap procedures are unaffected by violations of parametric assumptions and have

higher type I error control and power than the normal theory approach [55,56]. In PROCESS, 76 different conceptual diagrams are available. Model number 4 is programmed to test a simple mediation. In order to test mediations, model 4 (10,000 samples) was used with three mediators operating in parallel: cognitive involvement, perceived active control, and cognitive load. In the analyses, the independent variable interactivity was defined as multicategorical; consequently, it was automatically dummy coded by PROCESS (D1: moderate interactivity condition; D2: high interactivity condition; reference category: control condition). The percentage mediated effect was calculated for each significant indirect effect separately by dividing the corresponding unstandardized regression coefficient of the ab path (indirect effect) by the unstandardized regression coefficient of the c path (total effect) and multiplying it by 100. To examine differences between subgroups regarding the mediations (high vs low health literacy, high vs low need for cognition), model 4 (10,000 samples) was run for each group separately. Subgroups were created based on a mean split. This approach of testing moderated mediation effects was chosen for a technical reason. PROCESS v2 had limited features for dealing with multicategorical variables, which meant that it was

not possible to calculate an interaction term of a multicategorical and a numerical variable and put it in the model as a variable [57]. All analyses were conducted with adjustments for gender, age, educational level, dietary supplement use, involvement with the topic vitamin B6, being on a diet, following a certain rule of life (eg, anthroposophy), meat consumption, and duration of website visit. Analyses were conducted with SPSS version 23.

## Results

### Sample Characteristics

Slightly more female (530/983, 53.9%) than male (453/983, 46.1%) individuals participated in the study. Participants were on average aged 53.2 (SD 15.31) years, and most of them held either a medium (402/983, 40.9%) or high educational level (367/983, 37.3%; Table 2). More than half of the sample (524/983, 53.3%) had used dietary supplements in the last 12 months. Participants can be regarded as neutral toward the topic vitamin B6 since they were moderately involved (mean 3.42 [SD 1.54]).

**Table 2.** Sample characteristics (n=983).

Characteristic	Value
Gender (male), n (%)	453 (46.1)
Age in years, mean (SD)	53.20 (15.31)
<b>Educational level, n (%)</b>	
Low	214 (21.8)
Medium	402 (40.9)
High	367 (37.3)
Living according to a specific mode of life (yes), n (%)	92 (9.4)
Meat consumption (yes), n (%)	945 (96.1)
Being on a diet (yes), n (%)	201 (20.4)
Dietary supplement use in the last 12 months (yes), n (%)	524 (53.3)
Involvement with the topic vitamin B6 <sup>a</sup> , mean (SD)	3.42 (1.54)

<sup>a</sup>Involvement was measured on a 7-point Likert scale. The higher the score, the more involved participants were.

### Manipulation Check of Interactivity

The analysis of variance revealed significant differences between the three versions of the website regarding interactivity ( $F_{2,980}=24.99$ ,  $P<.001$ ,  $\eta^2=.05$ ). Post hoc Bonferroni tests indicated that all three versions differed significantly from each other. Participants rated the level of interactivity as low in the control condition (mean 3.31, SE .10, n=364), as moderate in the moderate interactivity condition (mean 3.89, SE .10, n=321), and as high in the high interactivity condition (mean 4.30, SE .10, n=298). Thus, the manipulation was successful.

### Descriptive Statistics of User Actions Within Conditions

Users spent the most time (minutes) browsing the website in the control condition (median 2.62 [IQR 1.54-3.96]), followed

by the moderate (median 0.8 [IQR 0.45-2.84]) and high interactivity condition (median 0.83 [IQR 0.43-3.23]). In the control condition, 78.7% (203/258) of participants scrolled all the way down the website and viewed all website content. In the moderate interactivity condition, 14.8% (21/142) of users clicked on all dropdown menus and viewed the complete content, whereas in the high interactivity condition only 8.9% (11/124) did so. Within the high interactivity condition, the proportion of participants who used all infographics was higher (36/124, 29.0%) than of those who used all dropdown menus (11/124, 8.9%). Descriptive statistics of user actions within conditions are presented in Table 3.

**Table 3.** Indicators of user activity per condition (n=524).

Variable	Control condition (n=258)	Moderate interactivity condition (n=142)	High interactivity condition (n=124)
Duration of website visit in minutes, mean (SD) [range]	366.19 (2146.69) [0.08-21714.72]	112.54 (1103.02) [0.08-13083.52]	78.05 (893.48) [0.10-11599.07]
Scrolled to end of browser window <sup>a</sup> , % (modus) [median]	78.7 (1) [1]	N/A <sup>b</sup>	N/A
Used all dropdown menus <sup>c</sup> , % (modus) [median]	N/A	14.8 (7) [6]	8.9 (7) [6]
Used all infographics <sup>d</sup> , % (modus) [median]	N/A	N/A	29.2 (4) [2]

<sup>a</sup>Scrolled scale ranges from 0 (no scroll down) to 1 (scrolled to end of window at least once).

<sup>b</sup>Not applicable.

<sup>c</sup>Dropdown menu use scale ranges from 0 (no clicks on menus) to 9 (all menus were used at least once).

<sup>d</sup>Infographic use scale ranges from 0 (none of the infographics were used) to 4 (all infographics were used at least once).

### Main Effect of Interactivity on Recall

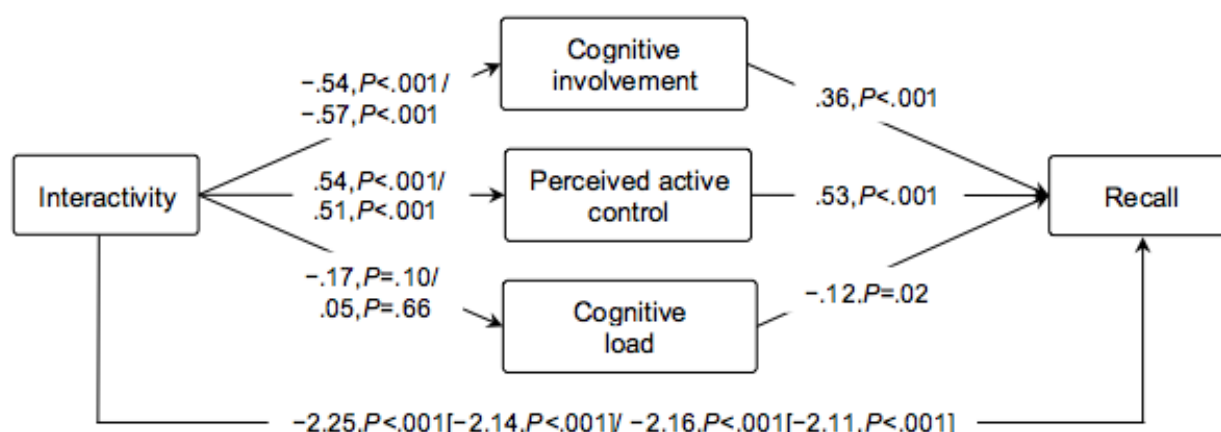
Analysis of variance showed that recall score differed significantly between the conditions ( $F_{2,980}=82.329$ ,  $P<.001$ ,  $\eta^2=.144$ ). According to the Bonferroni post hoc test, participants recalled significantly more information in the control condition (mean 5.63 [SD 2.18]) compared with the moderate (mean 3.48 [SD 2.71]) and high (mean 3.52 [SD 2.64]) interactivity conditions.

### Mediating Effect of Cognitive Involvement, Perceived Active Control, and Cognitive Load

Results showed that the effects of levels of interactivity were mediated by cognitive involvement (moderate interactivity:

$b=-.20$ , 95% CI  $-0.31$  to  $-0.10$ , 9% mediated effect; high interactivity:  $b=-.21$ , 95% CI  $-0.33$  to  $-0.10$ , 10% mediated effect) and perceived active control (moderate interactivity:  $b=.28$ , 95% CI  $0.18$  to  $0.40$ , 13% mediated effect; high interactivity:  $b=.27$ , 95% CI  $0.16$  to  $0.40$ , 13% mediated effect) but not by cognitive load (moderate interactivity:  $b=.02$ , 95% CI  $0$  to  $0.07$ ; high interactivity:  $b=-.01$ , 95% CI  $-0.04$  to  $0.02$ ). The mediations were partial as there was a remaining significant direct effect of interactivity on recall (moderate interactivity:  $b=-2.25$ , 95% CI  $-2.59$  to  $-1.90$ ; high interactivity:  $b=-2.16$ , 95% CI  $-2.51$  to  $-1.81$ ). Levels of interactivity and the three mediators explained 34% of the variance in recall ( $F_{14,953}=35.76$ ,  $P<.001$ ). The unstandardized path coefficients of the direct and total effects are presented in Figure 5.

**Figure 5.** Unstandardized path coefficients of the direct and total effects (in brackets) of moderate and high interactivity on the mediators and recall compared with the control condition (n=968).



### Conditional Indirect Effect of Interactivity on Recall According to Two Levels of Need for Cognition

The study sample was split into high (ie, score  $\geq 4.95$ , 519/968) versus low (ie, score  $\leq 4.94$ , 449/968) need for cognition using a mean split. Results showed that in individuals with low need for cognition exposed to the moderate interactivity condition, the effects of levels of interactivity were mediated by cognitive

involvement ( $b=-.18$ , 95% CI  $-0.36$  to  $-0.05$ , 8% mediated effect) and perceived active control ( $b=.14$ , 95% CI  $0.01$  to  $0.30$ , 6% mediated effect) but not by cognitive load ( $b=.12$ , 95% CI  $-0.03$  to  $0.08$ ). In low need for cognition individuals exposed to the high interactivity condition, none of the proposed mediations were significant.

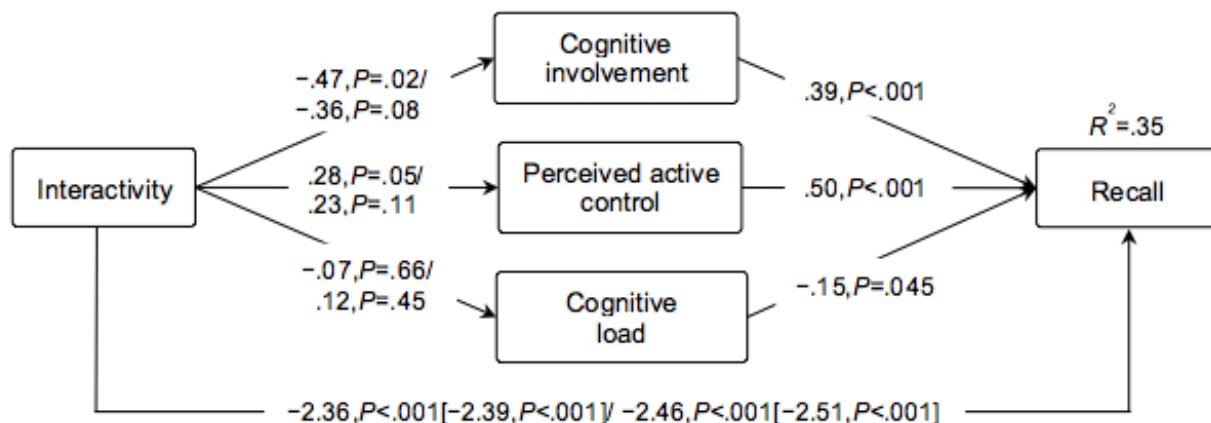
In individuals with high need for cognition, in the moderate interactive condition a significant mediation effect of cognitive

involvement ( $b=-.21$ , 95% CI  $-0.38$  to  $-0.08$ , 11% mediated effect) and active control ( $b=.43$ , 95% CI  $0.27$  to  $0.64$ , 22% mediated effect) was found. The mediation effects of cognitive involvement ( $b=-.28$ , 95% CI  $-0.46$  to  $-0.14$ , 16% mediated effect) and perceived active control ( $b=.43$ , 95% CI  $0.25$  to  $0.65$ , 4% mediated effect) were also significant in the high interactivity condition. Regardless of the condition, no mediation effect of cognitive load was found in high need for cognition individuals. As presented in Figure 6, levels of interactivity had a significant direct effect on recall.

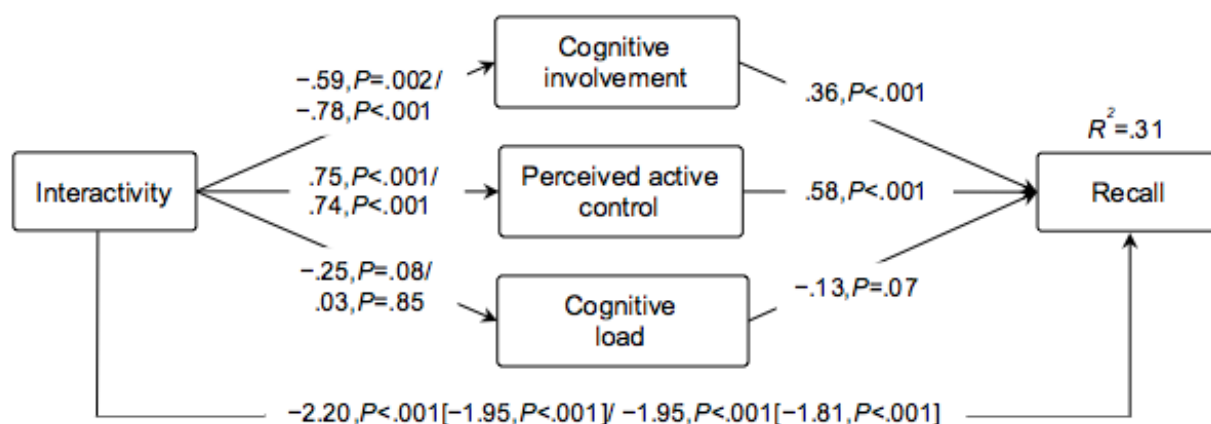
no mediation effects were found in low health literate individuals.

**Figure 6.** Unstandardized path coefficients of the direct and total effects (in brackets) of moderate and high interactivity on the mediators and recall compared with the control condition in the low ( $n=449$ ) and high ( $n=519$ ) need for cognition group.

### Low need for cognition group



### High need for cognition group



### Conditional Indirect Effect of Functional Interactivity on Recall According to Two Levels of Health Literacy

In order to test health literacy as a potential moderator of the mediation effects, individuals were categorized as having a low (ie, score  $\leq 5.14$ , 379/968) or high (ie, score  $\geq 5.15$ , 589/968) health literacy level using a mean split. Results showed that in low health literate individuals, interactivity effects on recall were mediated significantly only by cognitive involvement ( $b=-.18$ , 95% CI  $-0.37$  to  $-0.03$ , 7% mediated effect) in the moderate interactivity condition. When levels of interactivity were high,

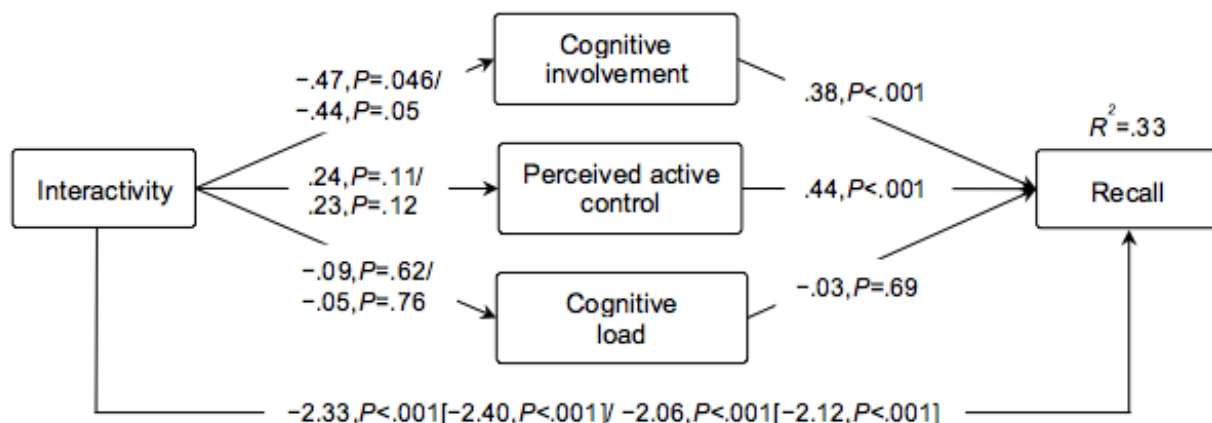
no mediation effects were found in low health literate individuals.

In high health literate individuals exposed to the moderate interactivity condition, interactivity effects were mediated by cognitive involvement ( $b=-.21$ , 95% CI  $-0.36$  to  $-0.09$ , 10% mediated effect) and perceived active control ( $b=.42$ , 95% CI  $0.28$  to  $0.60$ , 18% mediated effect). In the high interactivity condition, cognitive involvement ( $b=-.25$ , 95% CI  $-0.41$  to  $-0.12$ , 11% mediated effect) and perceived active control ( $b=.40$ , 95% CI  $0.22$  to  $0.61$ , 20% mediated effect) also partially mediated the effect. Direct and total effects are presented in Figure 7.

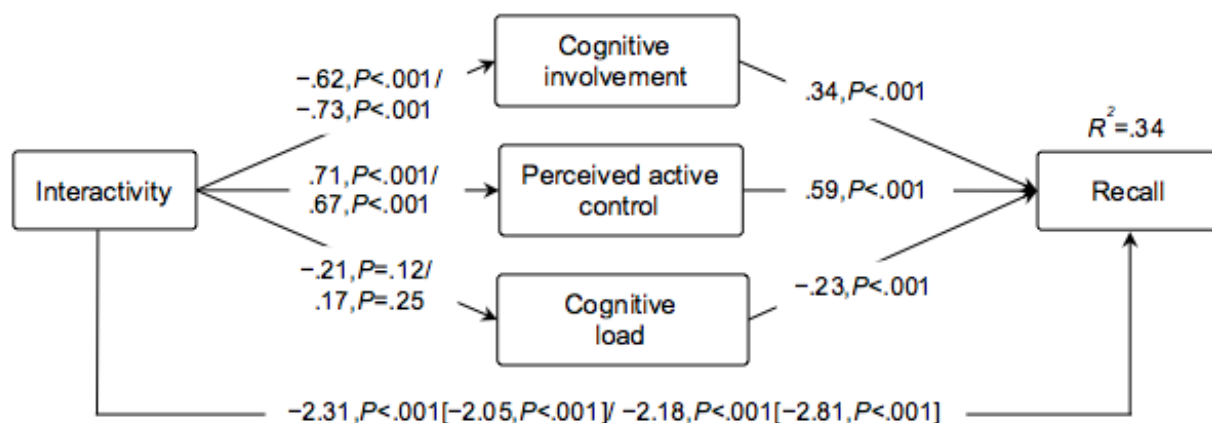


**Figure 7.** Unstandardized path coefficients of the direct and total effects (in brackets) of moderate and high interactivity on the mediators and recall compared with the control condition in the low (n=379) and high (n=589) health literacy group.

### Low health literacy group



### High health literacy group



## Discussion

### Principal Findings

Our study aimed to examine whether and how cognitive involvement, perceived active control, and cognitive load influenced the relationship between moderate and high levels of interactivity and recall. In addition, we looked at whether those pathways differed according to individuals' level of need for cognition and health literacy. Two out of the three mediation effects we tested were significant: a negative indirect effect of interactivity on recall was found through cognitive involvement and a positive indirect effect was found through active control. Cognitive load did not mediate the relation between interactivity and recall. In addition, the mediations found were partial since beside indirect effects, interactivity also had direct negative effects on recall. Participants in the moderate and high interactivity conditions remembered less information compared with the control condition.

With regard to cognitive involvement, we found a small, negative indirect effect that indicates that moderate and high levels of interactivity reduced recall through reduced cognitive involvement (ie, fewer thoughts generated). According to the

ELM, individuals' cognitive responses such as the number of thoughts generated increase if information processing follows the central route [58]. Our results may indicate that higher levels of interactivity hinder systematic information processing, resulting in less recall. This assumption is in line with previous research in which participants generated significantly fewer thoughts when exposed to interactive features (vs noninteractive features) [59]. A possible explanation for these findings might be that when individuals are exposed to interactive media that require different types of user actions (eg, reading, navigating), individuals may split their cognitive resources between the tasks. Since information processing capacities are already limited, this may lead to less conscious thinking about the message content [22,60]. Jeong and Hwang [61] found that media multitasking hindered systematic information processing resulting in reduced levels of attention, comprehension, and recall.

Our results showed that moderate and high levels of interactivity improved recall indirectly through enhanced perceptions of active control. In terms of dual process models, this may provide additional evidence that users tend to take the peripheral route of information processing when using interactive media. Sundar and Limperos [62] argue that new media offers several types

of technical affordances (eg, navigability) that may serve as cues (ie, snap judgments). For instance, the browsing heuristic refers to users' online information-seeking behavior in which they skim the site content and check out the menus or hyperlinks superficially [63]. Indeed, our data on user activity showed that participants did not make use of all interactive functions of the moderately and highly interactive websites. However, the magnitude of the mediation effect of perceived active control was comparable in both experimental conditions. As previously suggested, perceived active control may not be a function of the quantity but of the type of interactive features presented on a website [64].

Since we found no mediation effect of cognitive load, we suppose that moderate and high levels of interactivity are not more cognitively demanding in terms of information processing than static content (ie, control condition). Website complexity is a function of page length; amount of information presented; and number of pictures, hyperlinks, or other elements embedded in the website [65]. Our nonsignificant finding might be explained by the fact that we did not vary website content across conditions; we only varied the amount of interactive tools through which participants could interact with the website.

The direct, negative effect of interactivity on recall might be explained by user activity data revealing that both time spent on the website and amount of content visited were the highest in the control condition. This may imply that if users are provided with interactive features, their information search becomes more purposive, and the increased selectivity exposes them to less information, resulting in less recall.

It should be noted that in our study we examined the effects of cognitive involvement, perceived active control, and cognitive load in three separate pathways between interactivity and recall. However, these concepts should be examined in relation to each other as well. For instance, users' sense of high active control might be related to cognitive involvement. Active control entails autonomous user actions that might be driven by intrinsic interest, which is positively related to focused attention [31]. Therefore, future research should examine whether and how these concepts are related to each other.

### Differences in Mediation Effects With Regard to Need for Cognition and Health Literacy

The partial negative mediation effect of cognitive involvement and partial positive mediation effect of perceived active control

were of greater magnitude in individuals with high (vs low) need for cognition and in individuals with high (vs low) health literacy. While need for cognition and health literacy are generally associated with higher levels of elaboration and recall of (health) information [27,66,67], such associations were not found in research on interactive media [28,68], or negative associations were found [69]. Since literature suggests that interactive websites are preferred more by individuals with low need for cognition than their counterparts [29], we assume that interactivity distracts high need for cognition individuals from systematic information processing.

### Limitations

Our study had some limitations. First, we might have found less contrast between low and high health literacy groups since 87.2% of the sample (844/968) had adequate health literacy, according to the categorization of the newest vital sign [50,51]. Second, we did not measure recall of interactive and noninteractive website content separately. Accordingly, conclusions can be drawn only about individuals' overall recall of the website content. Third, this study was an online experiment that could be completed on participants' own device at home. Therefore, it is possible that not all participants paid full attention to the website as some outliers were found in the variable duration of website visit. However, analyses yielded comparable results when excluding those 66 outliers from the sample. Fifth, the topic of the study may seem very specific to some participants and appeal only to a selective group of individuals, namely those interested in nutrition and dietary supplements.

### Conclusions

Higher levels of interactivity decreased recall through reduced levels of cognitive involvement. At the same time, higher levels of interactivity increased recall through enhanced perceptions of active control. No significant mediation effects of cognitive load were found. In addition, the identified indirect effects were of greater magnitude in individuals with high (vs low) need for cognition and with high (vs low) health literacy. Beside the indirect effects, levels of interactivity decreased recall directly in all analyses.

### Acknowledgments

This research was supported by the Netherlands Food and Consumer Product Safety Authority. The authors would like to thank Done Digital Kft for website development and technical support.

### Authors' Contributions

The study was designed by PEM, SME, AO, and HdV. All authors participated in the development of the experimental stimulus and evaluated the results of the pilot tests. PEM conducted the statistical analyses, and results were interpreted with SME, AO, and HdV. PEM drafted the manuscript, and SME, AO, and HdV gave feedback on the manuscript. All authors approved the final version of the article.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Content and structure of the website about vitamin B6.

[DOCX File, 665 KB - [jmir\\_v22i10e14783\\_app1.docx](#)]

### Multimedia Appendix 2

CONSORT-EHEALTH checklist V1.6.1.

[PDF File (Adobe PDF File), 96 KB - [jmir\\_v22i10e14783\\_app2.pdf](#)]

## References

1. Sundar SS. Social psychology of interactivity in human-website interaction. In: Johnson AN, McKenna KYA, Postmes T, Reips UD, editors. Oxford Handbook of Internet Psychology. New York: Oxford University Press; 2007:89-102.
2. Oh J. The effect of interactivity on smokers' intention to quit: a linear or curvilinear relationship? *Comput Human Behav* 2017 Oct;75:845-854. [doi: [10.1016/j.chb.2017.05.049](#)]
3. Yang F, Shen F. Effects of web interactivity: a meta-analysis. *Comm Res* 2017 Mar 27;45(5):635-658. [doi: [10.1177/0093650217700748](#)]
4. van Rossum C, Buurma-Rethans E, Vennemann F, Beukers M, Brants HA, De Boer E. The diet of the Dutch: results of the first two years of the Dutch National Food Consumption Survey 2012-2016. National Institute for Public Health and the Environment. 2016. URL: <https://www.rivm.nl/bibliotheek/rapporten/2016-0082.pdf> [accessed 2020-05-30]
5. Gahche J, Bailey R, Burt V, Hughes J, Yetley E, Dwyer J, et al. Dietary supplement use among U.S. adults has increased since NHANES III (1988-1994). *NCHS Data Brief* 2011 Apr(61):1-8 [FREE Full text] [Medline: [21592424](#)]
6. Bucy EP, Tao C. The mediated moderation model of interactivity. *Media Psychol* 2007 May 15;9(3):647-672. [doi: [10.1080/15213260701283269](#)]
7. Sundar SS, Xu Q, Bellur S, Oh J, Jia H. Modality is the message: interactivity effects on perception and engagement. 2010 Presented at: 10th Conference on Human Factors in Computing Systems; 2010; Atlanta. [doi: [10.1145/1753846.1754110](#)]
8. Liu Y. Developing a scale to measure the interactivity of websites. *J Advert Res* 2003 Jun 01;43(2):207-216. [doi: [10.2501/jar-43-2-207-216](#)]
9. Rafaeli S. Interactivity: from new media to communication. In: Hawkins RP, Wiemann JM, Pingree S, editors. *Advancing Communication Science: Merging Mass and Interpersonal Processes*. Newbury Park: Sage; 1988:110-134.
10. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. In: Petty RE, Cacioppo JT, editors. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag; 1986:1-24.
11. Cacioppo JT, Petty RE. Social psychological procedures for cognitive response assessment: the thought listing technique. In: Merluzzi TV, Glass CR, Genest M, editors. *Cognitive Assessment*. New York: Guilford Press; 1981:309-342.
12. Jacobson MJ, Spiro RJ. Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: an empirical investigation. *J Educ Comput Res* 2016 Nov 28;12(4):301-333. [doi: [10.2190/4t1b-hbp0-3f7e-j4pn](#)]
13. Jonassen DH. Semantic network elicitation: tools for structuring hypertext. In: McAleese R, Green C, editors. *Hypertext State of the Art*. Oxford: Intellect; 1990:142-153.
14. Merrill MD. Learner control in computer based learning. *Computers & Education* 1980 Jan;4(2):77-95. [doi: [10.1016/0360-1315\(80\)90010-X](#)]
15. Patterson NG. Hypertext and the changing roles of readers. *English J* 2000 Nov;90(2):74-80. [doi: [10.2307/821221](#)]
16. Tremayne M. Lessons learned from experiments with interactivity on the web. *J Interact Advert* 2005 Mar;5(2):40-46. [doi: [10.1080/15252019.2005.10722100](#)]
17. McMillan SJ, Hwang JS, Lee G. Effects of structural and perceptual factors on attitudes toward the website. *JAR* 2003 Dec 01;43(4):400-409. [doi: [10.1017/S0021849903030393](#)]
18. Liu Y, Shrum LJ. What is interactivity and is it always such a good thing? Implications of definition, person, and situation for the influence of interactivity on advertising effectiveness. *J Advert* 2002 Dec;31(4):53-64. [doi: [10.1080/00913367.2002.10673685](#)]
19. Shrum LJ, Lowrey TM, Liu Y. Emerging issues in advertising research. In: Nabi RL, Oliver MB, editors. *The Sage Handbook of Media Processes and Effects*. Thousand Oaks: Sage; 2009:299-312.
20. Tremayne M, Dunwoody S. Interactivity, information processing, and learning on the world wide web. *Sci Comm* 2016 Aug 18;23(2):111-134. [doi: [10.1177/1075547001023002003](#)]
21. Sweller J. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* 1994 Jan;4(4):295-312. [doi: [10.1016/0959-4752\(94\)90003-5](#)]
22. Lang A. The limited capacity model of mediated message processing. *J Commun* 2000;50(1):46-70. [doi: [10.1111/j.1460-2466.2000.tb02833.x](#)]

23. Edwards MB, Gronlund SD. Task interruption and its effects on memory. *Memory* 1998 Nov;6(6):665-687. [doi: [10.1080/741943375](https://doi.org/10.1080/741943375)] [Medline: [10320869](https://pubmed.ncbi.nlm.nih.gov/10320869/)]
24. Lee J, Lin L, Robertson T. The impact of media multitasking on learning. *Learning Media Technol* 2011 Apr 27;37(1):94-104. [doi: [10.1080/17439884.2010.537664](https://doi.org/10.1080/17439884.2010.537664)]
25. Cacioppo JT, Petty RE. The need for cognition. *J Personality Soc Psychol* 1982 Jan;42(1):116-131. [doi: [10.1037/0022-3514.42.1.116](https://doi.org/10.1037/0022-3514.42.1.116)]
26. Cacioppo JT, Petty RE, Feng Kao C. The efficient assessment of need for cognition. *J Personality Assess* 2010 Jun 10;48(3):306-307. [doi: [10.1207/s15327752jpa4803\\_13](https://doi.org/10.1207/s15327752jpa4803_13)] [Medline: [16367530](https://pubmed.ncbi.nlm.nih.gov/16367530/)]
27. Cacioppo JT, Petty RE, Feinstein JA, Jarvis WBG. Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. *Psychol Bull* 1996;119(2):197-253. [doi: [10.1037/0033-2909.119.2.197](https://doi.org/10.1037/0033-2909.119.2.197)]
28. Sicilia M, Ruiz S, Munuera JL. Effects of interactivity in a web site: the moderating effect of need for cognition. *J Advert* 2005 Nov;34(3):31-44. [doi: [10.1080/00913367.2005.10639202](https://doi.org/10.1080/00913367.2005.10639202)]
29. Amichai-Hamburger Y, Kaynar O, Fine A. The effects of need for cognition on Internet use. *Comput Human Behav* 2007 Jan;23(1):880-891. [doi: [10.1016/j.chb.2006.03.002](https://doi.org/10.1016/j.chb.2006.03.002)]
30. Kaynar O, Amichai-Hamburger Y. The effects of need for cognition on internet use revisited. *Comput Human Behav* 2008 Mar;24(2):361-371. [doi: [10.1016/j.chb.2007.01.033](https://doi.org/10.1016/j.chb.2007.01.033)]
31. Srivastava K, Shukla A, Sharma N. Online flow experiences: the role of need for cognition, self-efficacy, and sensation seeking tendency. *Int J Bus Insights Transformat* 2010;3(2):93-100.
32. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012 Jan 25;12:80 [FREE Full text] [doi: [10.1186/1471-2458-12-80](https://doi.org/10.1186/1471-2458-12-80)] [Medline: [22276600](https://pubmed.ncbi.nlm.nih.gov/22276600/)]
33. Jensen JD, King AJ, Davis LA, Guntzville LM. Utilization of internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. *J Aging Health* 2010 Sep;22(6):804-826. [doi: [10.1177/0898264310366161](https://doi.org/10.1177/0898264310366161)] [Medline: [20495159](https://pubmed.ncbi.nlm.nih.gov/20495159/)]
34. Birru MS, Monaco VM, Charles L, Drew H, Njie V, Bierria T, et al. Internet usage by low-literacy adults seeking health information: an observational analysis. *J Med Internet Res* 2004 Sep 03;6(3):e25 [FREE Full text] [doi: [10.2196/jmir.6.3.e25](https://doi.org/10.2196/jmir.6.3.e25)] [Medline: [15471751](https://pubmed.ncbi.nlm.nih.gov/15471751/)]
35. Zarcadoolas C, Blanco M, Boyer JF, Pleasant A. Unweaving the Web: an exploratory study of low-literate adults' navigation skills on the World Wide Web. *J Health Commun* 2002 Jul;7(4):309-324. [doi: [10.1080/10810730290088157](https://doi.org/10.1080/10810730290088157)] [Medline: [12356289](https://pubmed.ncbi.nlm.nih.gov/12356289/)]
36. Lawless KA, Brown SW. Multimedia learning environments: issues of learner control and navigation. *Instruct Sci* 1997;25(2):117-131. [doi: [10.1023/A:1002919531780](https://doi.org/10.1023/A:1002919531780)]
37. Meppelink CS, van Weert JCM, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. *J Med Internet Res* 2015 Jan 13;17(1):e11 [FREE Full text] [doi: [10.2196/jmir.3979](https://doi.org/10.2196/jmir.3979)] [Medline: [25586711](https://pubmed.ncbi.nlm.nih.gov/25586711/)]
38. I&O Research. I&O Research Panel 2017. URL: <https://ioresearch.nl/Home/Aanbod> [accessed 2020-05-30]
39. O'Connor A, Stacey D, Boland L. Ottawa Decision Support Tutorial (ODST). The Ottawa Hospital Research Institute. 2015. URL: <https://decisionaid.ohri.ca/odst/pdfs/ODST.pdf> [accessed 2020-05-30]
40. Sundar SS, Bellur S, Oh J, Xu Q, Jia H. User experience of on-screen interaction techniques: an experimental investigation of clicking, sliding, zooming, hovering, dragging, and flipping. *Human Comput Interact* 2013 Dec 27;29(2):109-152. [doi: [10.1080/07370024.2013.789347](https://doi.org/10.1080/07370024.2013.789347)]
41. Kintsch W. The role of knowledge in discourse comprehension: a construction-integration model. *Psychol Rev* 1988 Apr;95(2):163-182. [Medline: [3375398](https://pubmed.ncbi.nlm.nih.gov/3375398/)]
42. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. *Educ Psychol* 2003 Mar;38(1):63-71. [doi: [10.1207/S15326985EP3801\\_8](https://doi.org/10.1207/S15326985EP3801_8)]
43. Eggemeier FT, Wilson GF. Performance-based and subjective assessment of workload in multi-task environments. In: Damos DL, editor. *Multiple-Task Performance*. London: Taylor & Francis; 1991:217-278.
44. Reid GB, Potter SS, Bressler JR. Subjective workload assessment technique (SWAT): a user's guide. Armstrong Aerospace Medical Research Laboratory, AAMRL-TR-89-023. 1989. URL: <http://www.dtic.mil/dtic/tr/fulltext/u2/a215405.pdf> [accessed 2020-05-30]
45. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. *Human Mental Workload*. Amsterdam: North Holland Press; 1988:139-184.
46. Pieters RGM, Verplanken B, Modde JM. "Neiging tot nadenken": Samenhang met beredeneerd gedrag ["Need for Cognition": the connection with reasoned behavior]. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden [Dutch Journal of Psychology and its Related Fields]* 1987;42(2):62-72.
47. Dornic S, Ekehammar B, Laaksonen T. Tolerance for mental effort: Self-ratings related to perception, performance and personality. *Personality and Individual Differences* 1991;12(3):313-319. [doi: [10.1016/0191-8869\(91\)90118-U](https://doi.org/10.1016/0191-8869(91)90118-U)]
48. Sadowski CJ. An examination of the short need for cognition scale. *J Psychol* 1993 Jul;127(4):451-454. [doi: [10.1080/00223980.1993.9915581](https://doi.org/10.1080/00223980.1993.9915581)]



49. Hevey D, Thomas K, Pertl M, Maher L, Craig A, Ni CS. Method effects and the need for cognition scale. *Int J Educ Psychol Assess* 2012;12(1):20-33.
50. Fransen MP, Leenaars KEF, Rowlands G, Weiss BD, Maat HP, Essink-Bot ML. International application of health literacy measures: adaptation and validation of the newest vital sign in The Netherlands. *Patient Educ Couns* 2014 Dec;97(3):403-409. [doi: [10.1016/j.pec.2014.08.017](https://doi.org/10.1016/j.pec.2014.08.017)] [Medline: [25224314](https://pubmed.ncbi.nlm.nih.gov/25224314/)]
51. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med* 2005;3(6):514-522 [FREE Full text] [doi: [10.1370/afm.405](https://doi.org/10.1370/afm.405)] [Medline: [16338915](https://pubmed.ncbi.nlm.nih.gov/16338915/)]
52. Sundar SS, Kalyanaraman S, Brown J. Explicating web site interactivity. *Commun Res* 2016 Jun 30;30(1):30-59. [doi: [10.1177/0093650202239025](https://doi.org/10.1177/0093650202239025)]
53. Clifton B. *Advanced Web Metrics with Google Analytics*. 3rd Edition. Indianapolis: John Wiley & Sons; 2012.
54. Zaichkowsky JL. The personal involvement inventory: reduction, revision, and application to advertising. *J Advertis* 1994 Dec;23(4):59-70. [doi: [10.1080/00913367.1993.10673459](https://doi.org/10.1080/00913367.1993.10673459)]
55. Hayes AF. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. 1st Edition. New York: Guilford Press; 2013.
56. Russell CJ, Dean MA. To log or not to log: bootstrap as an alternative to the parametric estimation of moderation effects in the presence of skewed dependent variables. *Organiz Res Meth* 2016 Jun 29;3(2):166-185. [doi: [10.1177/109442810032002](https://doi.org/10.1177/109442810032002)]
57. Hayes A. What's coming in PROCESS v3.0. 2017. URL: <https://processmacro.org/uploads/3/4/8/5/34858384/processv3overview.pdf> [accessed 2020-05-30]
58. Petty RE, Cacioppo JT, Strathman AJ, Priester JR. To think or not to think: exploring two routes of persuasion. In: Brock TC, Green MC, editors. *Persuasion: Psychological Insights and Perspectives*. 2nd Edition. Thousand Oaks: Sage; 2005:81-116.
59. Oh J, Bellur S, Sundar SS. Clicking, assessing, immersing, and sharing: an empirical model of user engagement with interactive media. *Commun Res* 2015 Sep 21;45(5):737-763. [doi: [10.1177/0093650215600493](https://doi.org/10.1177/0093650215600493)]
60. Wu LL, Lin JY. The match between information control and motivation in the online context. *Psychol Mark* 2012 Oct 11;29(11):822-835. [doi: [10.1002/mar.20567](https://doi.org/10.1002/mar.20567)]
61. Jeong S, Hwang Y. Media multitasking effects on cognitive vs. attitudinal outcomes: a meta-analysis. *Hum Commun Res* 2016 Jul 08;42(4):599-618. [doi: [10.1111/hcre.12089](https://doi.org/10.1111/hcre.12089)]
62. Sundar SS, Limperos AM. Uses and grats 2.0: new gratifications for new media. *J Broadcast Electr Media* 2013 Oct;57(4):504-525. [doi: [10.1080/08838151.2013.845827](https://doi.org/10.1080/08838151.2013.845827)]
63. Sundar SS. The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Metzger MJ, Flanagin AJ, editors. *Digital Media, Youth, and Credibility*. Cambridge: The MIT Press; 2008:73-100.
64. Voorveld HAM, Neijens PC, Smit EG. The relation between actual and perceived interactivity. *J Advert* 2011 Jul;40(2):77-92. [doi: [10.2753/JOA0091-3367400206](https://doi.org/10.2753/JOA0091-3367400206)]
65. Huang M. Designing website attributes to induce experiential encounters. *Comput Human Behav* 2003 Jul;19(4):425-442. [doi: [10.1016/S0747-5632\(02\)00080-8](https://doi.org/10.1016/S0747-5632(02)00080-8)]
66. McCarthy DM, Waite KR, Curtis LM, Engel KG, Baker DW, Wolf MS. What did the doctor say? Health literacy and recall of medical instructions. *Med Care* 2012 Apr;50(4):277-282 [FREE Full text] [doi: [10.1097/MLR.0b013e318241e8e1](https://doi.org/10.1097/MLR.0b013e318241e8e1)] [Medline: [22411440](https://pubmed.ncbi.nlm.nih.gov/22411440/)]
67. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011 Jul 19;155(2):97-107. [doi: [10.7326/0003-4819-155-2-201107190-00005](https://doi.org/10.7326/0003-4819-155-2-201107190-00005)] [Medline: [21768583](https://pubmed.ncbi.nlm.nih.gov/21768583/)]
68. Lustria MLA. Can interactivity make a difference? Effects of interactivity on the comprehension of and attitudes toward online health content. *J Am Soc Inf Sci* 2007 Apr;58(6):766-776. [doi: [10.1002/asi.20557](https://doi.org/10.1002/asi.20557)]
69. Broekhuizen T, Hoffmann A. Interactivity perceptions and online newspaper preference. *J Interact Advertis* 2012;12(2):29-43.

## Abbreviations

**ELM:** elaboration likelihood model  
**NCS:** Need for Cognition Scale  
**NVS-D:** Newest Vital Sign in Dutch



*Edited by G Eysenbach; submitted 03.07.19; peer-reviewed by C Short, M Laurent; comments to author 20.08.19; revised version received 14.12.19; accepted 24.01.20; published 28.10.20.*

*Please cite as:*

*Pajor EM, Eggers SM, de Vries H, Oenema A*

*Effects of Interactivity on Recall of Health Information: Experimental Study*

*J Med Internet Res 2020;22(10):e14783*

URL: <https://www.jmir.org/2020/10/e14783>

doi: [10.2196/14783](https://doi.org/10.2196/14783)

PMID: [33112245](https://pubmed.ncbi.nlm.nih.gov/33112245/)

©Emília Margit Pajor, Sander Matthijs Eggers, Hein de Vries, Anke Oenema. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study

Akira-Sebastian Poncette<sup>1,2</sup>, MD; Daniel Leon Glauert<sup>1</sup>; Lina Mosch<sup>1</sup>; Katarina Braune<sup>3</sup>, MD; Felix Balzer<sup>1,2</sup>, MSc, PhD, MD; David Alexander Back<sup>4,5</sup>, MD

<sup>1</sup>Department of Anesthesiology and Intensive Care Medicine, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

<sup>2</sup>Einstein Center Digital Future, Berlin, Germany

<sup>3</sup>Department of Paediatric Endocrinology and Diabetes, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

<sup>4</sup>Department of Traumatology and Orthopaedics, Septic and Reconstructive Surgery, Bundeswehr Hospital Berlin, Berlin, Germany

<sup>5</sup>Dieter Scheffner Center for Medical Education and Educational Research, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

**Corresponding Author:**

David Alexander Back, MD

Dieter Scheffner Center for Medical Education and Educational Research

Charité – Universitätsmedizin Berlin

Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health

Charitéplatz 1

Berlin, 10117

Germany

Phone: 49 30 2841 1240

Email: [david.back@charite.de](mailto:david.back@charite.de)

**Related Article:**

This is a corrected version. See correction statement: <https://www.jmir.org/2020/12/e25738/>

## Abstract

**Background:** Owing to an increase in digital technologies in health care, recently leveraged by the COVID-19 pandemic, physicians are required to use these technologies appropriately and to be familiar with their implications on patient care, the health system, and society. Therefore, medical students should be confronted with digital health during their medical education. However, corresponding teaching formats and concepts are still largely lacking in the medical curricula.

**Objective:** This study aims to introduce digital health as a curricular module at a German medical school and to identify undergraduate medical competencies in digital health and their suitable teaching methods.

**Methods:** We developed a 3-week curricular module on digital health for third-year medical students at a large German medical school, taking place for the first time in January 2020. Semistructured interviews with 5 digital health experts were recorded, transcribed, and analyzed using an abductive approach. We obtained feedback from the participating students and lecturers of the module through a 17-item survey questionnaire.

**Results:** The module received overall positive feedback from both students and lecturers who expressed the need for further digital health education and stated that the field is very important for clinical care and is underrepresented in the current medical curriculum. We extracted a detailed overview of digital health competencies, skills, and knowledge to teach the students from the expert interviews. They also contained suggestions for teaching methods and statements supporting the urgency of the implementation of digital health education in the mandatory curriculum.

**Conclusions:** An elective class seems to be a suitable format for the timely introduction of digital health education. However, a longitudinal implementation in the mandatory curriculum should be the goal. Beyond training future physicians in digital skills and teaching them digital health's ethical, legal, and social implications, the experience-based development of a critical digital health mindset with openness to innovation and the ability to assess ever-changing health technologies through a broad

transdisciplinary approach to translate research into clinical routine seem more important. Therefore, the teaching of digital health should be as practice-based as possible and involve the educational cooperation of different institutions and academic disciplines.

(*J Med Internet Res* 2020;22(10):e22161) doi:[10.2196/22161](https://doi.org/10.2196/22161)

## KEYWORDS

digital health; eHealth; mHealth; digital health education; elective module; eHealth education; curriculum, medical school, digital health mindset; qualitative research; interview; survey

## Introduction

### Background

With the progress in the introduction of digital solutions in patient care, such as electronic health records [1-3], artificial intelligence (AI) for decision support [4,5], telemedicine [6-8], or robotic surgery [9], the need for physicians and other health care professionals to get familiar with digital health is increasing. The recent COVID-19 pandemic has highlighted the advantages of remote care and puts pressure on health care professionals and infrastructure to adapt to a fast-developing, globalized world [10].

Although the introduction of new technologies in medicine is accompanied by public hope for better and more efficient patient care [11,12], experts agree that a new technology can only be as good as the physician using it [13]. Implementing digital health technologies into clinical settings remains a prolonged process [14], with one of the major barriers being the lack of health professionals' knowledge and awareness of the new technologies and the skills to use them [15].

Training physicians and nurses in the practical use of digital technologies at an early stage of their career is overdue to prepare them for their future challenges [3,13,16]. In this context, data literacy is considered a decisive skill for health care workers [17]. Digital literacy of students and young physicians, who are often referred to as the generation of digital natives, is discussed ambivalently in the literature. Although some authors postulate that growing up with digital services may lead to differentiated use in a professional context [18], other publications stress that the implication of the digital native stereotype would leave learners unsupported and technologies used in inappropriate ways, making further research in this area indispensable [19].

The relevance of implementing digital health education in medical curricula is evident [16]. Various pilot projects have been described in the literature, with differences in length and focus [20,21]. All of them reported a high level of satisfaction among students attending digital health-related courses, with expected positive influences on their skillset in later professional life [22].

According to a Europe-wide survey, medical students felt lacking digital literacy and demanded a wider implementation of digital health topics in their curricula [23,24]. Along with the need for education on ethics and technology specifics, a general introduction to the topic and the teaching of basic aspects of the field were asked for [23,24]. In 2019, the deans of 25 European medical universities agreed on the rapid implementation of digital health education in their respective

medical schools' curricula, focusing on interprofessional education, practical skills, and innovation [25]. To scale up the implementation of digital health in health care education, there is still a need to provide and exchange best practices of digital health in medical education. The ongoing public funding of digital health research projects has so far not been correlated with an increase in the number of corresponding courses in the medical curriculum [23]. The implementation of teaching initiatives for digital competencies should urgently be improved.

### Aim

We intended to deliver a proof of concept of teaching digital health at a medical school, describing the development, introduction, and evaluation of a 3-week elective module. We further aimed to identify undergraduate medical competencies in digital health and their suitable teaching methods.

## Methods

### Ethics Approval and Consent to Participate

Ethical approval for this study was granted by the ethics committee of the Charité—Universitätsmedizin Berlin (EA1/236/19). Participation in the study was voluntary. Before the study, all participants provided their consent.

### Study Setting

This mixed methods study took place at the medical school of a large German university hospital, the Charité—Universitätsmedizin Berlin, in the context of the development and deployment of a pioneer teaching module for digital health, which was realized in January 2020. At the time of the study, approximately 7500 students were enrolled at the medical school [26]. Although medical informatics was taught as an independent study program since 1972 in Heidelberg, Germany, and was mentioned in the Medical Licensure Act (*Approbationsordnung für Ärzte*) since 1989, medical informatics is only marginally taught in the German medical curriculum. The terms *digital* or *digital health* are still not mentioned in the Medical Licensure Act in Germany. By the end of the third year of the undergraduate modular medical track, students at the Charité have to choose an elective module from a variety of 3-week course programs covering different medical fields that are usually not covered by the mandatory curriculum. The modules are in direct competition for the students' interest and comprise 60 teaching hours (45 min each). Only the modules selected the most often by the students take place.

Over the course of 2 years, we developed a teaching module for digital health based on qualitative results of semistructured interviews with experts described below and by reviewing

similar existing projects. The module was designed to change over time, flexibly reacting to the feedback received and to new scientific findings. In multiple face-to-face meetings, calls, and emails with potentially interested parties, we identified 32 lecturers from 16 departments of our university hospital as well as from 6 academic and nonacademic partner organizations who were willing to participate.

The cooperation with the lecturers led to a high variety of digital health topics, such as lessons on law, ethics, and economics as well as on digital pharmacology or AI and big data in research. In addition to the theoretical teaching in seminars, our main focus was to provide hands-on experience with digital health technologies. Practice units on smart implants and wearables, symptom checker apps, telecardiology, mixed reality-assisted surgery, and video consultations were included with the help of experts and developers of the corresponding technologies. To enable the students to play the roles of different health system stakeholders, we encouraged them to compete in groups during a mini-hackathon. During the course of the whole module, each student group invented and eventually pitched a new product for a specific problem in health care. Further details and lessons can be found in the module timetable ([Multimedia Appendix 1](#)).

## Study Design

We chose a mixed methods approach consisting of an abductive, qualitative study based on semistructured interviews and a cross-sectional survey study using a web-based questionnaire. Qualitative data included the interview transcripts and results from the open-ended questions of the questionnaire. Items from the questionnaire with a five-point Likert-type scale as a response format were considered quantitative data.

## Data Collection

Between October and December 2019, DG conducted semistructured interviews with 5 experts in digital health and

medical education. Purposive sampling was deployed on the research team's professional external networks to select experts with complementary backgrounds in Europe.

We developed the interview guideline according to the research question ([Textbox 1](#)) and tested and adjusted the questions during pilot interviews within the research team. The interviews were conducted in English or German via phone calls and were recorded and transcribed verbatim by the interviewer (DG). Interview transcripts were reviewed by AP and LM. The median interview length was 31 min (range 27-50 min).

Furthermore, we collected feedback on the module and further digital health topics from the participating lecturers and via corresponding surveys. Survey items were generated through a literature review and informal research meetings ([Multimedia Appendix 2](#)). We grouped 17 items into 3 topics and chose a five-point Likert-type scale as an ordinal response format, with the options strongly agree (1), agree (2), neutral (3), disagree (4), and strongly disagree (5), and 4 open-ended questions. Respondents were also given space to comment on each topic. The 3 topics were (1) digital health at medical schools, (2) experience from the digital health module, and (3) feedback on the organization of the digital health module.

Pretests within the research team did not alter the questionnaire. Through pilot tests with associated research colleagues, clarity, relevance, and arrangement of the questionnaire items were improved.

The survey questionnaire was given to the participated students on the last day of the module as a paper version and slightly modified for participating lecturers of the module. For the latter, we sent an email invitation with multiple reminders to participate in the survey to all 32 lecturers. This web-based survey data were collected and managed using REDCap (Research Electronic Data Capture) tools hosted at Charité—Universitätsmedizin Berlin [27,28]. No incentive or compensation was given to the survey participants.

**Textbox 1.** Guideline questions for the semistructured interviews.

1. What is your relation to digital health and what are you currently dedicating to professionally?
2. How do you think will digitalization affect doctors' work in the future, and what are the biggest challenges?
3. What topics of digital health would you recommend to teach at German medical schools?
4. Which teaching methods would you imagine or consider to be the most effective for teaching digital health?
5. How can awareness for the challenges of digital health be maximized among medical students?
6. Based on your own experience, do you have further advice on what factors should be considered when teaching digital health?

## Data Analysis

We performed an abductive analysis of the interview transcripts and the results of the open-ended questions of the questionnaire from the lecturers and students to identify the predominant themes [26]. Predefined themes included teaching format and learning objectives. The theme consisted of multiple codes that were later defined as subthemes. The saturation of codes was achieved with a growing number of interviews to ensure covering the majority of the aspects connected to our research question. All coding was performed and reviewed by DG, AP,

and LM using MaxQDA qualitative data analysis software (MaxQDA 18.3.2; VERBI GmbH).

Descriptive data analysis of questionnaire items was conducted using Microsoft Excel 2020 (version 16.35).

## Results

### Overview

We constructed this mixed methods study based on 5 interviews with digital health experts and a questionnaire involving

lecturers and students. The sunburst diagram (Figure 1) visualizes the qualitative results, specifically the 4 themes in the inner ring with the most relevant subthemes in the middle ring that are specified in the outer ring.

**Figure 1.** This sunburst diagram represents the qualitative results. Within the 4 themes (inner ring), subthemes (middle ring) are assigned and specified (outer ring). ELSI: ethical, legal, and social implications.



The quantitative data were represented by the questionnaires, where the response rate was 91% (10/11) for the students and 100% (32/32) for the lecturers.

## Qualitative Results

### *Need for Digital Health*

#### **In the Health Care System**

According to the respondents, the positive effects of digital health on the health care system were clearly visible in areas such as communication, documentation, and patient empowerment. Digital health technologies would be already integrated into physicians' everyday lives and were

indispensable, as younger patients, in particular, would explicitly demand for them.

#### **In the Medical School Curriculum**

It was stated that the health care system of the future would be highly digitalized. It would, therefore, be necessary to prepare medical students accordingly as early as possible. Digital health would already be part of today's clinical practice, and concerns were expressed that a lack of knowledge in this field would lead to individual failure. Medical students would have to lose their initial reservations and become critical experts of digital health as much as their analog counterparts.



## ***Learning Objectives—Skills***

### **Digital Literacy**

Respondents emphasized that training with digital technologies from various perspectives, such as engineering, law, data protection, and ethics as well as statistical knowledge in the context of evidence-based medicine, should be part of the curriculum. The digital literacy acquired in this way would include an understanding of the meaningfulness and application areas of AI, robotics, big data, and telemedicine and thus help in one's own clinical work and in cooperation with other professional groups. Increasingly, patients would use digital services and apps, but convincing doctors to use them would remain to be a major challenge. As health apps are already being used in practice, their medical, legal, ethical, and economic implications should be a part of medical education and training.

Digital literacy would describe the sensitivity, confidence, and understanding with which physicians could apply new digital applications to promote health. Respondents pointed out that by gaining practical experiences with digital health, medical school graduates should be able to use a wide variety of digital health technologies. Interpersonal skills such as intuition and sensory experience in patient contact would gain importance, especially regarding diagnostics. Health care staff would have to be able to communicate in various novel ways (eg, apps, telemedicine). Therefore, expertise in a variety of communication methods with a special emphasis on remote care should be acquired in medical schools.

### **Innovation Capabilities and Entrepreneurial Skills**

Adapting to the constantly changing professional environment would require physicians to have their own inventive spirit and a lifelong willingness to progress and rethink. Physicians would be responsible for proactively shaping the transformation of medicine. This role could only be fulfilled by the corresponding conviction to be innovative. Innovation methods such as design thinking could be helpful to act visionary in this context.

Future medical school graduates should be trained to apply entrepreneurial skills in a digitalized and data-rich health care system to promote health care, according to the respondents. They should be able to think entrepreneurially and act innovatively, searching for new ways to sustainably integrate innovation into clinical workflows.

### **Transdisciplinary Cooperation and Open Access Mentality**

Problems could be approached and solved confidently in transdisciplinary cooperation. The respondents stated that this should also be taken into account in the teaching concept for digital medicine. Cooperation with other disciplines, such as computer science, nursing, administration, user experience design, industry, law, data protection, and ethics, should be trained at an early stage because mutual understanding of the roles would facilitate communication.

In our health care system, including medical schools, much more cooperation and sharing of knowledge is needed. E-learning should ideally be publicly available.

### **Patient-Centered Approach**

Physicians should be given the freedom to focus on patients rather than on technology. For this purpose, their focus should be on human interaction. For physicians, digital health would be about the ability to advise mature, proactive patients sensibly and to accompany them through the health care system appropriately through direct communication using digital health wherever appropriate.

### **Fearless of Change**

Lack of awareness and experience in using digital technologies would fuel reservations in health care workers, according to the respondents. Actors, especially in leading positions, would feel that their professional authority or seniority was being questioned. Trying out and getting to know new possibilities in a relaxed manner could reduce fears, even among older generations and up to executive levels, and lead to successful clinical implementation.

## ***Learning Objectives—Knowledge***

### **Data Literacy**

Clinical decisions were increasingly based on complex data. Identifying the relevant data and deriving good decisions from it would require a high degree of data competence from future physicians. Statistics would remain an important basis for scientific action in this context. It was stated that the responsible and sensible use of systems based on AI is an important medical skill and should become a compulsory learning objective for medical students.

### **Ethical, Legal, and Social Implications**

Medicine in its current form would be unethical according to the respondents. Digitalization and modern data science could improve this by analyzing patient data to make clinical care more based on objective evidence. Often, issues in digital health would lead to ethical questions, which should be discussed by experts and students using exemplary situations. Legal aspects, in particular regarding the regulation of medical practice and medical devices, should be taught by experts during the course of studies. The culture of societal change in the context of digitalization should be a central component of the curriculum.

### **Data Protection**

According to the interviewees, data protection, data security, and privacy were one of the greatest challenges for physicians in the course of digitalization, for example, in the case of mobile health apps. They should be taught transparently and according to clear guidelines by appropriate experts. As most patients care about the privacy of their health data, medical students should learn to ensure that their actions are in accordance with the correct data protection regulations.

### **Health Economics**

Basic economic knowledge and health economic aspects such as financing and resource optimization were missing in medical schools' curricula and should be taught specifically for digital health.

## Teaching Format

### Practical Experiences

The direct handling of new technologies would reduce fears and promote enthusiasm for digital technologies. *Learning by doing* would be the preferred form to learn about technical applications because it creates experiences, helps to shape interests, and ensures a *view beyond the end of one's own nose* (interview quote). The goal of teaching digital health would be its successful clinical implementation. Therefore, existing technologies should be included in the medical curriculum and demonstrated by experienced lecturers. Students should be trained practically involving direct contact with patients and technologies to handle unexpected new situations in relation to digital health and encouraged to be creative and culturally reflective by creating their own content, for example, in the course of a hackathon.

### Lectures

Short impulse lectures would be a useful teaching method, especially for large groups. Frontal lectures could be more effective for deepening a topic or as a supplement to practical training, especially if they do not last longer than 10 min.

### Open Discourse

An open-ended discourse on the values, visions, and competencies of medical students among themselves and with various stakeholders in digital health (patients, physicians, nurses, etc) would lead to a high degree of reflection and heightened awareness of the relevance of the topic. Thus, critics and enthusiasts would be simultaneously involved.

### Other Proposed Teaching Formats

Small group lessons in (carefully selected) groups of up to 10 students would ensure more enthusiasm for the course content, stated the respondents. Peer teaching would be very effective and bring individual experience into the curriculum. The creation of independent work is recommended as a teaching format. E-learning would be a suitable format, especially for theoretical content, and could be used in a scalable manner. In general, digital medicine teaching should be integrated longitudinally into the curriculum.

### Cooperation

#### Cross-Faculty

Digital health could be taught as a study component of various study programs, such as medicine, economics, or sociology.

According to respondents, transdisciplinary teaching was more attractive and stimulating. Interuniversity exchange would be profitable, and joint courses would be useful for students and lecturers.

### Cross-Institutional and Cross-Functional

In the teaching of digital health, lecturers, students, scientists, alumni, and experts from all participating study programs as well as external experts would be partners. It would also be advisable for the lecturers to consult with each other to ensure continuity and interlocking of the teaching content.

### Industry and Business

The respondents believe that integration of different actors from industry and business into the teaching of digital health at medical schools would help both students and companies to benefit and learn from each other.

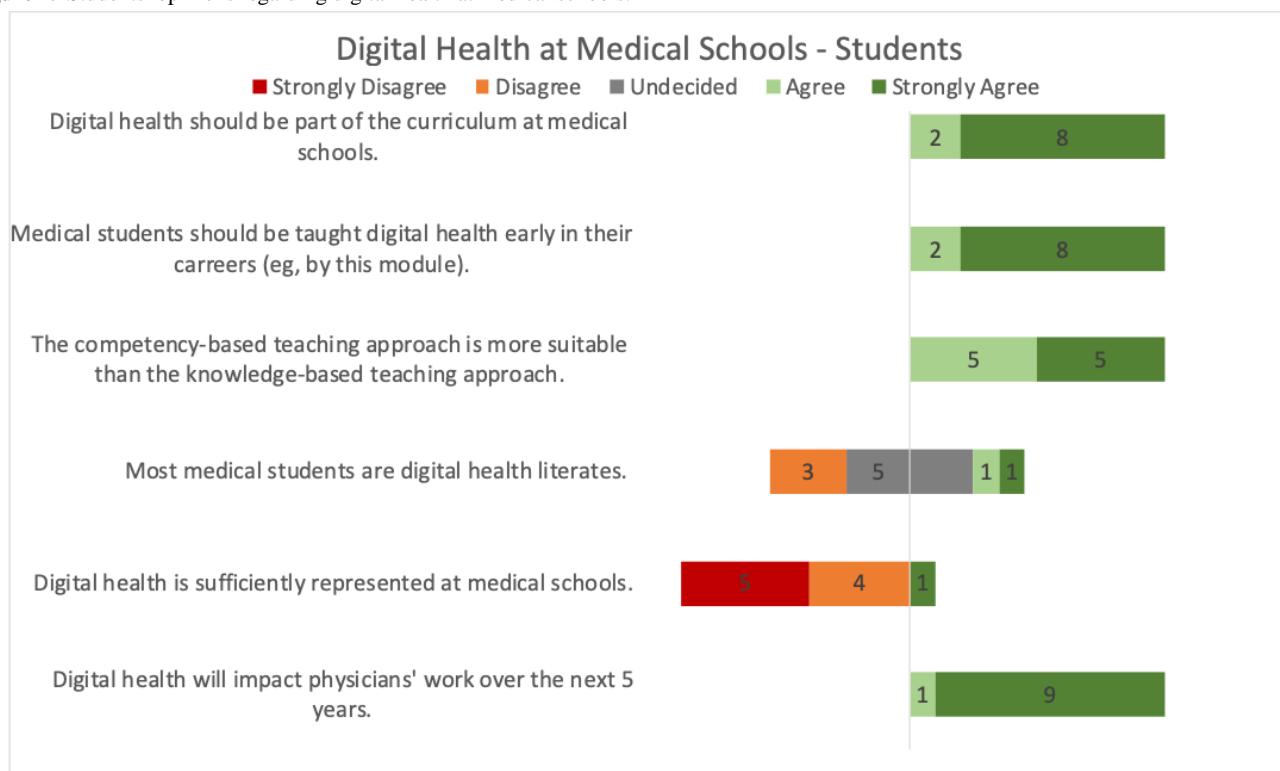
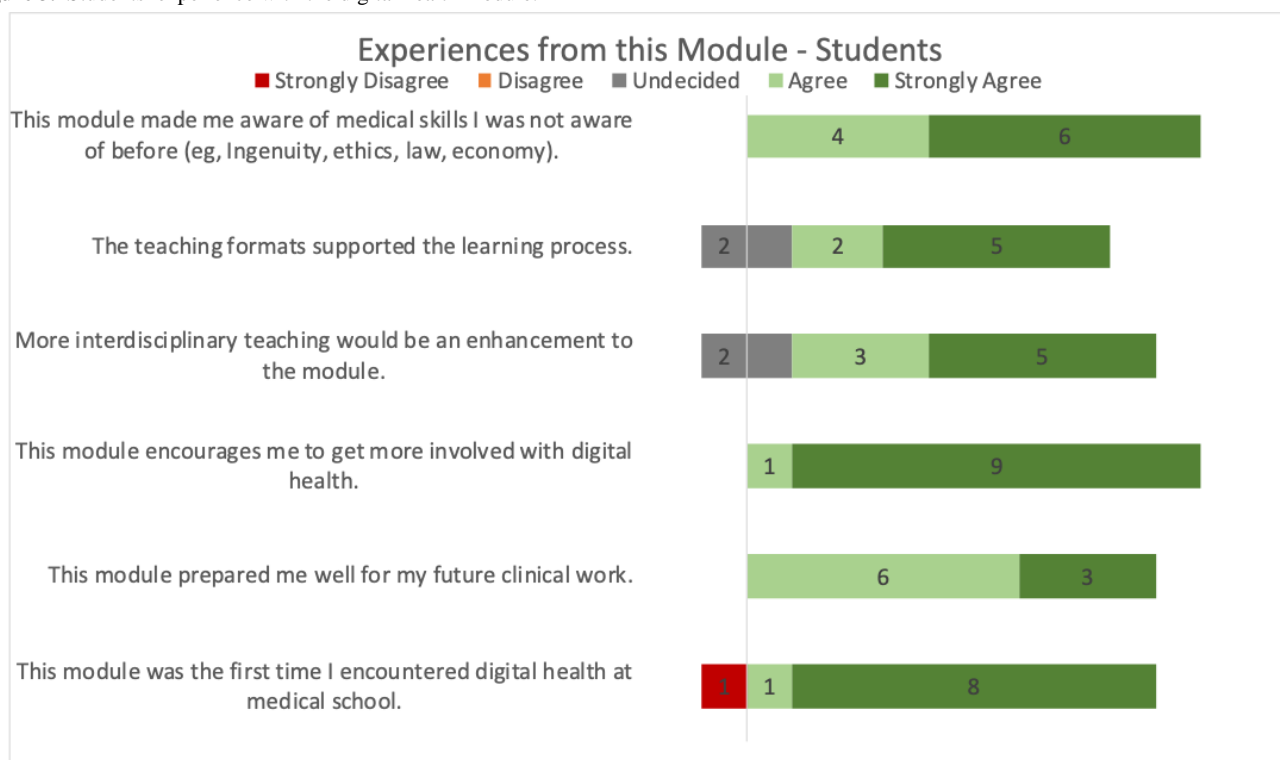
## Quantitative Results

### Students' Evaluation

All participating students agreed that digital health should be part of the mandatory curriculum at medical schools (8 strongly agree and 2 agree) and taught early in their careers (8 strongly agree and 2 agree) and that competencies are more relevant than knowledge for medical graduates (5 strongly agree and 5 agree; [Figure 2](#)). Regarding whether medical students are digitally literate, the opinion was split. Most students strongly disagreed or disagreed that digital health is sufficiently represented at medical schools (5 strongly disagree, 4 disagree, and 1 agree). All students agreed that digital health would impact physicians' work over the next 5 years (9 strongly agree and 1 agree).

Regarding the experiences from this module, all students agreed that this module made them aware of medical skills that they were not aware of before (6 strongly agree and 4 agree), that the teaching format helped them in the study process (5 strongly agree, 2 agree, and 2 neutral), and that interdisciplinary teaching formats would be an enhancement to the module (5 strongly agree, 3 agree, and 2 neutral; [Figure 3](#)). All students agreed that the module motivated them to get more involved with digital health (9 strongly agree and 1 agree) and that they felt well prepared for their future clinical work (3 strongly agree and 6 agree). For most students, this module was the first time they encountered digital health at medical school (8 strongly agree, 1 agree, and 1 strongly disagree).

There was a strong overall satisfaction with the organizational aspects of the module ([Multimedia Appendix 3](#)).

**Figure 2.** Students' opinions regarding digital health at medical schools.**Figure 3.** Students' experience with the digital health module.

### Lecturers' Evaluation

The majority of the lecturers stated that digital health should be part of the mandatory curriculum at medical schools (24 strongly agree, 6 agree, and 2 disagree) and that medical students should be taught digital health early in their career (25 strongly agree, 4 agree, 1 neutral, and 1 disagree; [Figure 4](#)). Most lecturers responded that a competency-centered teaching

approach is more suitable than a knowledge-based approach (17 strongly agree, 9 agree, 3 neutral, and 3 disagree). Regarding digital literacy among medical students, opinions were split. The majority of lecturers disagreed that digital health would be sufficiently represented at medical schools (14 strongly disagree, 8 disagree, 5 neutral, 1 agree, and 3 strongly agree). All lecturers

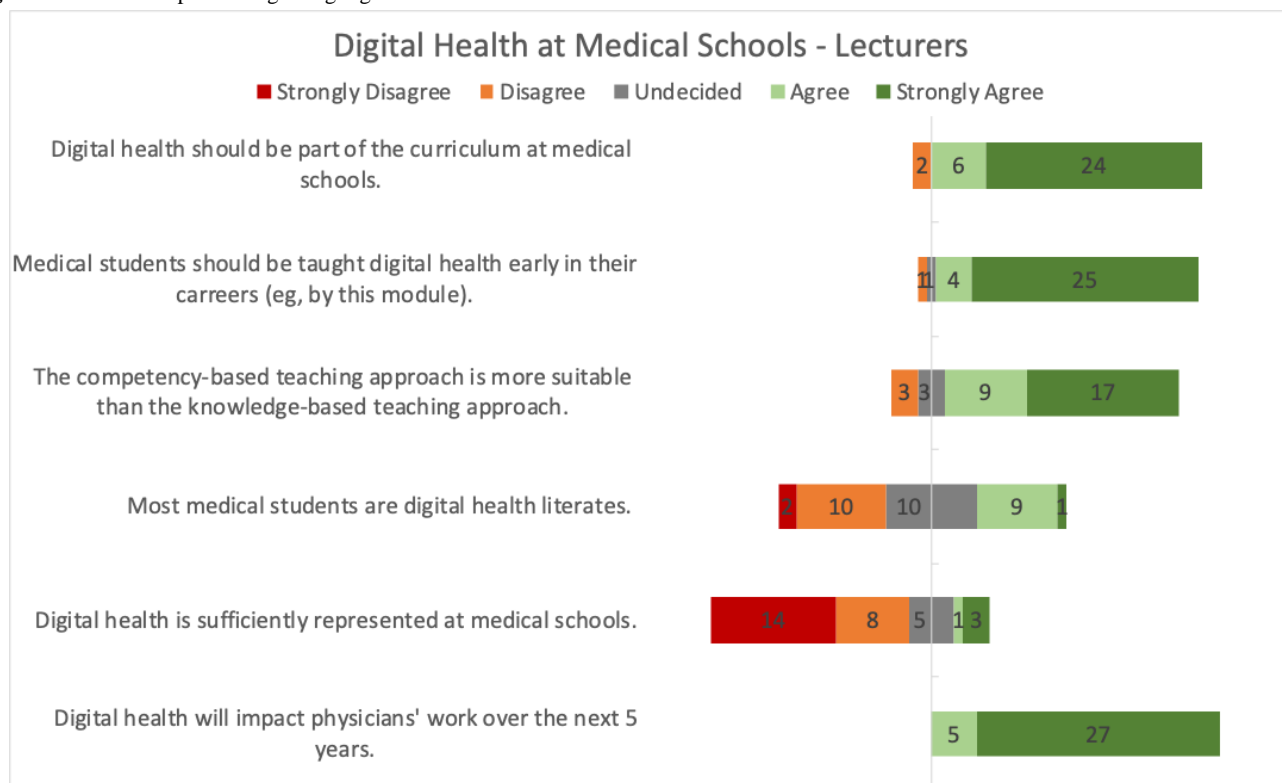
agreed that digital health will impact physicians' work over the next 5 years (27 strongly agree and 5 agree).

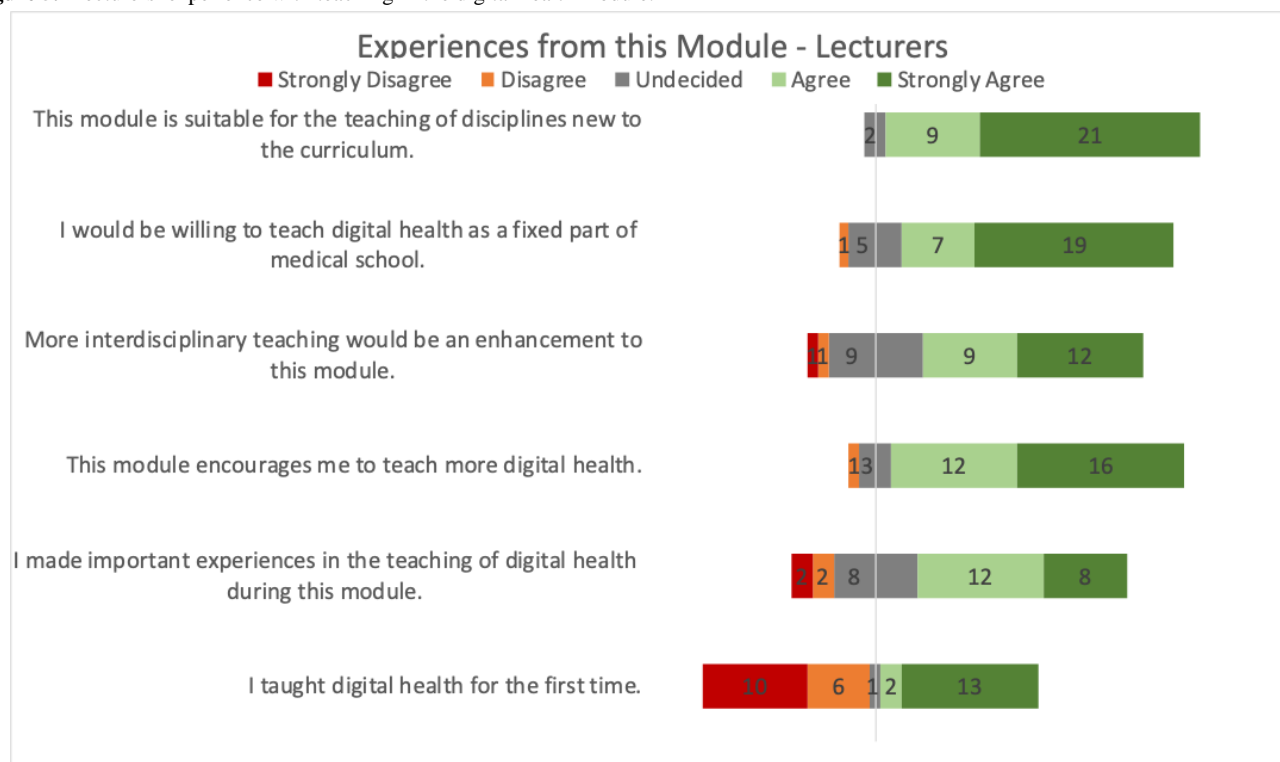
Regarding experiences from this module, lecturers agreed that this module over 3 weeks is suitable for teaching new disciplines (21 strongly agree, 9 agree, and 2 neutral; [Figure 5](#)). The majority of lecturers agreed that they would be willing to teach digital health as a fixed part of the medical curriculum (19 strongly agree, 7 agree, 5 neutral, and 1 disagree) and that more interdisciplinary teaching would enhance this module (12 strongly agree, 9 agree, 9 neutral, 1 disagree, and 1 strongly

disagree). Lecturers stated that this module animated them to teach more about digital health (16 strongly agree, 12 agree, 3 neutral, and 1 disagree). Regarding whether lecturers made important experiences during this module as well as whether lecturers taught digital health for the first time, opinions were split.

Overall, there was a high satisfaction rate for module organization, although in some individual cases, room for improvement was seen concerning the organizational communication ([Multimedia Appendix 3](#)).

**Figure 4.** Lecturers' opinions regarding digital health at medical schools.



**Figure 5.** Lecturers' experience with teaching in the digital health module.

## Discussion

### Principal Findings

In this mixed methods study, we provide valuable evidence on digital health teaching at medical schools. Our findings indicate that teaching digital health in medical schools not only comprises the knowledge transfer of technological advancements in health care. However, equally important is to shape the mindset of future physicians, teaching them to be open toward innovation, willing to develop and share novel insights through a transdisciplinary and patient-centered approach, and to be able to overcome fears of adopting digital health technology in clinical routine. Our results show that practical experiences, for example, with clinically implemented technology, bedside teaching, or by own creations as hackathons, are fundamental and should only be complemented with short impulse lectures. Small groups and peer teaching support the learning process.

The vast majority of lecturers and students stated that digital health is not sufficiently represented at medical schools and should be part of the curriculum. Both lecturers and students felt encouraged to get more involved with digital health after participating in the module.

### Need for Digital Health

The digital transformation of our health care system can increase the quality, accessibility, and affordability of health services [29]. By implementing digital technology for health care provision, we can apply health services more efficiently and individually. Within the next 20 years, it is expected that 90% of all jobs in the United Kingdom's National Health Service would require digital skills [30]. With an increasing flood of information for medical staff, there are already data reporting staff to be overwhelmed by alarms and notifications [31].

Intelligent systems may filter these and other digital information, making them better usable for patient care, which is more urgent than ever in some medical disciplines [32]. Furthermore, many processes could be simplified through the implementation of telemedicine in health care, thereby optimizing patient care [33,34].

Some sources name a lack of health care professionals' awareness and knowledge as a main barrier to a successful digital transformation of health systems [30,35]. The European Medical Students' Association recently identified a significant gap between students' willingness to act in a digitalized health care environment and the knowledge and skills acquired in their curriculum [23,24,36]. Medical staff and medical students can be regarded as ready and willing to use digital health technology in their clinical routine. However, adequate and regular training in digital devices as well as a structured curriculum in digital health are still lacking in most institutions [37,38], whereas some associations recommend up to 40 hours of education on biomedical and health informatics for medical and nursing students during their studies [39]. Our findings show that the need for digital health and its teaching is increasing, which is acknowledged by students, lecturers, and experts. Therefore, mandatory modules on digital health, medical informatics, or medical technologies should be included in the curricula of medical schools.

### Digital Health in Medical Schools

Elective classes and programs on digital health slowly find their way into medical curricula. This process is driven by individual pilot projects rather than by validated and coordinated guidelines or national regulations [16,38,40-48]. Pilot projects, adapted to the individual curricular conditions of medical faculties, can be the first step toward realizing a longitudinal interdisciplinary



approach to implement digital health in the overall curriculum [13,16,49-51].

In 2013, the German Association for Medical Informatics, Biometry and Epidemiology developed national competency-based learning objectives titled *Medical Informatics Learning Objectives for Undergraduate Medical Education*, which are consistent with the recommendations of the International Medical Informatics Association [50]. The objectives are in line with our findings implicating that medical graduates should have a broad knowledge of data literacy; be trained to assess ethical, legal, and social implications (ELSIs) and privacy aspects of digital health; and estimate the health economic consequences of novel technologies. Here, specialist expertise in digitalization beyond medicine could be offered in a local interdisciplinary and interprofessional network of universities. Even impressions from politics and business should be included to convey a comprehensive picture. Finally, addressing a peer-to-peer teaching approach, the motivation and competencies of many medical students as digital natives may have a very high potential to be included in medical curricular teaching offerings.

The national and international exchange of *best practice* examples and networking will be helpful in this respect. At the same time, however, urgent efforts should be made to define standards for the teaching of digital skills in medical studies at national levels [39].

### Undergraduate Medical Competencies in Digital Health

The future-proof physician must be able to apply digital health in the clinical routine. Our results emphasize that physicians should be open toward innovation and transdisciplinary cooperation throughout their careers and include the ability to recognize entrepreneurial potentials. Brunner et al [38] classified the digital health capabilities expected of medical graduates into 4 domains: (1) *digital technologies, systems, and policies*, covering digital literacy and ELSI; (2) *clinical practice and applications*, including the ability to integrate digital health into clinical routine; (3) *data analysis and knowledge creation*, including the ability to apply basic data analytics to unstructured digital data sets; and (4) *system and technology implementation*, suggesting that medical professionals should participate in the development and implementation of digital health. The latter aspect is also stressed by our results and a recent publication that demands physicians with dual competencies in clinical and data science expertise [49]. A medical graduate should be able to use digital health technology, interpret its results, and explain those to the patients.

Another main finding of our study is that teaching digital health should be about teaching a mindset that allows flexibility and suitability for an unknown and rapidly evolving future of our health care sector, rather than teaching the use of individual technologies or the right approach only for specific situations. This mindset includes openness to change, curiosity, and collaboration across functions, roles, professions, institutions, and hierarchical gradients. It also requires the willingness to actively anticipate and form the future of health care through reflection and innovation. Eventually, a strong awareness of

changing patients' and health care workers' needs and for new ELSIs should result from this mindset. It may be acquired by confrontation with a broad variety of different digital health aspects in terms of patient treatment, medical technologies, and the structure of health systems, as we did in our module.

Our results emphasize that passive knowledge transfer in the form of lectures is outdated, although it is the most used teaching method in medical schools [52]. A focus on students' engagement (eg, own creation in hackathons, peer teaching) and linking knowledge transfer to practical experiences (eg, bedside teaching with available digital technology) is desirable and effective. Students should be given the opportunity to deepen their knowledge and skills in special fields (eg, clinical data science, design thinking) in extracurricular activities or in the form of an elective module [49]. The exchange with other disciplines in workshops or hackathons can be the seed for future health innovations and entrepreneurial ventures. They are promising methods for introducing the inventor mindset to future physicians and offer the possibility of teaching medical students together with students of more information technology (IT)-focused and data-focused subjects. A close collaboration with IT faculties, for example, through an MD and computer scientist tandem, should be anticipated.

### Limitations

This mixed methods study provided novel insights into undergraduate medical competencies in light of digital transformation in health care and the development and deployment of a digital health module at a German medical school. However, a number of limitations apply.

First, the results are based on a limited number of participants. As a descriptive approach, quantification or generalization of the results is not possible but may still provide important impulses for further research and development in the field. Second, due to the selection bias of students and lecturers, the results of the questionnaire might only reflect the opinion of digital health advocates and are not representative. Third, the described undergraduate competencies in digital health are an additional source of information; therefore, there is no guarantee of completeness.

### Conclusions

Digital health competencies are key competencies for future-proof physicians and should therefore be included in the core curriculum at medical schools. For a first curricular implementation and to gain early local experiences, elective modules such as ours are suitable and can provide important information for existing didactic strengths in teaching this topic. The exchange of best teaching practices and institutional collaborations will further leverage the reach of this intent.

Digital health education should focus on building and sharpening a critical and experience-based *digital health mindset*, rather than passively transferring knowledge of technology specifics. Active teaching methods such as practice units, discussions with experts, and hackathons should be used to teach a broad variety of subjects. Interprofessional learning and teaching with nonmedical disciplines is recommended to add to the diversity

of perspectives and to prepare for increasingly interprofessional patient care.

## Acknowledgments

The authors express our gratitude to the digital health experts who participated in the interviews. Furthermore, the authors thank all students and lecturers for their feedback on this module and for sharing their opinions about digital health, especially the lecturers who also dedicated their time and effort for planning and performing new lessons within the context of the described module. The authors acknowledge support from Hacking Health Berlin, the German Research Foundation (DG), and the Open Access Publication Funds of Charité—Universitätsmedizin Berlin.

## Authors' Contributions

DB had the idea of developing and conducting a digital health module at the Charité—Universitätsmedizin Berlin. The study was conceived by AP, DB, DG, and FB. DG conducted data acquisition. Qualitative and quantitative data were analyzed by AP and DG, supported by KB and LM. AP and DG wrote the manuscript, supported by KB and LM. DB supervised all parts of the study. All authors critically reviewed and approved the manuscript.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Time table.

[DOCX File, 40 KB - [jmir\\_v22i10e22161\\_app1.docx](#)]

### Multimedia Appendix 2

Survey questionnaires.

[PDF File (Adobe PDF File), 345 KB - [jmir\\_v22i10e22161\\_app2.pdf](#)]

### Multimedia Appendix 3

Survey questionnaire results.

[PDF File (Adobe PDF File), 357 KB - [jmir\\_v22i10e22161\\_app3.pdf](#)]

## References

1. Wald HS, George P, Reis SP, Taylor JS. Electronic health record training in undergraduate medical education: bridging theory to practice with curricula for empowering patient- and relationship-centered care in the computerized setting. *Acad Med* 2014 Mar;89(3):380-386 [FREE Full text] [doi: [10.1097/ACM.000000000000131](#)] [Medline: [24448045](#)]
2. Bundesgesundheitsministerium. Die Elektronische Gesundheitskarte. URL: <https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/egk.html> [accessed 2019-07-15]
3. Tierney MJ, Pageler NM, Kahana M, Pantaleoni JL, Longhurst CA. Medical education in the electronic medical record (EMR) era: benefits, challenges, and future directions. *Acad Med* 2013 Jun;88(6):748-752. [doi: [10.1097/ACM.0b013e3182905ceb](#)] [Medline: [23619078](#)]
4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017 Feb 2;542(7639):115-118. [doi: [10.1038/nature21056](#)] [Medline: [28117445](#)]
5. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. *Yearb Med Inform* 2019 Aug;28(1):128-134 [FREE Full text] [doi: [10.1055/s-0039-1677903](#)] [Medline: [31022752](#)]
6. Waseh S, Dicker AP. Telemedicine training in undergraduate medical education: mixed-methods review. *JMIR Med Educ* 2019 Apr 8;5(1):e12515 [FREE Full text] [doi: [10.2196/12515](#)] [Medline: [30958269](#)]
7. Pathipati AS, Azad TD, Jethwani K. Telemedical education: training digital natives in telemedicine. *J Med Internet Res* 2016 Jul 12;18(7):e193 [FREE Full text] [doi: [10.2196/jmir.5534](#)] [Medline: [27405323](#)]
8. Krüger-Brand HE. Telemedizin: Vor dem Durchbruch. *Deutsches Ärzteblatt*. URL: <https://www.aerzteblatt.de/archiv/79318/Telemedizin-Vor-dem-Durchbruch> [accessed 2019-07-14]
9. Uslu Y, Altınbaş Y, Özercan T, van Giersbergen MY. The process of nurse adaptation to robotic surgery: a qualitative study. *Int J Med Robot* 2019 Aug;15(4):e1996. [doi: [10.1002/rcs.1996](#)] [Medline: [30884169](#)]
10. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. *Br Med J* 2020 Mar 12;368:m998. [doi: [10.1136/bmj.m998](#)] [Medline: [32165352](#)]
11. Fong DJ. Patient-centric Technology Improves Access, Efficiency, and Quality of Care | Clinical Drug Information Internet. Wolters Kluwer. URL: <https://www.wolterskluwer CDI.com/blog/patient-centric-technology/> [accessed 2019-07-19]

12. Digitalisierung im Gesundheitswesen: die 34-Milliarden-Euro-Chance für Deutschland. URL: <https://www.mckinsey.de/news/presse/2018-09-27-digitalisierung-im-gesundheitswesen> [accessed 2019-06-29]
13. Kuhn S, Kadioglu D, Deutsch K, Michl S. Data literacy in der medizin. *Onkologe* 2018 Feb 13;24(5):368-377. [doi: [10.1007/s00761-018-0344-9](https://doi.org/10.1007/s00761-018-0344-9)]
14. #SmartHealthSystems. Bertelsmann-Stiftung. 2018. URL: <https://www.bertelsmann-stiftung.de/en/publications/publication/did/smarthealthsystems-1> [accessed 2020-04-01]
15. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). *Implement Sci* 2016 Oct 26;11(1):146 [FREE Full text] [doi: [10.1186/s13012-016-0510-7](https://doi.org/10.1186/s13012-016-0510-7)] [Medline: [27782832](https://pubmed.ncbi.nlm.nih.gov/27782832/)]
16. Chandrashekar P. A digital health preclinical requirement for medical students. *Acad Med* 2019 Jun;94(6):749. [doi: [10.1097/ACM.0000000000002685](https://doi.org/10.1097/ACM.0000000000002685)] [Medline: [31136341](https://pubmed.ncbi.nlm.nih.gov/31136341/)]
17. Ridsdale C, Rothwell J, Smit M, Ali-Hassan H, Bliemel M, Irvine D, et al. Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report Internet. Semantic Scholar. 2015. URL: [http://www.mikesmit.com/wp-content/papercite-data/pdf/data\\_literacy.pdf](http://www.mikesmit.com/wp-content/papercite-data/pdf/data_literacy.pdf) [accessed 2020-10-19]
18. Loda T, Erschens R, Junne F, Stengel A, Zipfel S, Herrmann-Werner A. Undergraduate medical students' search for health information online: explanatory cross-sectional study. *JMIR Med Inform* 2020 Mar 2;8(3):e16279 [FREE Full text] [doi: [10.2196/16279](https://doi.org/10.2196/16279)] [Medline: [32130146](https://pubmed.ncbi.nlm.nih.gov/32130146/)]
19. Smith EE, Kahlke R, Judd T. Not just digital natives: integrating technologies in professional education contexts. *Australas J Educ Technol* 2020 Apr 27;1-14. [doi: [10.14742/ajet.5689](https://doi.org/10.14742/ajet.5689)]
20. Pontefract SK, Wilson K. Using electronic patient records: defining learning outcomes for undergraduate education. *BMC Med Educ* 2019 Jan 22;19(1):30 [FREE Full text] [doi: [10.1186/s12909-019-1466-5](https://doi.org/10.1186/s12909-019-1466-5)] [Medline: [30670000](https://pubmed.ncbi.nlm.nih.gov/30670000/)]
21. Blakemore LM, Meek SE, Marks LK. Equipping learners to evaluate online health care resources: longitudinal study of learning design strategies in a health care massive open online course. *J Med Internet Res* 2020 Feb 26;22(2):e15177 [FREE Full text] [doi: [10.2196/15177](https://doi.org/10.2196/15177)] [Medline: [32130120](https://pubmed.ncbi.nlm.nih.gov/32130120/)]
22. de Araújo Novaes M, Sá de Campos Filho A, Diniz PR. Improving education of medical students through telehealth. *Stud Health Technol Inform* 2019 Aug 21;264:1917-1918. [doi: [10.3233/SHTI190712](https://doi.org/10.3233/SHTI190712)] [Medline: [31438406](https://pubmed.ncbi.nlm.nih.gov/31438406/)]
23. Mosch L. EMSA Survey on EHealth Presentation Deans Meeting Rotterdam Internet. European Medical Students' Association. URL: <https://emsa-europe.eu/wp-content/uploads/2019/07/ESD-Deans-Meeting-Rotterdam-2019-1.pdf> [accessed 2019-07-15]
24. Machleid F, Kaczmarczyk R, Johann D, Balčiūnas J, Atienza-Carbonell B, von Maltzahn F, et al. Perceptions of Digital Health Education Among European Medical Students: Mixed Methods Survey. Toronto, Canada: European Medical Students' Association; 2019.
25. European Deans Meeting: Training Future-Proof Doctors for the Digital Society. European Medical Students' Association. 2019. URL: <https://www.erasmusmc.nl/en/patient-care/articles/deans-meeting-april-2019> [accessed 2019-07-19]
26. Dubois A, Gadde L. Systematic combining: an abductive approach to case research. *J Bus Res* 2002 Jul;55(7):553-560. [doi: [10.1016/s0148-2963\(00\)00195-8](https://doi.org/10.1016/s0148-2963(00)00195-8)]
27. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform* 2019 Jul;95:103208 [FREE Full text] [doi: [10.1016/j.jbi.2019.103208](https://doi.org/10.1016/j.jbi.2019.103208)] [Medline: [31078660](https://pubmed.ncbi.nlm.nih.gov/31078660/)]
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009 Apr;42(2):377-381 [FREE Full text] [doi: [10.1016/j.jbi.2008.08.010](https://doi.org/10.1016/j.jbi.2008.08.010)] [Medline: [18929686](https://pubmed.ncbi.nlm.nih.gov/18929686/)]
29. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. World Health Organization. URL: <https://www.who.int/publications-detail/who-guideline-recommendations-on-digital-interventions-for-health-system-strengthening> [accessed 2020-04-01]
30. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. *J Med Internet Res* 2017 Feb 16;19(2):e42 [FREE Full text] [doi: [10.2196/jmir.6900](https://doi.org/10.2196/jmir.6900)] [Medline: [28209558](https://pubmed.ncbi.nlm.nih.gov/28209558/)]
31. Bach TA, Berglund L, Turk E. Managing alarm systems for quality and safety in the hospital setting. *BMJ Open Qual* 2018;7(3):e000202. [doi: [10.1136/bmjopen-2017-000202](https://doi.org/10.1136/bmjopen-2017-000202)] [Medline: [30094341](https://pubmed.ncbi.nlm.nih.gov/30094341/)]
32. Komorowski M. Clinical management of sepsis can be improved by artificial intelligence: yes. *Intensive Care Med* 2020 Feb;46(2):375-377. [doi: [10.1007/s00134-019-05898-2](https://doi.org/10.1007/s00134-019-05898-2)] [Medline: [31834423](https://pubmed.ncbi.nlm.nih.gov/31834423/)]
33. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. *Lancet* 2018 Sep 22;392(10152):1047-1057. [doi: [10.1016/S0140-6736\(18\)31880-4](https://doi.org/10.1016/S0140-6736(18)31880-4)] [Medline: [30153985](https://pubmed.ncbi.nlm.nih.gov/30153985/)]
34. Koenig M. Telemedicine in the ICU. Cham, Switzerland: Springer International Publishing; 2019.
35. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. *J Am Med Inform Assoc* 2016 Jan;23(1):212-220. [doi: [10.1093/jamia/ocv052](https://doi.org/10.1093/jamia/ocv052)] [Medline: [26078410](https://pubmed.ncbi.nlm.nih.gov/26078410/)]

36. Machleid F, Kaczmarczyk R, Johann D, Balčiūnas J, Atienza-Carbonell B, von Maltzahn F, et al. Perceptions of digital health education among European medical students: mixed methods survey. *J Med Internet Res* 2020 Aug 14;22(8):e19827 [FREE Full text] [doi: [10.2196/19827](https://doi.org/10.2196/19827)] [Medline: [32667899](https://pubmed.ncbi.nlm.nih.gov/32667899/)]
37. Poncette A, Spies C, Mosch L, Schieler M, Weber-Carstens S, Krampe H, et al. Clinical requirements of future patient monitoring in the intensive care unit: qualitative study. *JMIR Med Inform* 2019 Apr 30;7(2):e13064 [FREE Full text] [doi: [10.2196/13064](https://doi.org/10.2196/13064)] [Medline: [31038467](https://pubmed.ncbi.nlm.nih.gov/31038467/)]
38. Brunner M, McGregor D, Keep M, Janssen A, Spallek H, Quinn D, et al. An ehealth capabilities framework for graduates and health professionals: mixed-methods study. *J Med Internet Res* 2018 May 15;20(5):e10229 [FREE Full text] [doi: [10.2196/10229](https://doi.org/10.2196/10229)] [Medline: [29764794](https://pubmed.ncbi.nlm.nih.gov/29764794/)]
39. Haag M, Igel C, Fischer MR, German Medical Education Society (GMA), Committee Digitization – Technology-Assisted LearningTeaching, Joint working group Technology-enhanced Teaching and Learning in Medicine (TeLL) of the German Association for Medical Informatics, BiometryEpidemiology (gmds)the German Informatics Society (GI). Digital teaching and digital medicine: a national initiative is needed. *GMS J Med Educ* 2018;35(3):Doc43 [FREE Full text] [doi: [10.3205/zma001189](https://doi.org/10.3205/zma001189)] [Medline: [30186953](https://pubmed.ncbi.nlm.nih.gov/30186953/)]
40. Timothy DA, Ravi P. Integrating digital health into the curriculum: considerations on the current landscape and future developments. Sage Publications. URL: <https://journals.sagepub.com/doi/full/10.1177/2382120519901275> [accessed 2020-04-01]
41. Gray K, Dattakumar A, Maeder A, Butler-Henderson K, Chenery H. Advancing Ehealth Education for the Clinical Health Professions: Final Report. Sydney, Australia: Bond University; 2014.
42. Digital Health Internet. UC San Diego Extension. URL: <https://extension.ucsd.edu/courses-and-programs/digital-health> [accessed 2020-04-10]
43. Academic Programs. Thomas Jefferson University. URL: <https://www.jefferson.edu/university/emerging-health-professions/centers/Digital-health/academic-programs.html> [accessed 2020-04-10]
44. Certificate in Digital Health Communication. Tufts Public Health. 2017. URL: <https://publichealth.tufts.edu/graduate-programs/professional-development/hcom-professional-development/digital-strategies-health> [accessed 2020-04-10]
45. Fernando J, Lindley J. Lessons learned from piloting mHealth informatics practice curriculum into a medical elective. *J Am Med Inform Assoc* 2018 Apr 1;25(4):380-384. [doi: [10.1093/jamia/ocx076](https://doi.org/10.1093/jamia/ocx076)] [Medline: [29024956](https://pubmed.ncbi.nlm.nih.gov/29024956/)]
46. Justus-Liebig-Univ Gießen. SPC Digitale Medizin, eHealth und Telemedizin. URL: <https://www.uni-giessen.de/fbz/fb11/studium/medizin/klinik/spc/spcehealth/index> [accessed 2019-06-29]
47. Mesko B, Györfy Z, Kollár J. Digital literacy in the medical curriculum: a course with social media tools and gamification. *JMIR Med Educ* 2015 Oct 1;1(2):e6 [FREE Full text] [doi: [10.2196/mededu.4411](https://doi.org/10.2196/mededu.4411)] [Medline: [27731856](https://pubmed.ncbi.nlm.nih.gov/27731856/)]
48. Prodekanat für Lehre - Modellstudiengang iMED. UKE. URL: <https://www.uke.de/studium-lehre/modellstudiengang-medizin-imed/index.html> [accessed 2019-06-29]
49. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? *NPJ Digit Med* 2020;3:86 [FREE Full text] [doi: [10.1038/s41746-020-0294-7](https://doi.org/10.1038/s41746-020-0294-7)] [Medline: [32577533](https://pubmed.ncbi.nlm.nih.gov/32577533/)]
50. Röhrig R, Stausberg J, Dugas M, GMDS project group Medical Informatics Education in Medicine". Development of national competency-based learning objectives 'medical informatics' for undergraduate medical education. *Methods Inf Med* 2013;52(3):184-188. [doi: [10.3414/ME13-04-0001](https://doi.org/10.3414/ME13-04-0001)] [Medline: [23615960](https://pubmed.ncbi.nlm.nih.gov/23615960/)]
51. Offergeld C, Neudert M, Emerich M, Schmidt T, Kuhn S, Giesler M. [Mediation of data literacy in curricular education in otorhinolaryngology: watch and wait or anticipatory obedience?]. *HNO* 2020 Apr;68(4):257-262. [doi: [10.1007/s00106-019-00745-8](https://doi.org/10.1007/s00106-019-00745-8)] [Medline: [31538215](https://pubmed.ncbi.nlm.nih.gov/31538215/)]
52. Dickman N, Schuster B. Active Education for Future Doctors. Cham, Switzerland: Springer International Publishing; 2020.

## Abbreviations

**AI:** artificial intelligence

**ELSI:** ethical, legal, and social implications

**IT:** information technology

*Edited by G Eysenbach; submitted 10.07.20; peer-reviewed by A Benis, L Genaro; comments to author 21.07.20; revised version received 05.08.20; accepted 06.08.20; published 29.10.20.*

*Please cite as:*

Poncette AS, Glauert DL, Mosch L, Braune K, Balzer F, Back DA

Undergraduate Medical Competencies in Digital Health and Curricular Module Development: Mixed Methods Study

*J Med Internet Res* 2020;22(10):e22161

URL: <http://www.jmir.org/2020/10/e22161/>

doi: [10.2196/22161](https://doi.org/10.2196/22161)

PMID: [33118935](https://pubmed.ncbi.nlm.nih.gov/33118935/)

©Akira-Sebastian Poncette, Daniel Leon Glauert, Lina Mosch, Katarina Braune, Felix Balzer, David Alexander Back. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Clinical Context–Aware Biomedical Text Summarization Using Deep Neural Network: Model Development and Validation

Muhammad Afzal<sup>1,2\*</sup>, PhD; Fakhare Alam<sup>2\*</sup>, MS; Khalid Mahmood Malik<sup>2</sup>, PhD; Ghaus M Malik<sup>3</sup>, MD

<sup>1</sup>Department of Software, Sejong University, Seoul, Republic of Korea

<sup>2</sup>Department of Computer Science & Engineering, School of Engineering and Computer Science, Oakland University, Rochester, MI, United States

<sup>3</sup>Department of Neurosurgery, Henry Ford Hospital, Detroit, MI, United States

\*these authors contributed equally

**Corresponding Author:**

Khalid Mahmood Malik, PhD

Department of Computer Science & Engineering

School of Engineering and Computer Science

Oakland University

Engineering Center 501

115 Library Drive

Rochester, MI, 48309

United States

Phone: 1 2483703542

Email: [mahmood@oakland.edu](mailto:mahmood@oakland.edu)

## Abstract

**Background:** Automatic text summarization (ATS) enables users to retrieve meaningful evidence from big data of biomedical repositories to make complex clinical decisions. Deep neural and recurrent networks outperform traditional machine-learning techniques in areas of natural language processing and computer vision; however, they are yet to be explored in the ATS domain, particularly for medical text summarization.

**Objective:** Traditional approaches in ATS for biomedical text suffer from fundamental issues such as an inability to capture clinical context, quality of evidence, and purpose-driven selection of passages for the summary. We aimed to circumvent these limitations through achieving precise, succinct, and coherent information extraction from credible published biomedical resources, and to construct a simplified summary containing the most informative content that can offer a review particular to clinical needs.

**Methods:** In our proposed approach, we introduce a novel framework, termed Biomed-Summarizer, that provides quality-aware Patient/Problem, Intervention, Comparison, and Outcome (PICO)-based intelligent and context-enabled summarization of biomedical text. Biomed-Summarizer integrates the prognosis quality recognition model with a clinical context–aware model to locate text sequences in the body of a biomedical article for use in the final summary. First, we developed a deep neural network binary classifier for quality recognition to acquire scientifically sound studies and filter out others. Second, we developed a bidirectional long-short term memory recurrent neural network as a clinical context–aware classifier, which was trained on semantically enriched features generated using a word-embedding tokenizer for identification of meaningful sentences representing PICO text sequences. Third, we calculated the similarity between query and PICO text sequences using Jaccard similarity with semantic enrichments, where the semantic enrichments are obtained using medical ontologies. Last, we generated a representative summary from the high-scoring PICO sequences aggregated by study type, publication credibility, and freshness score.

**Results:** Evaluation of the prognosis quality recognition model using a large dataset of biomedical literature related to intracranial aneurysm showed an accuracy of 95.41% (2562/2686) in terms of recognizing quality articles. The clinical context–aware multiclass classifier outperformed the traditional machine-learning algorithms, including support vector machine, gradient boosted tree, linear regression, K-nearest neighbor, and naïve Bayes, by achieving 93% (16127/17341) accuracy for classifying five categories: aim, population, intervention, results, and outcome. The semantic similarity algorithm achieved a significant Pearson correlation coefficient of 0.61 (0–1 scale) on a well-known BIOSSES dataset (with 100 pair sentences) after semantic enrichment, representing an improvement of 8.9% over baseline Jaccard similarity. Finally, we found a highly positive correlation among the evaluations performed by three domain experts concerning different metrics, suggesting that the automated summarization is satisfactory.

**Conclusions:** By employing the proposed method Biomed-Summarizer, high accuracy in ATS was achieved, enabling seamless curation of research evidence from the biomedical literature to use for clinical decision-making.

(*J Med Internet Res* 2020;22(10):e19810) doi:[10.2196/19810](https://doi.org/10.2196/19810)

## KEYWORDS

biomedical informatics; automatic text summarization; deep neural network; word embedding; semantic similarity; brain aneurysm

## Introduction

### Background

Automatic text summarization (ATS) is a leading topic in the field of information retrieval research, particularly in the medical and biomedical domains, which offers an efficient solution to access the ever-growing amount of scientific and clinical literature by summarizing the source documents while maintaining their most informative contents [1]. The large quantity of biomedical data now accessible to clinicians makes it challenging to locate the correct data rapidly; thus, automatic summarization can provide highlights particular to clinical needs [2]. Moreover, achieving precise, succinct, and coherent information extraction from credible published biomedical resources to construct a simplified summarization plays an imperative role in clinical decision-making, educating patients, and medical education. At the same time, automating this process will provide better opportunities for users to obtain the most critical points of required clinical knowledge without having to delve into an enormous amount of text, saving hours of searching. It is highly desirable to accurately identify scientifically sound published studies and summarize selected studies against a given type (eg, intervention and prognosis) of clinical query.

The primary requirements in carrying out this task of extracting noteworthy information are the efficiency of the process, contextualization, and precision of retrieved contents. Efficiency pertains to reducing human involvement and achieving the required materials in a timely manner; contextualization refers to the user objective relevant to a clinical task; and precision of contents applies to the correct identification of needed contents from trustworthy resources. Owing to the massive increase in the biomedical research literature, finding relevant, scientifically sound, and fit-to-context information has become more and more challenging for precise automatic text classification and summarization. The field of text classification and ATS is well-explored [3-7]; however, minimal work has been done in a clinically fit-to-context summary of biomedical text [8,9]. Medical text summarization poses a unique set of challenges compared to summarization in other domains [10].

The main limitations of existing medical text summarization are as follows: (a) inability to capture the clinical context while ranking the sentences for the summary; (b) lack of consideration in checking the quality of the documents before performing summarization; (c) inability of identifying context and implicit information present in the biomedical text, which cannot explicitly match with the user query; (d) lack of purpose-driven ranking and selection of passage for the final summary; and (e) nonuse of a large training set for training deep neural network models. To overcome these limitations, this paper introduces a

novel framework called Biomed-Summarizer, which provides a quality-aware Patient/Problem, Intervention, Comparison, and Outcome (PICO)-based intelligent and context-enabled summarization of contents of biomedical literature to satisfy the requirements listed above. Unlike the use of traditional features using a bag of words (BoW) approach, in which the text is tokenized into multiple words and each word is given a number representing the frequency of its use [3], more powerful word-embedding techniques have recently been introduced, such as word2vec [11,12] and GloVe [13], to combat the issues of data sparseness and context-overlooking by generating a vector space, typically of several hundred dimensions [14]. The word2vec approach creates vectors at the sentence or document level by employing the sentence2vec or doc2vec model [15]. The Biomed-Summarizer framework uses the Keras tokenizer [16] to generate a contextually rich set of features with a compact number of dimensions to enable semantics-based matching for precise extraction, summarization of contents, and avoiding sparseness. These features are then used as input to our bidirectional long-short term memory (Bi-LSTM) [17] classification model to identify the PICO sequence. PICO was initially proposed for formatting a clinical question. Subsequently, researchers have used the PICO structure for information retrieval and sentence identification in the text of biomedical documents [18-21]. In deciding which sequences to include in the final summary, we considered a comprehensive criterion that provides information on the quality of the study to which that sequence belongs, the relevance of the sequence to the user query, study type, credibility of the publishing venue, and freshness in term of the date of publication.

In summary, the main contributions of this paper are as follows. First, we introduce a novel framework, termed Biomed-Summarizer, for extractive multidocument ATS of biomedical documents. The summary construction is based on PICO elements identified in each record. Additionally, we employ contextual and quality parameters for the selection of a subset of a PICO sequence of sentences to include in the final summary.

Second, for quality recognition, a highly optimized multilayer feed-forward neural network model, multilayer perceptron (MLP), is presented to acquire significantly accurate results. This model offers binary classification to identify the soundness of a study based on data (title and abstract) and metadata (article type, authors, and publishing venue and date) features.

Third, for PICO elements classification, we propose a Bi-LSTM recurrent neural network (RNN) model trained on the vector representation of the text, which significantly boosts the performance compared to conventional machine-learning models such as support vector machine (SVM), logistic regression, decision tree, and naïve Bayes. Unlike previous studies that

focused on the detection of PICO elements one-by-one by employing a separate binary classifier for each PICO element, the proposed approach is a multiclassifier model, which classifies PICO sequences simultaneously from any given biomedical study.

Fourth, to accurately extract the PICO sentences to be included in the final summary, we present a novel method of calculating the similarity between the query and medical text using the Jaccard coefficient after semantically enriching the text using medical ontologies.

Finally, we offer a publicly available dataset [22] comprising thousands of abstracts related to intracranial aneurysm (also known as cerebral or brain aneurysm) curated from the biomedical literature for PICO-based classification. Additionally, another open-source dataset [22] is presented for the quality recognition model.

We aimed to achieve these contributions through the precise, succinct, and coherent information extraction from credible published biomedical resources, and to construct a simplified summary containing the most informative contents that provide a review particular to clinical needs.

## Related Works

### *ATS in the Biomedical Domain*

Summarization techniques are generally divided into two categories: abstractive and extractive [2,4]. Abstractive summarization methods examine the text and generate a new summarized text as a representative of the original text. In contrast, extractive summarization methods recognize the important part of the text, extract that component, and generate the summary verbatim. Extractive summarization approaches are classified into different categories such as statistical-, topic-, graph-, discourse-, and machine learning-based approaches [5]. Single and multidocument summarization are the two principal categories concerning the number of documents, whereas generic and query-specific are the two main types of summaries [5]; however, another possible criterion can be used to classify these studies into item set-based mining, and classification and ranking. The item set-based mining approach extracts domain concepts in the representative sentences to generate a graph-based summary [1,23,24]. The classification and ranking-based approach first detects key sentences and ranks them according to their importance in the text to produce a plain textual summary. For such a task, some researchers have used statistical features such as term frequency, sentence position, and similarity with the title [25,26], whereas other methods incorporated semantic information extracted from external linguistic resources [26,27].

### *Identification of PICO Elements in Text*

Minimal work has been done in the area of PICO-based retrieval of contents from the biomedical literature. The existing studies are categorized based on three aspects: individual PICO element identification [9,18,28], sentence classification [21,29], and question and answer with summarization [9,30]. In a proceeding, the authors presented a process of a PICO corpus at the individual element level and sentence level [31]. A hybrid

approach of combining machine learning and rule-based methods was proposed for the identification of PICO sentences and individual elements in successive order [20]. Another study on PICO sentence extraction was carried out with a supervised distance supervision approach that capitalizes on a small labeled dataset to mitigate noise in distantly derived annotations [32]. The authors developed a naïve Bayes-based classifier and reported that it is not sufficient to rely only on the first sentence of each section, particularly when high recall is required [28]. Boudin et al [33] used multiple supervised classification algorithms to detect PICO elements at the sentence level by training data on structured medical abstracts for each PICO element. The results showed that the detection accuracy was better for the Patient/Problem compared to the Intervention and Outcome elements.

### *Quality of Biomedical Studies*

Several promising approaches have been explored [34-41] to retrieve high-quality (ie, scientifically sound) studies from the PubMed database. Among these, some methods such as PubMed Clinical Queries rely on Boolean-based strategies using Medical Subject Heading (MeSH) terms and keywords [39,40]. Clinical Queries are often considered as a reference point for assessing the effectiveness of approaches intended to retrieve scientifically sound studies from PubMed. Some methods have also used supervised machine-learning algorithms such as SVM, naïve Bayes, and decision tree to identify scientifically rigorous studies from PubMed [34,37,38,42]. These approaches mainly depend on bibliometric features, semantic features, and MeSH terms. The limitation of approaches that use MeSH terms is the availability, as MeSH terms are added to PubMed citations with an average time gap of 20 to 252 days after an article is published [40,43]. Very recently, a deep-learning approach was used for the task of detecting quality articles in PubMed [40,41]. Deep learning-based approaches have proven to improve accuracy over the existing approaches of machine-learning models, PubMed Clinical Queries search, and McMaster text word search in terms of precision. These approaches were trained on a treatment-related dataset.

### *Sentence Scoring and Ranking for Summarization*

Various methods have been proposed to choose what text should be included in the final summary [1,4-8,23,24,44]. The most common method is the frequency-based approach, in which a sentence with a word holding the highest frequency is given more importance [45]. Some studies have found similarities in a sentence with the title of the document [5]. If a sentence encompasses words in the title, the sentence is assigned a score value of 1; otherwise, a score value of 0 is assigned. Another approach commonly followed for the selection of sentences to include in the summary is the cue words technique, in which the cue words are provided by the user [4,6,44]. A score value of 1 is assigned to the sentence if it contains these cue words; otherwise, a score value of 0 is assigned. Sentence position and length are also considered for sentence inclusion in the summary. These techniques are used to score a sentence by linearly combining the individual scores of each method [44]. Based on the scoring, a certain number of sentences with high ranks are picked up to include in the final summary. Recently,

deep-learning approaches have been widely used in text summarization [3,46-49]. A neural network-based general framework was proposed for single-document summarization using a hierarchical document encoder and attention-based extractor [48]. For the sentence ranking task of multidocument summarization, a ranking framework was presented using a recursive neural network (R2N2) [49]. A query-focus summarization system called AttSum [46] was proposed that determines ranking based on *both* saliency and relevance in contrast to previous approaches of learning to rank based on *either* relevance or saliency.

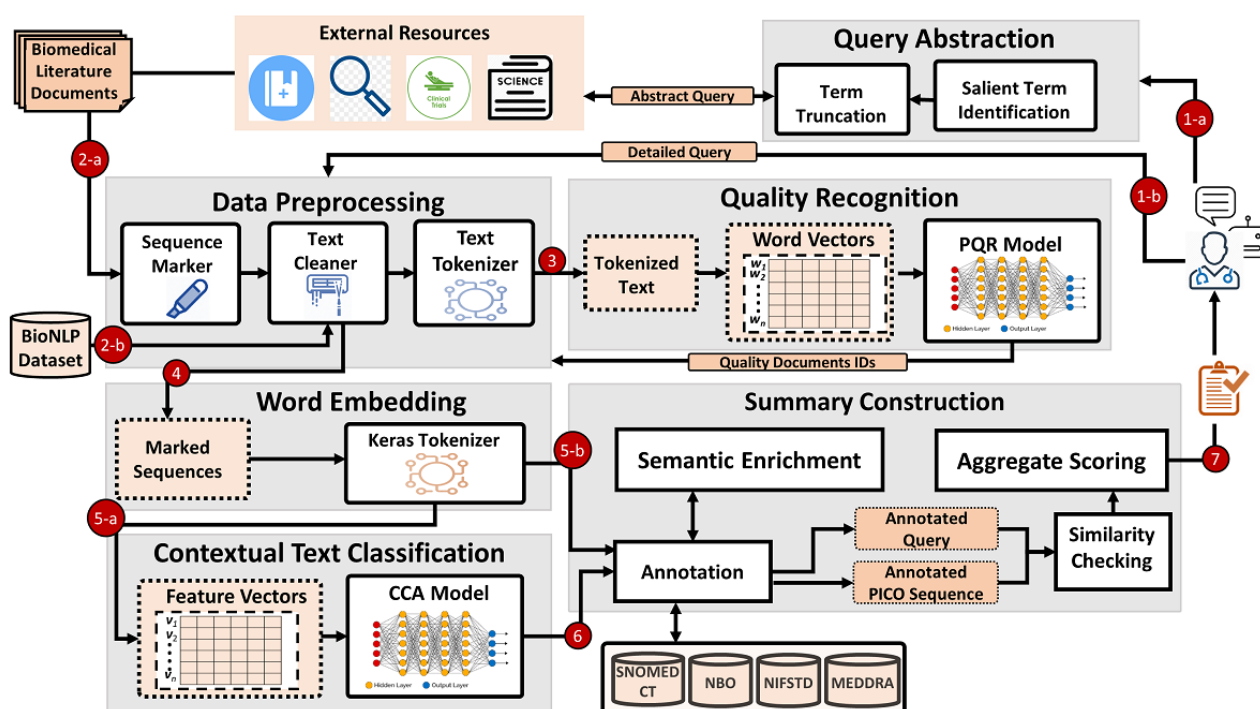
The prominent issues with existing approaches are as follows: (a) inability to capture the clinical context while ranking the sentences for creating summarization, (b) omitting the check on the quality of the documents used in the summary, (c) inability of identifying implicit information located in the body text of a study that is not explicitly matched with the user query, (d) lack of purpose-driven ranking and selection of passage for the final summary, and (e) unavailability of a large training set for training deep neural network models. Biomed-Summarizer circumvents all of these limitations.

## Methods

### Overall Design

Figure 1 shows the architecture of the proposed Biomed-Summarizer framework highlighting four main modules: data preprocessing, quality recognition, contextual text classification, and text summarization. The main tasks in data preprocessing are sequence marker, cleaning, tokenization, vector generation, and feature creation. The quality recognition module identifies the scientific soundness of a study using a feed-forward deep-learning binary classifier that is trained on biomedical data and metadata features. The contextual text classification classifies the PICO sequence of sentences in the selected quality documents using the Bi-LSTM model, an RNN trained on a large set of data acquired from BioNLP [21] and domain-specific Medline abstracts. Lastly, the summarization module generates a summary from the classified PICO sequences of multiple documents using a sequence scoring method consisting of semantic similarity (between a query and a text sequence extracted from a biomedical study), the publishing venue's credibility, and year of publication.

**Figure 1.** Proposed Biomed-Summarizer architecture with four major components: data preprocessing, quality recognition, context identification, and summary construction. PQR: prognosis quality recognition; CCA: clinical context-aware; PICO: Population/Problem, Intervention, Comparison, Outcome.



### Data Preprocessing

Biomed-Summarizer acquires data from two input sources: BioNLP dataset and PubMed. The BioNLP dataset is publicly available, whereas we retrieved the domain-specific abstracts from PubMed. The process used to parse raw abstracts retrieved from the PubMed and BioNLP is described below.

#### Sequence Marker

The task of sequence marker is to parse each raw abstract in the dataset to retrieve the headings based on the keywords listed in

the dictionary, as shown in Textbox 1. If a keyword is matched in the abstract, we extract the text (sequence of sentences) under that heading and mark it with the corresponding label. For instance, if a heading “objective” is found in an abstract, label “A” is assigned to the text. The dictionary is based on a previous study [21] with extension of a few more keywords (eg, Patient 1 and Patient 2). This process is repeated for all of the abstracts in the dataset of documents retrieved from PubMed.



**Textbox 1.** Master dictionary representing keywords that appear in headings in the structured abstracts of biomedical studies.

*dict = {'A': ['objective', 'background', 'background and objectives', 'context', 'background and purpose', 'purpose', 'importance', 'introduction', 'aim', 'rationale', 'goal', 'context', 'hypothesis'], 'P': ['population', 'participant', 'sample', 'subject', 'patient', 'patient 1', 'patient 2'], 'I': ['intervention', 'diagnosis'], 'O': ['outcome', 'measure', 'variable', 'assessment'], 'M': ['method', 'setting', 'design', 'material', 'procedure', 'process', 'methodology'], 'R': ['result', 'finding'], 'C': ['conclusion', 'implication', 'discussion', 'interpretation table']}*

We employed the following steps to prepare data for the clinical context-aware (CCA) classification model.

### Text Cleaning

The process of text cleaning removes special characters, punctuation, stop words, and URLs present in the text. For removing stop words, we used the Python NLTK library [50].

### Tokenization

This process splits the sentences retrieved after the text cleaning process into individual tokens. We used the Keras tokenizer [16] to create a list of tokens for each biomedical paragraph.

### Vector Generation

We considered a maximum of 50,000 words for each paragraph and generated a 250-dimension vector representation using Keras text [16] to sequence functionality. We made sure that each vector length is the same.

As a result of the above steps, we obtained a vector representation of each paragraph, which is then used for training and testing of the CCA classification model. An example of text sequences and corresponding class labels is shown in Table 1. It is important to mention that we do not require a process of sequence marking on BioNLP data because these sentences are premarked.

**Table 1.** Example sequence of sentences for an assigned category of Aim, Population, Methods, Interventions, Results, Conclusion, and Outcomes.

Sequence	Category
The aim of the present study was to evaluate whether the Anterior communicating artery (A com) aneurysms behave differently from the aneurysms located elsewhere with respect to size being a rupture risk. To this end, we examined the clinical data of ruptured A com aneurysms and analyzed other morphological parameters, including size parameter, providing adequate data for predicting rupture risk of the A com aneurysms.	Aim (A)
Between January 2010 and December 2015, a total of 130 consecutive patients at our institution with the A com aneurysms- 86 ruptured and 44 unruptured-were included in this study. The ruptured group included 43 females (50%) and 43 males (50%) with the mean age of 56 years (range, 34-83 years). The unruptured group included 23 females (52%) and 21 males (48%) with the mean age of 62 years (range, 28-80 years). All patients underwent either digital subtraction angiography or 3-dimensional computed tomography angiography. The exclusion criteria for this study were the patients with fusiform, traumatic, or mycotic aneurysm. There were preexisting known risk factors, such as hypertension in 73 patients, who required antihypertensive medication; other risk factors included diabetes mellitus (9 patients), coronary heart disease (9 patients), previous cerebral stroke (18 patients), and end-stage renal disease (3 patients) in the ruptured group. In the unruptured group, 38 patients had hypertension, 4 had diabetes mellitus, 5 had coronary heart disease, 10 had a previous cerebral stroke, and 2 had end-stage renal disease.	Population (P)
Four intracranial aneurysms cases were selected for this study. Using CT angiography images, the rapid prototyping process was completed using a polyjet technology machine. The size and morphology of the prototypes were compared to brain digital subtraction arteriography of the same patients.	Methods (M)
After patients underwent dural puncture in the sitting position at L3-L4 or L4-L5, 0.5% hyperbaric bupivacaine was injected over two minutes: group S7.5 received 1.5 mL, Group S5 received 1.0 mL, and group S4 received 0.8 mL. Interventions after sitting for 10 minutes, patients were positioned for surgery.	Intervention (I)
The ruptured group consisted of 9 very small (<2 mm), 38 small (2-4 mm), 32 medium (4-10 mm), and 7 large (>10 mm) aneurysms; the unruptured group consisted of 2 very small, 16 small, 25 medium, and one large aneurysms. There were 73 ruptured aneurysms with small necks and 13 with wide necks (neck size >4 mm), and 34 unruptured aneurysms with small necks and 10 with wide necks.	Results (R)
The method which we develop here could become surgical planning for intracranial aneurysm treatment in the clinical workflow.	Conclusion (C)
The prevailing view is that larger aneurysms have a greater risk of rupture. Predicting the risk of aneurysmal rupture, especially for aneurysms with a relatively small diameter, continues to be a topic of discourse. In fact, the majority of previous large-scale studies have used the maximum size of aneurysms as a predictor of aneurysm rupture.	Outcome (O)

### Quality Recognition Model

The proposed model, as shown in Figure 2, comprises multilayer feed-forwarded neural networks as a so-called MLP. The MLP is further optimized with an ensemble method using adaptive boosting (AdaBoost). The final AdaBoost-MLP, termed the prognosis quality recognition (PQR) model, was trained on a

dataset acquired automatically through PubMed searches based on the following two criteria: selecting the “Clinical Query prognosis” filter and choosing scope as “narrow.” To build and evaluate the model, we performed the steps involved in the preparation of a dataset for training the deep-learning models, followed by training and tuning the deep-learning models. Additionally, we compared the performance of the proposed

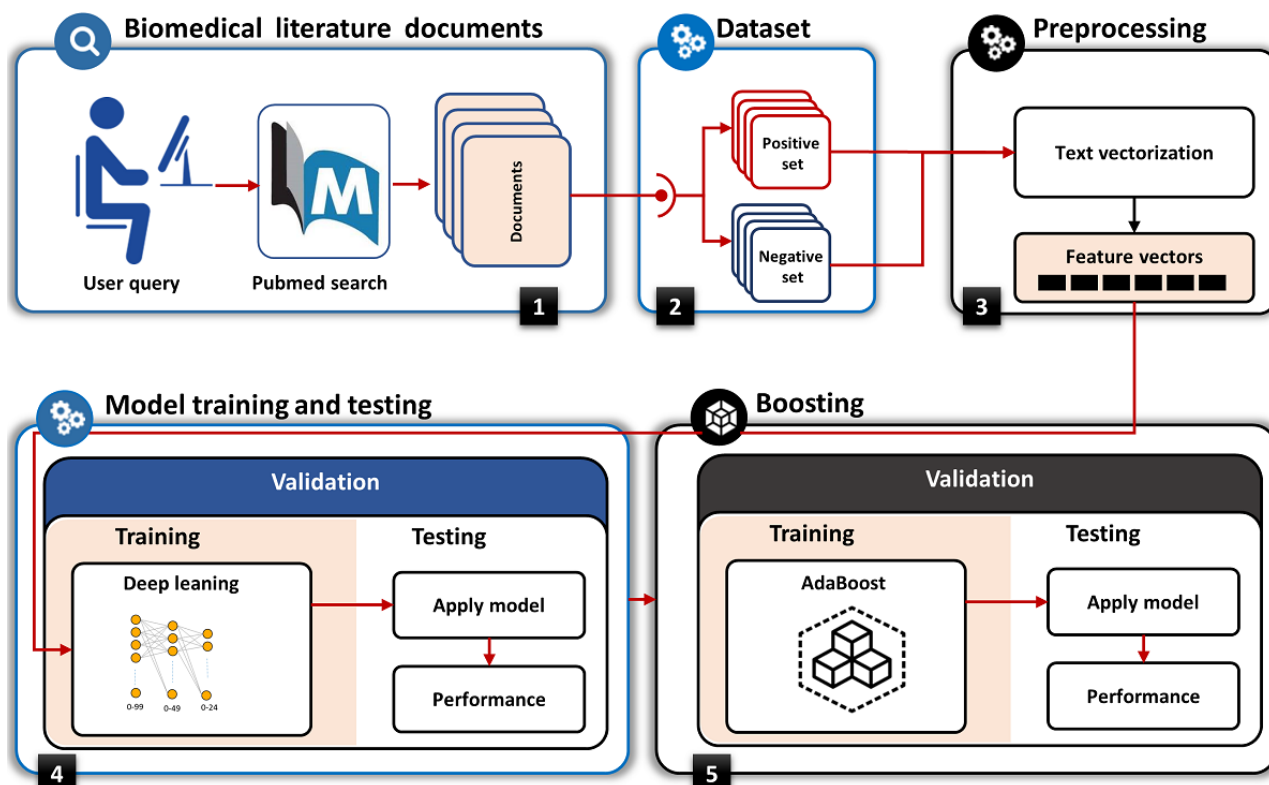


model with those of shallow machine-learning models in terms of precision and recall.

For training and testing the PQR model, we collected a dataset consisting of a total of 2686 Medline abstracts, 697 of which

were considered to be positive studies (ie, scientifically sound). To retrieve positive studies, we used the interface of Clinical Queries [33]. To retrieve negative studies (ie, not scientifically sound), we used the interface of PubMed and retrieved studies that were not included in the positive set.

**Figure 2.** Process steps of proposed prognosis quality recognition (PQR) model training and testing.



### Training and Tuning the Deep-Learning Model

The PQR model is trained using five textual features, including two data features (title and abstract) and three metadata features (article type, publishing journal, and authors). The data features were preprocessed through applying the steps described above. The model consists of 5 layers with 3 hidden layers of size 100, 50, and 25, respectively. The input layers take the BoW vectors generated from data features. The “Maxout” activation function was used with 10 epochs. The data were split into a 70:30 ratio for training and testing.

### Comparison with Shallow Machine-Learning Models

To identify the best performer on our dataset, we compared a set of machine-learning algorithms, including SVM, naïve Bayes, k-nearest neighbor, and decision tree. These algorithms were tested in other studies [35,39] for addressing similar problems.

### CCA Classification Model

The CCA classification model aims to develop a multiclass classifier that can accurately classify text from selected quality documents (given by the quality model) into one of the following 7 classes: Aim, Methods, Population, Intervention, Results, Outcome, and Conclusion. We then merged Methods and Population into a single category (P) and Outcome and Conclusion into a single category (O) because of their strong

correlation found in the text. Although our focus was to utilize PICO classes in the summarization task, we retained the other categories to enable additional non-PICO clinical applications such as summarization for medical education and deriving rules from clinical trials for developing the knowledge base of a clinical decision support system.

PICO is a well-known terminology in evidence-based medicine. Our proposed CCA classifier incorporates the Patient/Problem, Intervention, and Outcome PICO components in addition to two more classes, Aim and Results. PICO detection is a sequential sentence classification task rather than a single sentence problem; therefore, the sequence of sentences is jointly predicted. In this way, more complete context is captured from multiple sentences, which then improves the classification accuracy of predicting the current sentence. The steps followed to build and evaluate the model were: (a) preparation of a dataset for training the deep-learning model, (b) training and tuning the deep-learning model, and (c) comparison with traditional machine-learning models in terms of precision and recall.

### Model Building

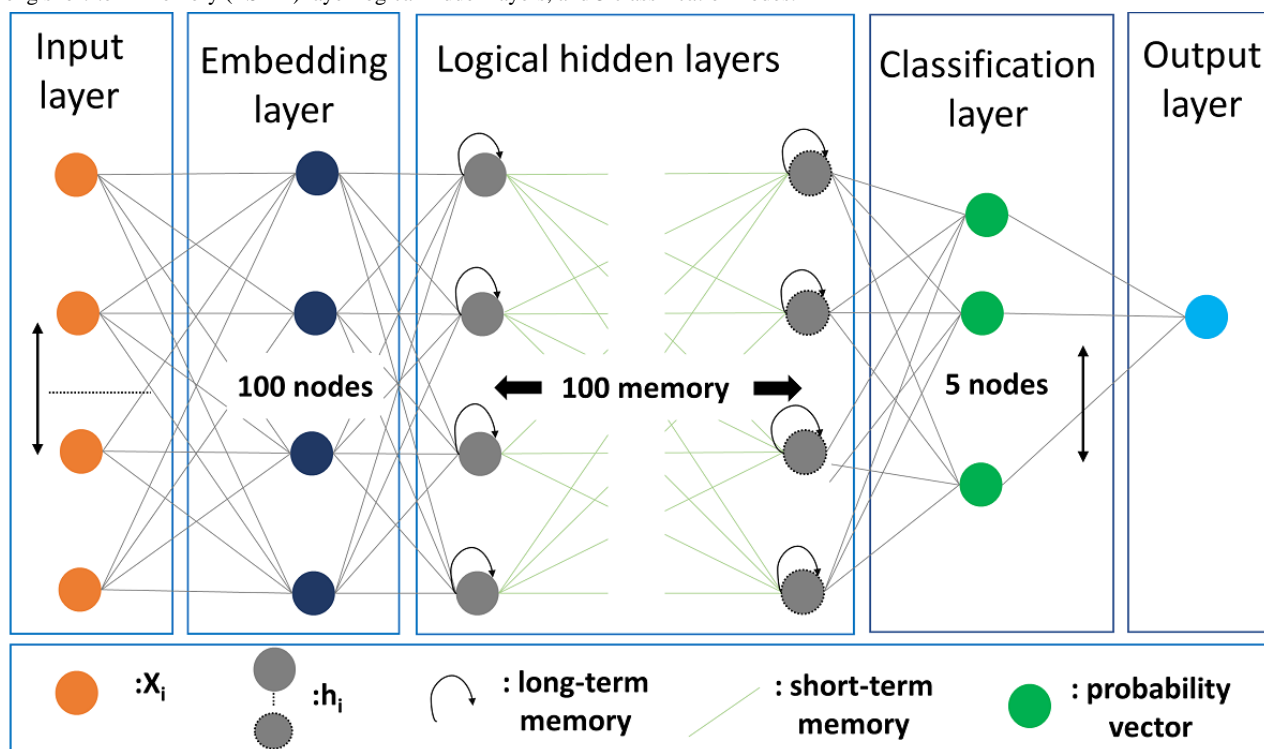
The neural network model is heavily used in text processing due to the ability to process arbitrary-length sequences. RNNs are immensely popular in multiclass text classification. To build and evaluate the classification model, we employed the Bi-LSTM model as a type of RNN. This model preserves the

long-term dependency of the text and is one of the most popular models in text classification.

As shown in Figure 3, the input layer consists of 250 dimensions showing the numeric features created using the Keras tokenizer, embedding layer, LSTM logical hidden layers, classification layer, and output layer. In brief, the trained CCA model has the following features: (1) an initial training dataset comprising 173,401 records, 90% (n=156,060) of which were used for training and 10% (n=17,341) of which were used for testing; (2) the first hidden layer is the embedding layer that uses

100-dimension vectors to represent each paragraph; (3) the next layer is the LSTM layer with 100 memory units, and the recurrent dropout rate was set to 0.2; (4) the next layer is the dense layer with 5 nodes, and we used the SoftMax activation function for multiclassification, and categorical\_crossentropy as a loss function; (5) the training dataset is further divided into two parts, 10% of which was used to validate the category and minimize the loss function, resulting in 140,454 records used for training and 15,606 used for validation; (6) the model was then trained across 5 epochs with a minibatch size of 64.

**Figure 3.** Clinical context-aware (CCA) classifier trained on 250-dimension feature vectors, 100 nodes at the embedding layer, 100 memory units of the long short-term memory (LSTM) layer logical hidden layers, and 5 classification nodes.



## ATS

For the ATS task, we developed a sentence-scoring mechanism based on a multifeatured matrix, consisting of 4 features (relevance, study type, venue credibility, and freshness), where each feature was learned with a specific method as explained in the following sections.

First, an individual score value is assigned to each feature, which is then aggregated in the final column as a final score using the formula in Equation 1:

$$\text{Aggregate Score} = \beta R_{score} + \gamma (ST_{rank} + VC_{score} + f_{score}) \quad (1)$$

where  $R_{score}$  is the relevance score,  $ST_{rank}$  is the study type rank,  $VC_{score}$  is the publishing venue credibility score,  $f_{score}$  is the freshness score, and  $\beta=70$  and  $\gamma=10$  are the scaling constants to keep the aggregate score in the range of 0 to 100.

## Relevance Score

For the relevance score, we developed a semantic similarity algorithm, Jaccard similarity with semantic enrichments (JS<sup>2</sup>E) as a 6-step algorithm that uses the BioPortal application

programming interface to access biomedical ontologies (SNOMED CT, MedDRA, NBO, NIFSTD) for obtaining semantic enrichment and the Jaccard similarity metric to find similarities between two texts.

In step 1, an individual sequence of text is obtained from the query, which is preprocessed to remove the special characters, separate characters, and different formatting tags. In step 2, the preprocessed text is annotated using BioPortal. In step 3, each token is enriched using “definition,” “synonyms,” and “prefLabel” relations from selected ontologies. In step 4, the annotations of text are retrieved using the “definition” relationship, along with the preprocessing (step 1) and annotation (step 2) procedures. In step 5, the annotated tokens received in step 2 are combined with the “synonyms” and “preflabel” obtained in step 3 and the annotated tokens received in step 4, and the data structure of metatokens is constructed. Finally, in step 6, the Jaccard similarity between the metatokens of the text sequence and the query is calculated using Equation 2:

$$R_{score} = S_m \cap Q_m / S_m \cup Q_m \quad (2)$$

where  $S_m$  are the metaset tokens of the text sequence in a document and  $Q_m$  are the metaset tokens of the query.

Study Type

The study type plays a vital role in proving a study’s usefulness for a user concerning a clinical task. For instance, if a surgeon wants to look for advances in successful surgical treatments, randomized control trials or a meta-analysis of randomized controlled trials will be the priority. The priority will change if the user is interested in prognosis-related studies. Study types can be used to grade evidence concerning quality and other elements [34,51,52]. To find the priority of different study types in the category of prognosis, we conducted a questionnaire-based survey among domain experts who were asked to assign a score value for each study type. An example of a filled-in questionnaire by a physician is provided in Multimedia Appendix 1.

The final rank value is then learned from the average value of rank values assigned by the domain experts, as shown in Equation 3.



where  $V_i$  represents the rank values assigned to a study type by each domain expert,  $n$  is the total number of domain experts that participated in the questionnaire, and  $\alpha=0.1$  is a scaling constant to keep the final score value in the range of 0 to 1.

Table 2. Assigned weights for research study year of publication.

Year of publication	Rank
Previous 1-5 years	1
Previous 6-10 years	2
Previous 11-15 years	3
Other	4

The final score was calculated according to Equation 5:

$F_{score}=\alpha(R_i\rightarrow R_f)$  (5),

where  $R_i$  is the initial value assigned to each year obtained from Table 2,  $R_f$  is the mapped score obtained through mappings {1→10, 2→9, 3→8, and 4→7}, and  $\alpha=0.1$  is a scaling constant to keep the final score value in the range of 0 to 1.

Text Selection for Summary

The aggregate score calculated using Equation 1 provides the final rank values for a text sequence. Two types of summary structures can be generated: (1) a PICO-based summary, in which we select the top  $k$  text sequences in each part of PICO out of a total  $n$  number of sentences to be included in the final summary as  $|n/k|$  sentences from the Patient/Population component,  $|n/k|$  sentences from the Intervention component,

Venue Credibility

The credibility of a study publishing venue, including journals, proceedings, and books, is also an important parameter; however, it is more of a subjective matter. Therefore, it is necessary to consult with the stakeholders of the service. For this study, we sought to obtain a list of valuable publishing venues. We involved resident doctors to rank various venues concerning their credibility. An example of a filled questionnaire by a physician is provided in Multimedia Appendix 1. The final credibility score was determined from the average value of rank values assigned by the domain experts, as shown in Equation 4:



where  $S_i$  is the initial rank assigned by the domain experts,  $S_f$  is the mapped score obtained through mappings {1→10, 2→9, 3→8, 4→7, 5→6, and 6→5}, and  $\alpha=0.1$  is a scaling constant to keep the final score value in the range of 0 to 1. We applied the majority vote method before using the mapping function.

Freshness

Freshness represents the date of publication of a study, which is useful to consider to keep up with the advancement in a domain. We included this attribute in summarization with a higher rank assigned to more recent studies, following the less current studies, as described in Table 2.

and  $|n/k|$  from the Outcome component; and (2) a non-PICO-based summary, where we select the top  $k$  text sequences without considering their classification.

Summary Presentation

Our proposed model of automatic summarization can generate a summary for a single document or multiple documents simultaneously. For summaries of a single document, the selected set of representative PICO sequences is arranged in a template of Population/Problem, Intervention, and Outcome, as shown in Textbox 2. The templates are displayed according to the number of documents retained after a quality check. For instance, if we have a set of 5 studies retrieved against a user query, our algorithm will generate 5 individual summaries presented in the order of the most recent document summary followed by the others.

Textbox 2. Example summary of a biomedical document represented with the Patient/Problem (P), Intervention (I), and Outcome (O) sequence.

P: One hundred and fifty patients ...duration at a frequency of at least once per week
I: After patients underwent dural puncture ... patients were positioned for surgery
O: Number of follow-up appointments attended ... occurrence of secondary ocular hypertension

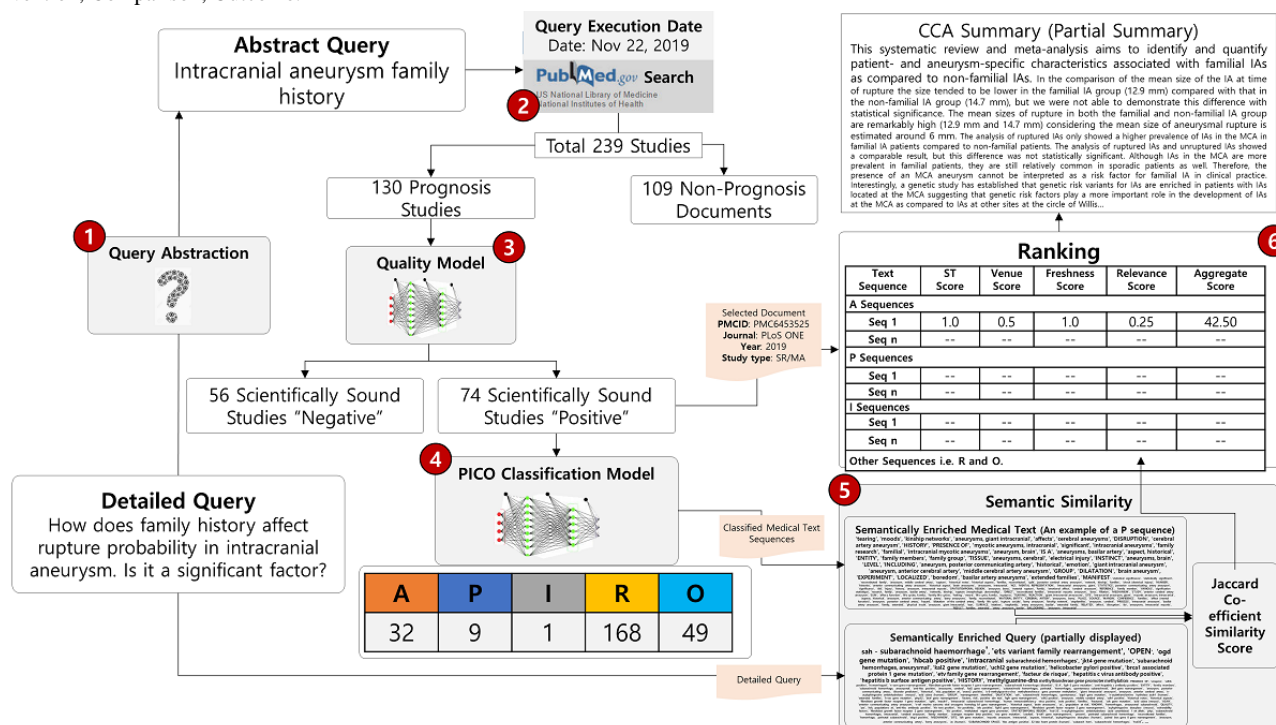
In the case of a multidocument summary, our algorithm first identifies PICO sequences in all documents, finds the score for each sequence, selects the highest-scoring sequence in each category, and concatenates all scored sequences to obtain a combined summary of each type. In this case, the sequences included in the summary may or may not belong to a single document; however, each sequence is linked to its corresponding documents for audit, transparency, and education.

### Example Case: Step-by-Step Scenario of Summary Generation

To clarify the steps of the proposed Biomed-Summarizer framework, we take the following example query: “How does family history affect rupture probability in intracranial aneurysms; is it a significant factor?” (Abstract Query-Intracranial aneurysm family history). As shown in Figure 4, the user query was first abstracted from a detailed user

query to increase the recall of retrieving studies. Second, the query was run on the PubMed search service, which returned a total of 239 studies, 130 of which were prognosis studies. Third, these studies were run through a quality model, “PQM,” which identified 74 studies as scientifically sound and the rest were filtered out. Fourth, the set of 74 studies was given to the PICO classification model in the form of text sequences, which were classified into five classes: Aim (32), Patients (9), Intervention (1), Results (168), and Outcome (49). Fifth, the PICO text sequences and the detailed query were passed through semantic similarity, in which the texts were first enriched semantically using medical ontologies, and the similarity score was calculated using the Jaccard coefficient. Sixth, the documents were ranked according to the accumulated score of four parameters: relevance score of the query and text, study type, venue’s credibility score, and freshness score. Finally, the required summary is created and presented to the user.

**Figure 4.** Step-by-step scenario of query execution, retrieval of documents, quality checking, clinical context-aware (CCA) classification, semantic similarity, ranking, and summary creation. A: Aim; P: Population/Patients/Problem; I: Intervention; R: Results; O: Outcome; PICO: Patient/Problem, Intervention, Comparison, Outcome.



## Results

### Dataset Preparation

To evaluate the performance of the proposed framework, we prepared two different datasets to test the proposed Biomed

Summarizer framework. One dataset was used for the development and testing of the PQR model (D1), whereas the other was used for the development and testing of the CCA model (D2). The preparation protocols of both datasets are outlined in Table 3. These datasets are available to the research community via a GitHub link [22].



**Table 3.** Dataset preparation protocols.

Preparation Protocol	PQR <sup>a</sup> dataset (D1)	CCA <sup>b</sup> dataset (D2)
Description	This dataset was created for the quality assessment of biomedical studies related to the prognosis of brain aneurysm.	This dataset was curated for the use of PICO <sup>c</sup> sequence classification. The final dataset was specific to the prognosis of brain aneurysm.
Purpose	To select only published documents that are scientifically rigorous for final summarization.	To identify a sentence or a group of sentences for discovering the clinical context in terms of population, intervention, and outcomes.
Methods	The manual preparation of the dataset is a cumbersome job, and thus AI <sup>d</sup> models were used. For development of an AI model, a massive set of annotated documents is needed. Annotation is a tedious job; therefore, PubMed Clinical Queries (narrow) as a surrogate were used to obtain scientifically rigorous studies.	N/A <sup>e</sup>
Data sources	PubMed Database (for positive studies, the “Narrow[filter]” parameter was enabled).	First, we collected a publicly available dataset, BioNLP 2018 [21], which was classified based on the PICO sequence in addition to “Method” and “Results” elements. To increase the dataset size, we added more sentences related to brain aneurysm created from Medline abstracts retrieved using the NCBI <sup>f</sup> PubMed service Biopython Entrez library [53].
Query	The term “(Prognosis/Narrow[filter]) AND (intracranial aneurysm)” was used as a query string.	The term “Intracranial aneurysm” (along with its synonyms “cerebral aneurysm” and “brain aneurysm”) were used as a query string.
Size	2686 documents, including 697 positive (ie, scientifically rigorous) records	A total of 173,000 PICO sequences (131,000 BioNLP+42,000 Brain Aneurysm) were included in the dataset.
Inclusion/exclusion	Only studies that were relevant and passed the criteria to be “Prognosis/Narrow[filter]” were included in the positive set. The other relevant studies not in the positive set were included in the negative set. All other studies were excluded from the final dataset.	Only structured abstracts identified with at least one of the PICO elements were considered to extract the text sequence.
Study types	RCTs <sup>g</sup> , systematic reviews, and meta-analysis of RCTs were given more importance.	RCTs, systematic reviews, and meta-analysis of RCTs were given more importance.

<sup>a</sup>PQR: prognosis quality recognition.

<sup>b</sup>CCA: clinical context–aware.

<sup>c</sup>PICO: Patient/Problem, Intervention, Comparison, Outcome.

<sup>d</sup>AI: artificial intelligence.

<sup>e</sup>N/A: not applicable.

<sup>f</sup>NCBI: National Center of Biotechnology Information.

<sup>g</sup>RCT: randomized controlled trial.

## Experimental Setup

To evaluate the CCA and PQR classification models, we performed experiments on a 64-bit Windows operating system with an Intel Core i5 CPU, 3.20-Hz processor with 8-GB memory using dataset D1 and D2, respectively. The experiment was performed using RapidMiner studio [54] to train and test the PQR models, whereas the python deep learning library Keras was used for the CCA classification model [55].

## PQR Model

The aim of this experiment was to quantify the comparative analysis of the proposed deep-learning model with other

machine-learning models. Using the default settings of RapidMiner, the SVM model followed the kernel type “dot,” kernel cache of 200, and complexity constant  $c$  of 0.0; the DT criterion was “gain\_ratio” with a maximum depth of 10; the  $k$  of the  $k$ -nearest neighbor was 5 with the mixed measure of “MixedEuclideanDistance.” The comparison of the performance of these models with that of a deep-learning model was assessed in terms of the F1-score, accuracy, and area under the curve (AUC) value, as shown in Table 4.

We tuned the hyperparameters of the MLP by varying the hidden layers, size of layers, activation function, and epochs, and obtained varied results as shown in Table 5.



**Table 4.** Comparative results of the deep-learning model with shallow machine-learning models.

Algorithm	F1-score	Accuracy	AUC <sup>a</sup>
Naïve Bayes	90.83	87.47	0.987
Decision tree	85.10	74.07	0.50
k-nearest neighbor	46.53	48.39	0.829
General linear model	89.34	82.38	0.904
Support vector machine	86.96	77.79	0.983
Deep learning (MLP <sup>b</sup> )	93.17	90.20	0.967

<sup>a</sup>AUC: area under the receiver operating characteristic curve.

<sup>b</sup>MLP: multilayer perceptron.

**Table 5.** Results of multilayer perceptron with varied hyperparameter settings.

Hidden layers (n)	Hidden layer size	BoW <sup>a</sup>	Activation	Epochs (n)	Recall	Precision	F1-score	Accuracy	AUC <sup>b</sup>
2	50, 50	No	Rectifier	10	90.28	96.25	93.17	90.2	0.967
3	100, 50, 25	No	Rectifier	10	93.47	96.71	95.06	92.8	0.969
3	100, 50, 25	No	Maxout	10	96.82	96.01	96.41	94.67	0.976
3	100, 50, 25	No	Maxout with Dropout	10	98.16	93.61	95.83	93.67	0.963
3	100, 50, 25	No	Tanh	10	90.62	97.65	94.00	91.44	0.978
3	100, 50, 25	Yes	Rectifier	10	93.47	98.24	95.80	93.92	0.999
3	100, 50, 25	Yes	Maxout	10	94.47	97.41	95.92	94.04	0.977
3	50, 50, 50	No	Rectifier	10	87.77	96.86	92.09	88.83	0.958
3	200, 100, 50	No	Rectifier	10	92.96	96.86	94.87	92.56	0.975
4	200, 100, 50, 25	No	Rectifier	10	93.63	96.05	94.82	92.43	0.973

<sup>a</sup>BoW: bag of words.

<sup>b</sup>AUC: area under the receiver operating characteristic curve.

As highlighted in [Table 5](#), the highest accuracy and F1-score were obtained with the setting of 3 hidden layers consisting of 100, 50, and 25 neurons, respectively. The activation function was set to Maxout, and the model was trained on 10 epochs. Finally, we boosted the performance of the selected model with

an ensemble approach using AdaBoost. The results of the optimized version of the proposed model are shown in [Table 6](#), demonstrating an F1-score of about 97% and an accuracy of 95.41% with an AUC of 0.999.

**Table 6.** Ensembling of deep-learning models.

Boosting Model	Recall	Precision	F1-score	Accuracy	AUC <sup>a</sup>
Ensemble voting (MLP <sup>b</sup> , DT <sup>c</sup> , NB <sup>d</sup> )	95.81	97.28	96.54	94.91	0.955
Proposed model (AdaBoost <sup>e</sup> -MLP)	97.99	95.9	96.93	95.41	0.999

<sup>a</sup>AUC: area under the receiver operating characteristic curve.

<sup>b</sup>MLP: multilayer perceptron.

<sup>c</sup>DT: decision tree.

<sup>d</sup>NB: naïve Bayes.

<sup>e</sup>AdaBoost: adaptive boosting.

## CCA Model

We experimented with the combination of dataset D1 described above. The classification model results are shown in [Table 7](#).

The individual decision class performance of the proposed model is reported in [Table 8](#). We found high F1-score values ( $\geq 80$ ) for classes with top support (Aim, Outcome, Results, Population) as compared to the scores of the minor classes, which indicated that the dataset size needed to be increased for each of these minor classes to obtain a higher F1-score.

**Table 7.** Comparative results of deep learning with traditional machine-learning models.

Model	Recall	Precision	F1-score	Accuracy
Logistic Regression	0.42	0.34	0.36	0.42
AdaBoost <sup>a</sup>	0.49	0.48	0.46	0.49
Gradient Boost	0.50	0.59	0.45	0.50
ANN <sup>b</sup>	0.29	0.08	0.13	0.29
kNN <sup>c</sup>	0.35	0.36	0.37	0.35
Proposed Bi-LSTM <sup>d</sup> model	0.93	0.94	0.94	0.93

<sup>a</sup>AdaBoost: adaptive boosting.

<sup>b</sup>ANN: artificial neural network.

<sup>c</sup>KNN: k-nearest neighbor.

<sup>d</sup>Bi-LSTM: bidirectional long-short term memory.

**Table 8.** Precision, recall, F1-score, and support for individual classes of the proposed deep-learning model.

Class	Precision	Recall	F1-score	Support
Aim	0.94	0.95	0.95	3133
Intervention	0.84	0.94	0.89	1238
Outcome	0.96	0.94	0.95	5036
Result	0.96	0.94	0.95	4852
Population	0.94	0.95	0.94	3082

### Proposed Semantic Similarity Algorithm (JS<sup>2</sup>E)

We measured the correlation strength based on the standard guideline of the Pearson correlation coefficient in the biomedical domain [20]. We evaluated different similarity approaches, including string similarity (Jaccard), term frequency (count vectorizer), and pretrained word embedding using GloVe, Google word2vec, and fastText. The results showed that semantic enrichment is crucial to find the similarity between the texts because it significantly increases the size of the token

set of texts by including synonyms, definitions, and preLabel from ontologies. We tested this method on a well-known biomedical dataset, BIOSSES [56], which contained 100 pairs of sentences manually annotated by the experts. As shown in Table 9, the Pearson correlation coefficient after semantic enrichment increased by 8.9% relative to that of the best-performing Jaccard similarity. Concerning the pretrained word-embedding model GloVe, word2vec (Google), and fastText (Facebook), the correlation improved by 41.8%, 60.5%, and 17.3%, respectively.

**Table 9.** Comparison of similarity approaches.

Methods	Pearson correlation coefficient (0.0-1.0)
Jaccard similarity	0.56
Cosine similarity (Count Vectorizer)	0.54
GloVe Embedding	0.43
Word2Vec (Google)	0.38
fastText (Facebook)	0.52
Jaccard Similarity after semantic enrichment (JS <sup>2</sup> E)	0.61

### Expert Evaluation of Candidate and Reference Summaries

The PICO-based summary obtained after classification and performing JS<sup>2</sup>E was then compared with the classical summary obtained using JS<sup>2</sup>E without classification. We considered the JS<sup>2</sup>E summary without classification as a baseline summary. Evaluation and comparison of two summaries were performed

by three independent evaluators and scored between 0 and 5 on the following three metrics: summary relevance to the inbound query (M1); Aim, Population, Intervention, Results, and Outcome classification representation in the summary (M2); and model summary better than the baseline summary (M3).

Table 10 shows the Pearson correlation coefficients of the scores of each evaluator (A, B, and C) concerning the average scores of the remaining two evaluators for each evaluation metric (M1,

M2, and M3). There was a strong association among the scores of each evaluator concerning each metric, suggesting that automated summarization performed best on all three parameters. The lowest correlation coefficient was 0.728, which is still considered to be high on the correlation scale.

The distribution of the scores by each of the evaluators concerning each metric is described in Table 11. The distribution suggests that there are sufficient instances for each of the scores in the evaluation dataset.

**Table 10.** Correlation matrix (Pearson correlation coefficients) of similarity approaches among evaluators for summaries according to the three metrics.

Metric	Evaluator A	Evaluator B	Evaluator C
M1 <sup>a</sup>	0.728	0.767	0.837
M2 <sup>b</sup>	0.826	0.924	0.841
M3 <sup>c</sup>	0.772	0.843	0.804

<sup>a</sup>M1: summary relevance to the inbound query.

<sup>b</sup>M2: aim, population, intervention, results, and outcome classification representation in the summary.

<sup>c</sup>M3: model summary better than the baseline summary.

**Table 11.** Frequency distribution of scores with respect to each metric by all evaluators.

Metric	Score	Frequency
M1 <sup>a</sup>	2	4
M1	3	10
M1	4	12
M1	5	4
M2 <sup>b</sup>	2	3
M2	3	10
M2	4	10
M2	5	7
M3 <sup>c</sup>	3	12
M3	4	16
M3	5	2

<sup>a</sup>M1: summary relevance to the inbound query.

<sup>b</sup>M2: aim, population, intervention, results, and outcome classification representation in the summary.

<sup>c</sup>M3: model summary better than the baseline summary.

## Discussion

### Principal Findings

The proposed Biomed-Summarizer framework generates extractive summaries for single and multiple documents with acceptable accuracy. The evaluation results signify that the proposed framework performs significantly better than existing approaches, which was also evident from the correlation matrix generated by comparing the candidate and reference summaries obtained for each defined parameter. The PQR model trained on a large dataset of biomedical literature of intracranial aneurysm showed an accuracy of 95.41% in terms of recognizing quality articles. The CCA multiclass classification model outperformed the traditional machine-learning algorithms and achieved 93% accuracy for classifying five categories: Aim, Population, Intervention, Results, and Outcome. The semantic similarity algorithm demonstrated a significant Pearson correlation coefficient of 0.61 (on a 0-1 scale) from a well-known BIOSSES dataset after semantic enrichment,

representing an improvement of 8.9% over the baseline Jaccard similarity score.

An accurate clinical summarization is expected to revolutionize the field of domain-specific knowledge graphs [57] and clinical decision support systems. For example, an individual PICO-extracted element from a biomedical text could be represented as a relationship in a knowledge graph, which could then be used for various clinical applications or could be directly supplied to clinical decision support systems for physicians to link to internal data-driven instances. The importance of linking external evidence (extracted from the biomedical literature) to internal evidence is important when the internal data are unable to capture all critical risk factors. One of our motivations was to extract the evidence from a PICO-based summarization of documents relevant to intracranial aneurysm (also known as cerebral or brain aneurysm). Since neither decision support systems nor knowledge graphs providing external evidence for complex neurological diseases such as intracranial aneurysm exist, there is a need to achieve automated summarization for

such conditions to enhance translational research and clinical decision making synergistically. The proposed automated summarization framework will be pivotal to develop a hybrid machine-learning and knowledge-based clinical decision support system, termed NeuroAssist [58], which aims to predict aneurysmal rupture (ie, subarachnoid hemorrhage). The other possible application areas of the proposed framework include automation of writing systematic reviews over a set of published clinical trials, extraction of evidence for evidence-based medicine, precision medicine, and development of clinical assistants/chatbots.

The current work in the biomedical domain is mainly focused on issues of concept- or sentence-based relevance without relating a concept or sentence to a clinical context. Although the sentence-based classification approach is well-regarded for ATS, relying on only the relevance of a sentence without capturing its clinical context and semantics may lead to a clinically undesirable summary. Some of the summarization work focuses on creating summaries from abstracts only, which may result in low recall due to missing an important sequence of text that exists only in the body of the document. In addition, in previous works, the dataset used for training and testing [29] contained only 1000 abstracts, which is not sufficient for a deep-learning model to be generalized. Recently, a new dataset was published in the BioNLP 2018 proceedings [21], but it does not consider the quality evaluation of the source documents used for extracting the text for summarization. Therefore, we curated a dataset comprising over tens of thousands of abstracts from Medline and combined it with the BioNLP dataset.

PICO-based ATS remains an unexplored area; however, work has been done on individual aspects such as PICO elements identification in the text [21,29,30], quality recognition of therapy-related biomedical studies [40,41], and sentence-based summarization without PICO and quality evaluation [3,32]. To the best of our knowledge, the proposed Biomed-Summarizer framework for biomedical ATS is the first of its kind to use a

quality recognition model, PICO-based classification using LSTM-based recurrent deep neural network model for key sentence identification trained on contextual vectors, and the JS<sup>2</sup>E algorithm for semantic matching of the user query and PICO text sequences.

The proposed approach offers a few potential benefits compared to existing methods. First, unlike traditional machine-learning approaches that depend on features that are well-structured and less noisy, deep learning deals well with an unstructured noisy text. Therefore, our deep-learning model can be reused for domains using data of the same nature. Additionally, the accuracy of the proposed deep-learning model is expected to increase further if the volume of the dataset is extended. Second, our approach considers the quality evaluation of the documents being used in summarization in addition to publishing venue credibility, which offers two-fold benefits: it enhances the confidence of physicians on the system-generated summary and has applicability in real clinical settings, and, because it filters out the documents with the lowest quality from the retrieval set, the computational time of summary creation is reduced with respect to checking the similarity of the text with the query. Third, our proposed model is based on the PICO structure, which provides additional semantics as compared to a non-PICO approach in terms of understanding the problem, interventions, and outcomes.

Traditional approaches are not cognizant of capturing the clinical context; therefore, the resultant summary includes sentences based on a high similarity score and sentence position, and is therefore less clinically effective. **Textbox 3** shows an example of a summary generated by the proposed method considering the clinical context; sequence 1 represents the aim of the study, sequence 2 represents the patient population included in the study, and sequence 3 represents the outcome of the study. In contrast, the conventional method selects the top three sequences based on high relevancy but misses the clinical context.

**Textbox 3.** Effectiveness of Biomed-Summarizer in terms of clinical significance.

#### Conventional Method

*Sequence 1: The prevailing view is that larger aneurysms have a greater risk of rupture. Predicting the risk of aneurysmal rupture, especially for aneurysms with a relatively small diameter...*

*Sequence 2: Alongside with the posterior communicating cerebral artery and middle cerebral artery of bifurcation, the anterior communicating cerebral artery (A com) is one ....*

*Sequence 3: The A com artery is known to be one of the common sites of aneurysm rupture. Given the diversity of geometry, the difference of the dominance of A1 arteries ....*

#### Proposed Method

*Sequence 1: The aim of the present study was to evaluate whether the A com aneurysms behave differently from the aneurysms located elsewhere with respect to size being a rupture risk ....*

*Sequence 2: Between January 2010 and December 2015, a total of 130 consecutive patients at our institution with the A com aneurysms-86 ruptured and 44 unruptured-were included in this study ....*

*Sequence 3: The prevailing view is that larger aneurysms have a greater risk of rupture. Predicting the risk of aneurysmal rupture, especially for aneurysms with a relatively small diameter....*

## Limitations and Future Directions

The training datasets could be reviewed for the noise generated during the creation of automated annotation to obtain even more accurate results in PQR and CCA models. The current summary is a textual summary that contains sentences in the original form as they are presented in the source documents without any further processing to extract statistical information for an easy catch up of the contents.

## Conclusion

Compared to traditional approaches, state-of-the-art deep neural network models can achieve high accuracy for an ATS task when trained on nonsparse semantically enriched features. Additionally, the automated pipeline for seeking research evidence can enable the seamless curation of voluminous biomedical literature to use in clinical decisions. By designing the proposed Biomed-Summarizer framework, we employed a

set of methods for information extraction from credible published biomedical resources to construct a simplified summary that is precise, relevant, and contextually suited to clinical needs. We designed the framework to provide openness for other researchers to use, extend, or even make use of a subpart of it and extend for designing their own systems and services. Alongside the framework, we created a specialized dataset containing PICO elements and a few other text sequences such as Aim, Method, and Result for researchers to use in their experiments in the domain of brain aneurysm. The PICO dataset was extended using a custom data-mining process by incorporating the rigorous text processing techniques on PubMed research documents. This method can be further used to create a PICO dataset in other related biomedical domains by obtaining related research papers from PubMed. The evaluation results signify that the proposed Biomed-Summarizer framework performs significantly better than existing approaches.

## Acknowledgments

This material is based upon work supported by the Brain Aneurysm Foundation and Oakland University, MI.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey questionnaire.

[DOCX File, 16 KB - [jmir\\_v22i10e19810\\_app1.docx](#)]

## References

1. Nasr Azadani M, Ghadiri N, Davoodijam E. Graph-based biomedical text summarization: An itemset mining and sentence clustering approach. *J Biomed Inform* 2018 Aug;84:42-58 [FREE Full text] [doi: [10.1016/j.jbi.2018.06.005](#)] [Medline: [29906584](#)]
2. Schulze F, Neves M. Entity-Supported Summarization of Biomedical Abstracts. In: The COLING 2016 Organizing Committee. Presented at: Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016); December 2016; Osaka, Japan p. 40-49 URL: <https://www.aclweb.org/anthology/W16-5105/>
3. Alami N, Meknassi M, En-nahnahi N. Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning. *Exp Syst App* 2019 Jun;123:195-211. [doi: [10.1016/j.eswa.2019.01.037](#)]
4. Allahyari M, Pouriyeh S, Assefi M, Safaei S, D. E, B. J, et al. Text Summarization Techniques: A Brief Survey. *arXiv* 2017 Jul 28:1-9 [FREE Full text]
5. Gambhir M, Gupta V. Recent automatic text summarization techniques: a survey. *Artif Intell Rev* 2016 Mar 29;47(1):1-66. [doi: [10.1007/s10462-016-9475-9](#)]
6. Khatri C, Singh G, Parikh N. Abstractive and Extractive Text Summarization using Document Context Vector and Recurrent Neural Networks. *arxiv* 2018 Jul 29:1807.08000 [FREE Full text]
7. Mishra R, Bian J, Fiszman M, Weir C, Jonnalagadda S, Mostafa J, et al. Text summarization in the biomedical domain: a systematic review of recent research. *J Biomed Inform* 2014 Dec;52:457-467 [FREE Full text] [doi: [10.1016/j.jbi.2014.06.009](#)] [Medline: [25016293](#)]
8. Afantenos S, Karkaletsis V, Stamatopoulos P. Summarization from medical documents: a survey. *Artif Intell Med* 2005 Feb;33(2):157-177. [doi: [10.1016/j.artmed.2004.07.017](#)] [Medline: [15811783](#)]
9. Bui DDA, Del Fiore G, Hurdle JF, Jonnalagadda S. Extractive text summarization system to aid data extraction from full text in systematic review development. *J Biomed Inform* 2016 Dec;64:265-272 [FREE Full text] [doi: [10.1016/j.jbi.2016.10.014](#)] [Medline: [27989816](#)]
10. Pivovarov R, Elhadad N. Automated methods for the summarization of electronic health records. *J Am Med Inform Assoc* 2015 Sep;22(5):938-947 [FREE Full text] [doi: [10.1093/jamia/ocv032](#)] [Medline: [25882031](#)]
11. Mikolov T, Yih W, Zweig G. Linguistic Regularities in Continuous Space Word Representations Internet. : Association for Computational Linguistics; 2013 Jun Presented at: Proceedings of the 2013 Conference of the North American Chapter



- of the Association for Computational Linguistics: Human Language Technologies; June 2013; Atlanta, Georgia p. 746-751 URL: <https://www.aclweb.org/anthology/N13-1090/>
12. Mikolov T, Chen KA, Corrado GS, Dean JA. Computing numeric representations of words in a high-dimensional space. US Patent. 2015 May 19. URL: <https://patents.google.com/patent/US9740680B1/en> [accessed 2019-08-05]
  13. Pennington J, Socher R, Manning CD. Glove: Global Vectors for Word Representation. : Association for Computational Linguistics; 2014 Oct Presented at: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP); 2014; Doha, Qatar p. 1532-1543. [doi: [10.3115/v1/d14-1162](https://doi.org/10.3115/v1/d14-1162)]
  14. Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, et al. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform 2018 Nov;87:12-20 [FREE Full text] [doi: [10.1016/j.jbi.2018.09.008](https://doi.org/10.1016/j.jbi.2018.09.008)] [Medline: [30217670](https://pubmed.ncbi.nlm.nih.gov/30217670/)]
  15. Chen W, Liu C, Yin J, Yan H, Zhang Y. Mining E-commercial data: A text-rich heterogeneous network embedding approach. : IEEE; 2017 May 14 Presented at: International Joint Conference on Neural Networks (IJCNN); May 14-19, 2017; Anchorage, AK p. 1403-1410. [doi: [10.1109/IJCNN.2017.7966017](https://doi.org/10.1109/IJCNN.2017.7966017)]
  16. Text data preprocessing. Keras. URL: <https://keras.io/preprocessing/text/> [accessed 2020-10-07]
  17. Sherstinsky A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena 2020 Mar;404:132306. [doi: [10.1016/j.physd.2019.132306](https://doi.org/10.1016/j.physd.2019.132306)]
  18. Boudin F, Shi L, Nie J. Improving medical information retrieval with PICO element detection. : Springer; 2010 Mar Presented at: European Conference on Information Retrieval; March 2010; Berlin, Heidelberg. [doi: [10.1007/978-3-642-12275-0\\_8](https://doi.org/10.1007/978-3-642-12275-0_8)]
  19. Huang KC, Liu CCH, Yang SS, Xiao F, Wong JM, Liao CC, et al. Classification of PICO elements by text features systematically extracted from PubMed abstracts. : IEEE; 2011 Nov 08 Presented at: 2011 IEEE International Conference on Granular Computing; 2011; Kaohsiung. [doi: [10.1109/grc.2011.6122608](https://doi.org/10.1109/grc.2011.6122608)]
  20. Chabou S, Iglewski M. PICO Extraction by combining the robustness of machine-learning methods with the rule-based methods. : IEEE; 2015 Jun Presented at: 2015 World Congr Inf Technol Comput Appl WCITCA 2015 Internet IEEE; June 2015; Hammamet, Tunisia. [doi: [10.1109/wcitca.2015.7367038](https://doi.org/10.1109/wcitca.2015.7367038)]
  21. Jin D, Szolovits P. PICO Element Detection in Medical Text via Deep Neural Networks. In: BioNLP 2018 Workshop.: Association for Computational Linguistics; 2018 Jul Presented at: Proceedings of the BioNLP 2018 workshop; July 2018; Melbourne, Australia URL: <https://www.aclweb.org/anthology/papers/W/W18/W18-2308/> [doi: [10.18653/v1/w18-2308](https://doi.org/10.18653/v1/w18-2308)]
  22. BioMed\_Summarizer. Brain\_Aneurysm\_Research.: GitHub URL: [https://github.com/smileslab/Brain\\_Aneurysm\\_Research/tree/master/BioMed\\_Summarizer](https://github.com/smileslab/Brain_Aneurysm_Research/tree/master/BioMed_Summarizer) [accessed 2020-10-07]
  23. Moradi M. Small-world networks for summarization of biomedical articles. arXiv 2019 Mar 7:1903.02861 [FREE Full text]
  24. Moradi M, Ghadiri N. Quantifying the informativeness for biomedical literature summarization: An itemset mining method. Comput Methods Programs Biomed 2017 Jul;146:77-89. [doi: [10.1016/j.cmpb.2017.05.011](https://doi.org/10.1016/j.cmpb.2017.05.011)] [Medline: [28688492](https://pubmed.ncbi.nlm.nih.gov/28688492/)]
  25. Luhn HP. The Automatic Creation of Literature Abstracts. IBM J Res Dev 1958 Apr;2(2):159-165. [doi: [10.1147/rd.22.0159](https://doi.org/10.1147/rd.22.0159)]
  26. Ferreira R, de Souza Cabral L, Freitas F, Lins RD, de França Silva G, Simske SJ, et al. A multi-document summarization system based on statistics and linguistic treatment. Exp Syst Appl 2014 Oct;41(13):5780-5787. [doi: [10.1016/j.eswa.2014.03.023](https://doi.org/10.1016/j.eswa.2014.03.023)]
  27. Lynn HM, Choi C, Kim P. An improved method of automatic text summarization for web contents using lexical chain with semantic-related terms. Soft Comput 2017 Apr 27;22(12):4013-4023. [doi: [10.1007/s00500-017-2612-9](https://doi.org/10.1007/s00500-017-2612-9)]
  28. Huang K, Chiang I, Xiao F, Liao C, Liu C, Wong J. PICO element detection in medical text without metadata: are first sentences enough? J Biomed Inform 2013 Oct;46(5):940-946 [FREE Full text] [doi: [10.1016/j.jbi.2013.07.009](https://doi.org/10.1016/j.jbi.2013.07.009)] [Medline: [23899909](https://pubmed.ncbi.nlm.nih.gov/23899909/)]
  29. Kim S, Martinez D, Cavedon L, Yencken L. Automatic classification of sentences to support Evidence Based Medicine. BMC Bioinformatics 2011;12(Suppl 2):S5. [doi: [10.1186/1471-2105-12-s2-s5](https://doi.org/10.1186/1471-2105-12-s2-s5)]
  30. Demner-Fushman D, Lin J. Answering Clinical Questions with Knowledge-Based and Statistical Techniques. Comp Ling 2007 Mar;33(1):63-103. [doi: [10.1162/coli.2007.33.1.63](https://doi.org/10.1162/coli.2007.33.1.63)]
  31. Zlabinger M, Andersson L, Hanbury A, Andersson M, Quasnik V, Brassey J. Medical entity corpus with pico elements and sentiment analysis. In: European Language Resources Association (ELRA). 2018 May Presented at: Eleventh International Conference on Language Resources and Evaluation (LREC 2018); 2018; Miyazaki, Japan.
  32. Wallace B, Kuiper J, Sharma A, Zhu M, Marshall I. Extracting PICO Sentences from Clinical Trial Reports using Supervised Distant Supervision. J Mach Learn Res 2016;17:132 [FREE Full text] [Medline: [27746703](https://pubmed.ncbi.nlm.nih.gov/27746703/)]
  33. Boudin F, Nie J, Bartlett JC, Grad R, Pluye P, Dawes M. Combining classifiers for robust PICO element detection. BMC Med Inform Decis Mak 2010 May 15;10(1):29 [FREE Full text] [doi: [10.1186/1472-6947-10-29](https://doi.org/10.1186/1472-6947-10-29)] [Medline: [20470429](https://pubmed.ncbi.nlm.nih.gov/20470429/)]
  34. Kilicoglu H, Demner-Fushman D, Rindfleisch TC, Wilczynski NL, Haynes RB. Towards automatic recognition of scientifically rigorous clinical research evidence. J Am Med Inform Assoc 2009;16(1):25-31 [FREE Full text] [doi: [10.1197/jamia.M2996](https://doi.org/10.1197/jamia.M2996)] [Medline: [18952929](https://pubmed.ncbi.nlm.nih.gov/18952929/)]

35. Wilczynski NL, Morgan D, Haynes RB, Hedges Team. An overview of the design and methods for retrieving high-quality studies for clinical care. *BMC Med Inform Decis Mak* 2005 Jun 21;5:20 [FREE Full text] [doi: [10.1186/1472-6947-5-20](https://doi.org/10.1186/1472-6947-5-20)] [Medline: [15969765](https://pubmed.ncbi.nlm.nih.gov/15969765/)]
36. Wilczynski N, Haynes R, Hedges Team. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Med* 2004 Jun 09;2:23 [FREE Full text] [doi: [10.1186/1741-7015-2-23](https://doi.org/10.1186/1741-7015-2-23)] [Medline: [15189561](https://pubmed.ncbi.nlm.nih.gov/15189561/)]
37. Aphinyanaphongs Y, Tsamardinou I, Statnikov A, Hardin D, Aliferis CF. Text categorization models for high-quality article retrieval in internal medicine. *J Am Med Inform Assoc* 2005;12(2):207-216 [FREE Full text] [doi: [10.1197/jamia.M1641](https://doi.org/10.1197/jamia.M1641)] [Medline: [15561789](https://pubmed.ncbi.nlm.nih.gov/15561789/)]
38. Aphinyanaphongs Y, Statnikov A, Aliferis CF. A comparison of citation metrics to machine learning filters for the identification of high quality MEDLINE documents. *J Am Med Inform Assoc* 2006;13(4):446-455 [FREE Full text] [doi: [10.1197/jamia.M2031](https://doi.org/10.1197/jamia.M2031)] [Medline: [16622165](https://pubmed.ncbi.nlm.nih.gov/16622165/)]
39. Wilczynski NL, McKibbin KA, Walter SD, Garg AX, Haynes RB. MEDLINE clinical queries are robust when searching in recent publishing years. *J Am Med Inform Assoc* 2013;20(2):363-368 [FREE Full text] [doi: [10.1136/amiainl-2012-001075](https://doi.org/10.1136/amiainl-2012-001075)] [Medline: [23019242](https://pubmed.ncbi.nlm.nih.gov/23019242/)]
40. Del Fiol G, Michelson M, Iorio A, Cotoi C, Haynes R. A Deep Learning Method to Automatically Identify Reports of Scientifically Rigorous Clinical Research from the Biomedical Literature: Comparative Analytic Study. *J Med Internet Res* 2018 Jun 25;20(6):e10281 [FREE Full text] [doi: [10.2196/10281](https://doi.org/10.2196/10281)] [Medline: [29941415](https://pubmed.ncbi.nlm.nih.gov/29941415/)]
41. Afzal M, Hussain M, Malik KM, Lee S. Impact of Automatic Query Generation and Quality Recognition Using Deep Learning to Curate Evidence From Biomedical Literature: Empirical Study. *JMIR Med Inform* 2019 Dec 09;7(4):e13430 [FREE Full text] [doi: [10.2196/13430](https://doi.org/10.2196/13430)] [Medline: [31815673](https://pubmed.ncbi.nlm.nih.gov/31815673/)]
42. Afzal M, Hussain M, Haynes R, Lee S. Context-aware grading of quality evidences for evidence-based decision-making. *Health Informatics J* 2019 Jun;25(2):429-445. [doi: [10.1177/1460458217719560](https://doi.org/10.1177/1460458217719560)] [Medline: [28766402](https://pubmed.ncbi.nlm.nih.gov/28766402/)]
43. Irwin AN, Rackham D. Comparison of the time-to-indexing in PubMed between biomedical journals according to impact factor, discipline, and focus. *Res Social Adm Pharm* 2017;13(2):389-393. [doi: [10.1016/j.sapharm.2016.04.006](https://doi.org/10.1016/j.sapharm.2016.04.006)] [Medline: [27215603](https://pubmed.ncbi.nlm.nih.gov/27215603/)]
44. Morid MA, Fiszman M, Raja K, Jonnalagadda SR, Del Fiol G. Classification of clinically useful sentences in clinical evidence resources. *J Biomed Inform* 2016 Apr;60:14-22 [FREE Full text] [doi: [10.1016/j.jbi.2016.01.003](https://doi.org/10.1016/j.jbi.2016.01.003)] [Medline: [26774763](https://pubmed.ncbi.nlm.nih.gov/26774763/)]
45. Goularte FB, Nassar SM, Fileto R, Saggion H. A text summarization method based on fuzzy rules and applicable to automated assessment. *Exp Syst Appl* 2019 Jan;115:264-275. [doi: [10.1016/j.eswa.2018.07.047](https://doi.org/10.1016/j.eswa.2018.07.047)]
46. Cao Z, Li W, Li S, Wei F, Li Y. AttSum: Joint learning of focusing and summarization with neural attention. *arXiv* 2016 Sep 16:1604.00125 [FREE Full text]
47. Nallapati R, Zhai F, Zhou B. SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents. *arxiv* 2016 Nov 14:1611.04230 [FREE Full text]
48. Cheng J, Lapata M. Neural Summarization by Extracting Sentences and Words. *arXiv* 2016 Mar 23:1603.07252 [FREE Full text]
49. Cao Z, Wei F, Dong L, Li S, Zhou M. Ranking with Recursive Neural Networks and Its Application to Multi-Document Summarization. 2015 Jan 25 Presented at: AAAI; 2015; Austin, Texas, USA p. 2153-2159.
50. Loper E, Bird S. NLTK: The Natural Language Toolkit. *arXiv* 2002 May 17:0205028 [FREE Full text]
51. Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 2005 Mar 23;5(1):25 [FREE Full text] [doi: [10.1186/1472-6963-5-25](https://doi.org/10.1186/1472-6963-5-25)] [Medline: [15788089](https://pubmed.ncbi.nlm.nih.gov/15788089/)]
52. Kitto SC, Chesters J, Grbich C. Quality in qualitative research. *Med J Austral* 2008 Feb 18;188(4):243-246. [doi: [10.5694/j.1326-5377.2008.tb01595.x](https://doi.org/10.5694/j.1326-5377.2008.tb01595.x)]
53. Bio.Entrez package. Biopython. URL: <https://biopython.org/DIST/docs/api/Bio.Entrez-module.html> [accessed 2019-08-20]
54. RapidMiner. URL: <https://rapidminer.com/> [accessed 2019-08-21]
55. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Machine Learn Res* 2011 Nov:2825-2830 [FREE Full text]
56. Sogancioglu G, Öztürk H, Özgür A. BIOSSES: a semantic sentence similarity estimation system for the biomedical domain. *Bioinformatics* 2017 Jul 15;33(14):i49-i58 [FREE Full text] [doi: [10.1093/bioinformatics/btx238](https://doi.org/10.1093/bioinformatics/btx238)] [Medline: [28881973](https://pubmed.ncbi.nlm.nih.gov/28881973/)]
57. Malik KM, Krishnamurthy M, Alobaidi M, Hussain M, Alam F, Malik G. Automated domain-specific healthcare knowledge graph curation framework: Subarachnoid hemorrhage as phenotype. *Exp Syst Appl* 2020 May;145:113120. [doi: [10.1016/j.eswa.2019.113120](https://doi.org/10.1016/j.eswa.2019.113120)]
58. Neuroassist: an intelligent secure decision support system for prediction of brain aneurysm rupture, funded by Brain Aneurysm Foundation and Oakland University. SmilesLab. 2020. URL: <https://www.secs.oakland.edu/~Mahmood/details.html#one> [accessed 2020-10-07]

## Abbreviations

**AdaBoost:** adaptive boosting  
**ATS:** automatic text summarization  
**AUC:** area under the receiver operating characteristic curve  
**Bi-LSTM:** bidirectional long-short term memory  
**BoW:** bag of words  
**CCA:** clinical context-aware  
**JS<sup>2</sup>E:** Jaccard similarity with semantic enrichments  
**MeSH:** Medical Subject Heading  
**MLP:** multilayer perceptron  
**PICO:** Patient/Problem, Intervention, Comparison, and Outcome  
**PQR:** prognosis quality recognition  
**RNN:** recurrent neural network  
**SVM:** support vector machine

*Edited by G Eysenbach; submitted 02.05.20; peer-reviewed by F ALi, S Chen; comments to author 27.07.20; accepted 24.09.20; published 23.10.20.*

*Please cite as:*

Afzal M, Alam F, Malik KM, Malik GM

Clinical Context-Aware Biomedical Text Summarization Using Deep Neural Network: Model Development and Validation

J Med Internet Res 2020;22(10):e19810

URL: <http://www.jmir.org/2020/10/e19810/>

doi: [10.2196/19810](https://doi.org/10.2196/19810)

PMID: [33095174](https://pubmed.ncbi.nlm.nih.gov/33095174/)

©Muhammad Afzal, Fakhare Alam, Khalid Mahmood Malik, Ghaus M Malik. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Correlation of Online Physician Rating Subscores and Association With Overall Satisfaction: Observational Study of 212,933 Providers

Hanson Hanqing Zhao<sup>1</sup>, MD; Michael Luu<sup>1,2</sup>, MPH; Brennan Spiegel<sup>2,3,4</sup>, MD; Timothy John Daskivich<sup>1,2</sup>, MSHPM, MD

<sup>1</sup>Division of Urology, Cedars-Sinai Medical Center, Los Angeles, CA, United States

<sup>2</sup>Center for Outcomes Research and Education, Cedars-Sinai Medical Center, Los Angeles, CA, United States

<sup>3</sup>Division of Health Services Research, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, United States

<sup>4</sup>Department of Health Policy and Management, University of California, Los Angeles, Los Angeles, CA, United States

**Corresponding Author:**

Timothy John Daskivich, MSHPM, MD

Division of Urology

Cedars-Sinai Medical Center

8635 West 3rd Street, Suite 1070W

Los Angeles, CA

United States

Phone: 1 310 423 4700

Email: [timothy.daskivich@csmc.edu](mailto:timothy.daskivich@csmc.edu)

## Abstract

**Background:** Online physician rating websites commonly ask consumers to rate providers across multiple physician-based (eg, spending sufficient time, listening) and office-based (eg, appointment scheduling, friendliness) subdimensions of care in addition to overall satisfaction. However, it is unclear if consumers can differentiate between the various rated subdimensions of physicians. It is also unclear how each subdimension is related to overall satisfaction.

**Objective:** The objectives of our study were to determine the correlation of physician-based and office-based subdimensions of care and the association of each with overall satisfaction.

**Methods:** We sampled 212,933 providers from the Healthgrades website and calculated average provider metrics for overall satisfaction (likelihood to recommend doctor), physician-based subdimensions (trust in physician, ability to explain, ability to listen and answer questions, and spending adequate time), and office-based subdimensions (ease of scheduling, office environment, staff friendliness, and wait time). We used Spearman rank correlation to assess correlation between subdimension ratings. Factor analysis was used to identify potential latent factors predicting overall satisfaction. Univariate and multivariable linear regression were performed to assess the effect of physician and office-based factors on overall satisfaction.

**Results:** Physician-based metrics were highly correlated with each other ( $r=.95$  to  $.98$ ,  $P<.001$ ), as were office-based metrics ( $r=.84$  to  $.88$ ,  $P<.001$ ). Correlations between physician-based and office-based ratings were less robust ( $r=.79$  to  $.81$ ,  $P<.001$ ). Factor analysis identified two factors, clearly distinguishing between physician-based metrics (factor loading = 0.84 to 0.88) and office-based metrics (factor loading = 0.76 to 0.84). In multivariable linear regression analysis, the composite factor representing physician-based metrics (0.65, 95% CI 0.65 to 0.65) was more strongly associated with overall satisfaction than the factor representing office-based metrics (0.42, 95% CI 0.42 to 0.42). These factors eclipsed other demographic variables in predicting overall satisfaction.

**Conclusions:** Consumers do not differentiate between commonly assessed subdimensions of physician-based care or subdimensions of office-based care, but composite factors representing these broader categories are associated with overall satisfaction. These findings argue for a simpler ratings system based on two metrics: one addressing physician-based aspects of care and another addressing office-based aspects of care.

(*J Med Internet Res* 2020;22(10):e11258) doi:[10.2196/11258](https://doi.org/10.2196/11258)



**KEYWORDS**

online ratings; Healthgrades; physician ratings

**Introduction**

Online physician ratings websites have become an increasingly influential source of information for health care consumers [1-3]. While rating websites provide transparency and a platform for patient feedback, there is little data supporting their validity and utility in identifying high-quality care [4-6]. Nevertheless, studies suggest that consumers believe these sites are important in choosing a physician [7]. A survey of 1000 surgical patients at Mayo Clinic found that 75% would choose a physician and 88% would avoid a physician based on ratings alone [8]. As testament to the growing acceptance and trust in consumer ratings, payers and institutions now list commercial consumer ratings as part of online provider listings, and a percentage of Medicare payments are redistributed to hospitals with higher quality metrics and better patient evaluations [9-11]. Of the top 20 hospitals in a recent US News & World Report ranking, 10 of them currently display ratings for their providers [12].

Most ratings websites prominently feature overall satisfaction scores on a 5-star Likert scale [13,14]. To improve clarity, some websites additionally ask consumers to rate physicians on specific subdimensions of care related to the physician's bedside manner (eg, level of trust in provider's decisions, how well the provider explains medical conditions, how well the provider listens and answers questions, spending sufficient time with patients) and office (eg, ease of scheduling, staff friendliness, total wait time, office environment) [15-17]. Although there is good face validity in asking consumers to rate physicians on these discrete, service-related aspects of care, there is a lack of data showing that patients actually distinguish between these subdimensions. Furthermore, the individual contribution of each subdimension to overall satisfaction is unknown.

In this study, we analyzed a large, heterogeneous sample of quantitative online reviews to determine if consumers are able to parse different components of the patient experience and identify physician and office characteristics that predict higher overall satisfaction scores. We hypothesized that all physician-related scores would be highly correlated, since patients are asked to rate subdimensions that are all related to bedside manner, while office-based scores may vary, as they measure distinct aspects of care that are unrelated.

**Methods****Data Source**

We sampled online consumer reviews for providers in the United States from the Healthgrades website using a method that has previously been described [18]. The dataset consisted of 2.7 million reviews for 830,308 providers up to March 31, 2017. These data were linked with demographic information from the US Centers for Medicare & Medicaid Services Physician Compare website using National Provider Identifier numbers: this information included medical specialty, region, gender, and year of graduation from medical school. In order to sample

physicians with an adequate number of reviews, we excluded physicians with 4 or fewer reviews ( $n=611,013$ ). We also excluded physicians who were missing information on their primary specialty ( $n=1813$ ) or who were identified as nursing or nonclinical specialty providers ( $n=4549$ ). Our final analytic sample comprised 212,933 physicians. The study was approved by the Cedars-Sinai IRB.

**Physician Rating Selection**

Average provider metrics across all reviews on a 5-star Likert scale were collected for overall satisfaction and subdimensions of perceived physician quality including level of trust in provider's decisions, how well the provider explains medical conditions, how well the provider listens and answers questions, and spending the appropriate amount of time with patients (see [Multimedia Appendix 1](#) for a screenshot of the homepage and a sample review). Office-based metrics were also collected across subdimensions of ease of scheduling urgent appointments, office environment, staff friendliness and courteousness, and total wait time.

**Statistical Analysis**

Physician demographics were described using median and interquartile range for continuous variables and counts with percentages for categorical variables.

Spearman rank correlation coefficient was used to assess the correlation between overall satisfaction, physician-based ratings subdimensions, and office-based ratings subdimensions. Additionally, a scatter plot matrix was used to visually depict the strength of association between pairings of overall satisfaction, physician-based, and office-based metrics.

To identify potential latent factors among our physician-based and office-based metrics, exploratory and confirmatory factor analysis was conducted [19]. Sampling adequacy was confirmed by the Kaiser-Meyer-Olkin statistic (0.93) to examine the appropriateness of the sample size for conducting exploratory factor analysis [20]. Sampling adequacy values between .80 to .90 are considered excellent, where values between .50 and .60 are considered marginal, and below .50 considered unacceptable [21]. The Bartlett test of sphericity was also conducted to test the null hypothesis of an identity matrix and the suitability for factor extraction ( $P<.001$ ) [22].

To determine the number of potential latent factors in our sample, we conducted the Horn parallel analysis [23]. Parallel analysis involves the generation of a random dataset with the same number of observations and variables. The eigenvalues and correlation matrix are computed from this dataset and compared with results of the eigenvalues from factor extraction. The point at which the eigenvalues from the random data exceed the values from factor extraction indicates that any further factors encompass primarily random noise. We identified the point at which the decrease in eigenvalues became negligible on the scree plot ([Figure 1](#)), which revealed a 2-factor solution. Factor extraction was conducted based on the orthogonal varimax rotation and extracted using the maximum likelihood.



Exploratory factor analysis was performed using a 1-, 2-, and 3-factor solution with the resulting cumulative variance explained of .840, .929, and .932, respectively.

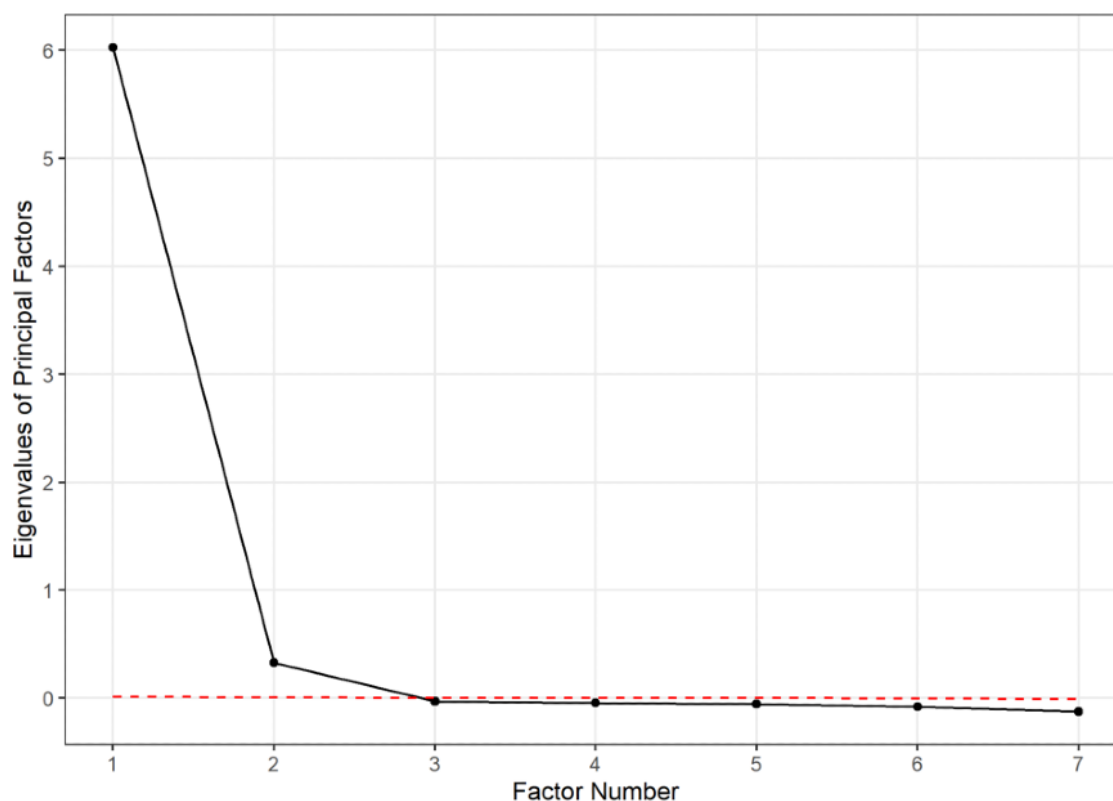
Confirmatory factor analysis was further conducted to test 1-, 2-, and 3-factor models of overall patient satisfaction. The hypothesized latent factor structure is overall physician satisfaction measured by physician ratings subdomains. The confirmatory factor analysis model was fit using lavaan version 0.5-23 (Rosseel, 2012) and showed acceptable goodness of fit (Tucker-Lewis index 0.989, comparative fit index .993, root mean square error of approximation 0.087 [90% CI 0.086 to 0.088], goodness of fit index 0.975, adjusted goodness of fit index 0.945, and standardized root mean square residual 0.01) with all subdomains loading significantly on their hypothesized latent factors ( $P < .001$ ). Furthermore, discriminatory validity of the measures has been assessed using composite reliability and the average variance extracted for the 2-factor solution. Composite reliability among the physician satisfaction measures (0.99) and office staff satisfaction measures (0.95) both showed

high internal consistency [24]. The average variance extracted also showed a high amount of variance captured by the two factors solution for both physician satisfaction (0.97) and office staff satisfaction (0.87) [25]. In contrast, the fit of the 1-factor model provided a lower chi-square (262,257 vs 20,950), lower Akaike information criterion (219,065 vs 460,370), lower root mean square error of approximation (0.087 vs 0.297), and higher comparative fit index (0.993 vs 0.912) and goodness of fit index (0.975 vs 0.730) in comparison with the 2-factor model.

To assess the relative impact of physician-based and office-based metrics on overall satisfaction, univariate linear regression was performed [26]. Additionally, a multivariable linear regression model was performed regressing the saved factor scores extracted from exploratory factor analysis on overall satisfaction adjusting for all physician demographics.

All statistical analyses were performed using R version 3.5.1 (R Foundation for Statistical Computing) with 2-sided test and significance level of .05 [27].

**Figure 1.** Scree plot.



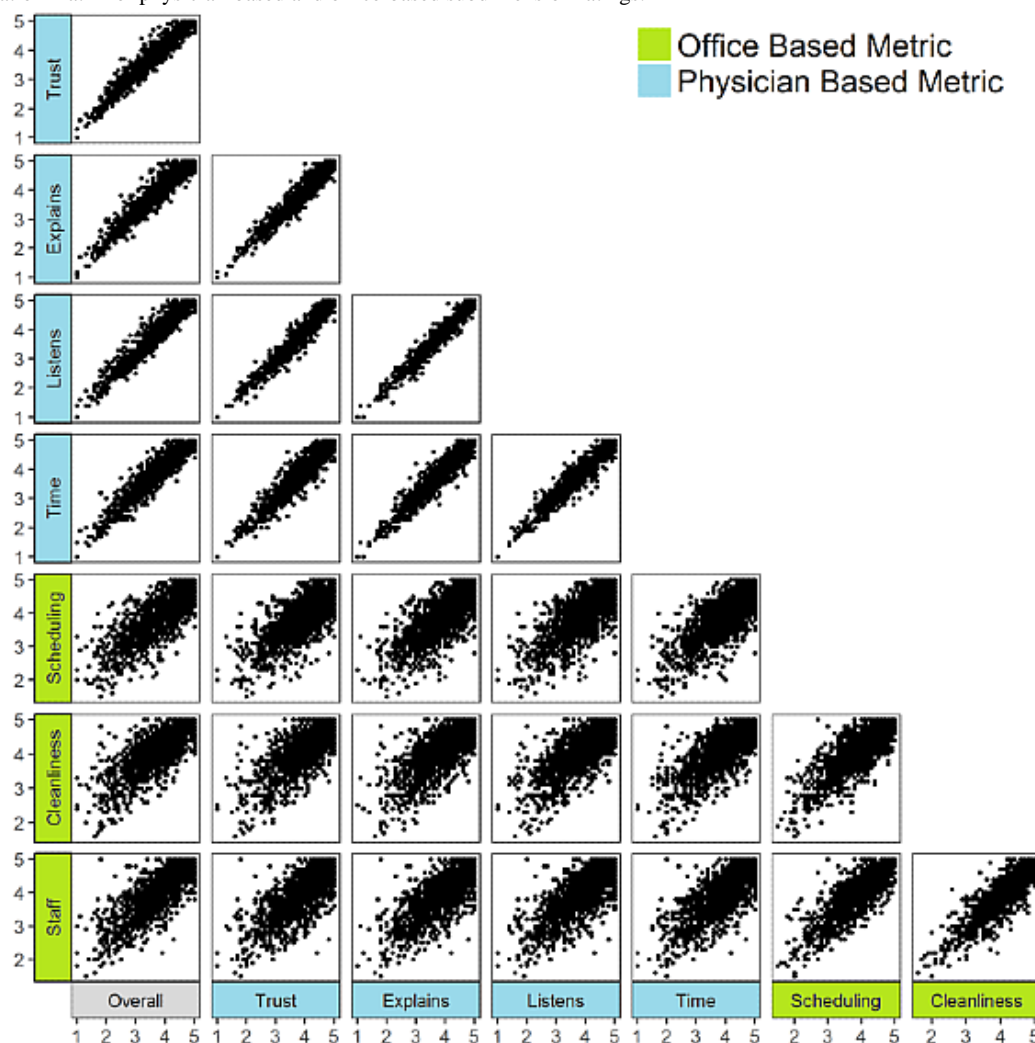
## Results

Physician characteristics are summarized in Table 1. The majority of our sample were medical specialists (128,678/212,933, 60.43%) from the southern United States (80,751/212,701, 37.96%) who graduated from medical school after 1985 (146,246/209,095, 69.94%). Median scores for overall satisfaction, physician-based metrics, and office-based metrics were universally high (range 4.1 to 4.3) and interquartile ranges for scores were narrow. The majority of wait times were within 10 to 15 minutes (113,51/212,921, 53.31%).

Physician-based metrics were highly correlated with each other ( $r = .95$  to  $.98$ ,  $P < .001$ ), as were office-based metrics ( $r = .84$  to  $.88$ ,  $P < .001$ ; Figure 2). Correlations between physician-based and office-based ratings were less robust ( $r = .79$  to  $.81$ ,  $P < .001$ ). Overall patient satisfaction correlated more strongly with physician-based metrics ( $r = .95$  to  $.97$ ,  $P < .001$ ) than office-based metrics ( $r = .82$  to  $.84$ ,  $P < .001$ ). Distributions of subdimension scores for providers with overall satisfaction scores of 5.0, 4.0, 3.0, and 2.0 were narrow, with interquartile range of subdimensions spanning a maximum of 0.4 points for physician-based subdimensions and 0.8 points for office-based subdimensions (Table 2).

**Table 1.** Physician demographics (n=212,933).

Characteristics	Values
<b>Physician specialty group, n (%)</b>	
Medical specialties	128,678 (60.43)
Allied health providers	11,724 (5.51)
Surgical specialties	72,531 (34.06)
<b>Geographical region, n (%)</b>	
Midwest	44,069 (20.72)
Northeast	45,616 (21.45)
South	80,751 (37.96)
West	42,265 (19.87)
<b>Year of graduation, n (%)</b>	
1945-1954	57 (0.03)
1955-1964	1579 (0.74)
1965-1974	1,3475 (6.33)
1975-1984	47,738 (22.42)
1985-1994	64,498 (30.29)
1995-2004	61,338 (28.81)
2005-2014	20,349 (9.56)
2015-2016	61 (0.03)
Unknown	3838 (1.80)
Overall patient satisfaction, median (IQR)	4.10 (3.40, 4.60)
<b>Physician-based subdomains, median (IQR)</b>	
Trust (level of trust in provider's decision)	4.20 (3.60, 4.60)
Explains (how well provider explains medical conditions)	4.20 (3.60, 4.60)
Listens (how well provider listens and answers questions)	4.20 (3.60, 4.60)
Time (spends appropriate amount of time with patients)	4.20 (3.60, 4.60)
<b>Office-based subdomains, median (IQR)</b>	
Scheduling (ease of scheduling urgent appointments)	4.20 (3.60, 4.60)
Cleanliness (office environment, cleanliness, comfort)	4.30 (3.90, 4.70)
Staff (staff friendliness and courteousness)	4.20 (3.70, 4.60)
<b>Total wait time in minutes, n (%)</b>	
<10	31,177 (14.64)
10-15	113,517 (53.31)
16-30	54,412 (25.55)
31-45	12,907 (6.06)
Over 45	908 (0.43)
Unknown	12 (0.01)

**Figure 2.** Correlation matrix of physician-based and office-based subdimension ratings.**Table 2.** Distribution of subdimension scores for providers with overall satisfaction scores of 5.0, 4.0, 3.0, and 2.0.

Subdimension	5.0	4.0	3.0	2.0
<b>Physician-based, median (IQR)</b>				
Trust	5.0 (5.0-5.0)	4.1 (4.0-4.2)	3.2 (3.0-3.3)	2.3 (2.1-2.4)
Explains	5.0 (5.0-5.0)	4.1 (4.0-4.3)	3.2 (3.0-3.3)	2.2 (2.1-2.4)
Listens	5.0 (5.0-5.0)	4.1 (4.0-4.2)	3.2 (3.0-3.3)	2.2 (2.0-2.3)
Time spent	5.0 (5.0-5.0)	4.1 (4.0-4.3)	3.2 (3.0-3.4)	2.3 (2.0-2.4)
<b>Office-based, median (IQR)</b>				
Scheduling	4.8 (4.7-5.0)	4.1 (3.9-4.3)	3.4 (3.1-3.7)	2.7 (2.3-3.1)
Office cleanliness	4.9 (4.8-5.0)	4.3 (4.1-4.5)	3.7 (3.4-4.0)	3.0 (2.2-2.4)
Staff friendliness	4.9 (4.8-5.0)	4.1 (3.9-4.4)	3.5 (3.2-3.8)	2.8 (2.4-3.1)

Factor analysis was used to identify latent clusters of variables predicting overall patient satisfaction. One-, 2-, and 3-factor solutions were tested. Although a 1-factor solution explains the majority of the variance, the second factor explains an additional ~10% variance, where the third factor provides negligible information, which supports the results of the parallel analysis (Figure 1, Table 3). In the 2-factor model, two discrete clusters

of variables exceeded the a priori defined loading cutoff of  $\geq 0.70$  (Table 3). These clusters corresponded directly with physician-based metrics (factor 1 loading values 0.84 to 0.88) and office-based metrics (factor 2 loading values 0.76 to 0.84). Confirmatory factor analysis showed that the individual subdomains loaded successfully upon the a priori hypothesized latent factor structure (Table 4).

**Table 3.** Factor loadings from exploratory factor analysis.

Physician ratings subdimension	1-Factor solution	2-Factor solution		3-Factor solution		
	Factor 1	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3
Level of trust in provider's decision	0.99	0.86	0.48	0.84	0.52	0.11
How well provider explains medical conditions	0.99	0.87	0.47	0.85	0.51	N/A <sup>a</sup>
How well provider listens and answers questions	0.99	0.88	0.46	0.86	0.50	N/A
Spends appropriate amount of time with patients	0.98	0.84	0.49	0.82	0.53	N/A
Ease of scheduling urgent appointments	0.81	0.50	0.76	0.47	0.78	N/A
Office environment, cleanliness, comfort, etc	0.81	0.49	0.78	0.46	0.80	N/A
Staff friendliness and courteousness	0.82	0.46	0.84	0.43	0.86	N/A
Proportion variance explained	0.84	0.53	0.40	0.49	0.44	0.00
Cumulative variance explained	0.84	0.53	0.93	0.49	0.93	0.93

<sup>a</sup>N/A: Not applicable.**Table 4.** Factor loadings from confirmatory factor analysis.

Latent factor and items	Loadings	P value
<b>Physician-based metrics</b>		
Level of trust in provider's decision	0.98	<.001
How well provider explains medical conditions	0.99	<.001
How well provider listens and answers questions	0.99	<.001
Spends appropriate amount of time with patients	0.98	<.001
<b>Office-based metrics</b>		
Ease of scheduling urgent appointments	0.92	<.001
Office environment, cleanliness, comfort, etc.	0.93	<.001
Staff friendliness and courteousness	0.95	<.001

In univariable linear regression analysis, all physician-based metrics and office-based metrics were associated with overall satisfaction (Table 5). The physician-based subdimensions most strongly associated with overall satisfaction were trust in physician and ability to explain, with overall satisfaction ratings increasing by 1.05 (95% CI 1.04 to 1.05) and 1.03 (95% CI 1.03 to 1.03) points for each point increase in subdimension score, respectively. The office-based subdimension most strongly associated with overall satisfaction was office cleanliness, with overall satisfaction ratings increasing by 1.09 (95% CI 1.09 to 1.1) for each point increase in office cleanliness score. Stepwise increases in office wait times were strongly associated with worsening overall satisfaction ratings; for example, compared with those with total wait time under 10 minutes, a wait time

of 31 to 45 minutes was associated with a decrease of –1.35 (95% CI –1.37 to –1.34) in overall satisfaction score. Since subdimension scores were highly correlated, latent factors identifying physician-based metrics and office-based metrics were used in multivariable analysis. In multivariable linear regression, physician-based metrics (0.65; 95% CI 0.65 to 0.65,  $P<.001$ ) were more strongly associated with overall satisfaction than office-based metrics (0.42; 95% CI 0.42 to 0.42,  $P<.001$ ), and the association of office wait times was strikingly diminished (Table 5).

While physician demographics such as practice region and years in practice were also associated with overall satisfaction score in univariable analysis, none were meaningfully associated with overall satisfaction in multivariable analysis (Table 5).

**Table 5.** Univariable and multivariable linear regression model predicting overall satisfaction.

Characteristics	Univariable		Multivariable	
	$\beta$ (95% CI)	<i>P</i> value	$\beta$ (95% CI)	<i>P</i> value
<b>Physician specialty group, n (%)</b>				
Medical specialties	Reference		Reference	
Allied health providers	0.48 (0.47 to 0.50)	<.001	0.00 (–0.01 to 0.00)	.20
Surgical specialties	0.18 (0.17 to 0.19)	<.001	0.01 (0.01 to 0.01)	<.001
<b>Geographical region, n (%)</b>				
Midwest	Reference		Reference	
Northeast	0.04 (0.03 to 0.05)	<.001	0.00 (0.00 to 0.00)	.30
South	0.00 (–0.01 to 0.01)	.53	–0.01 (–0.01 to 0.00)	<.001
West	–0.08 (–0.09 to –0.06)	<.001	0.00 (0.00 to 0.00)	.01
<b>Year of graduation, n (%)</b>				
1945-1954	Reference		Reference	
1955-1964	–0.06 (–0.27 to 0.15)	.60	0.01 (–0.03 to 0.05)	.61
1965-1974	–0.04 (–0.25 to 0.16)	.68	0.01 (–0.03 to 0.05)	.63
1975-1984	0.04 (–0.16 to 0.25)	.68	0.01 (–0.04 to 0.05)	.79
1985-1994	0.08 (–0.13 to 0.28)	.46	0.00 (–0.04 to 0.04)	.92
1995-2004	0.18 (–0.02 to 0.39)	.08	0.00 (–0.04 to 0.04)	.92
2005-2014	0.32 (0.11 to 0.53)	.002	0.01 (–0.04 to 0.05)	.77
2015-2016	0.72 (0.44 to 1.01)	<.001	0.01 (–0.04 to 0.07)	.62
<b>Physician-based metrics</b>			0.65 (0.65 to 0.65) <sup>a</sup>	<.001
Trust	1.05 (1.04 to 1.05)	<.001	N/A <sup>b</sup>	N/A
Explains	1.03 (1.03 to 1.03)	<.001	N/A	N/A
Listens	1.01 (1.01 to 1.01)	<.001	N/A	N/A
Time	1.02 (1.02 to 1.02)	<.001	N/A	N/A
<b>Office-based metrics</b>			0.42 (0.42 to 0.42) <sup>a</sup>	<.001
Scheduling	0.99 (0.99 to 0.99)	<.001	N/A	N/A
Cleanliness	1.09 (1.09 to 1.10)	<.001	N/A	N/A
Staff	1.02 (1.02 to 1.02)	<.001	N/A	N/A
<b>Total wait time in minutes</b>				
<10	Reference		Reference	
10-15	–0.43 (–0.44 to –0.42)	<.001	–0.01 (–0.01 to –0.01)	<.001
16-30	–0.89 (–0.89 to –0.88)	<.001	–0.02 (–0.03 to –0.02)	<.001
31-45	–1.35 (–1.37 to –1.34)	<.001	–0.06 (–0.06 to –0.05)	<.001
>45	–1.90 (–1.95 to –1.85)	<.001	–0.10 (–0.11 to –0.09)	<.001

<sup>a</sup>Factor score from exploratory factor analysis.<sup>b</sup>N/A: Not applicable.

## Discussion

### Principal Findings

Online physician ratings have been steadily gaining popularity, with physicians rated a median of 7 times across commercially available websites [1,28]. More consumers are aware of these

ratings and are using them as the primary source of information to guide their health care decisions [7,29]. In addition, more physicians are now being rated across multiple platforms [2,3,30], yet despite the groundswell of interest and uptake of online ratings, very little data exist to support their validity and utility in assisting consumers choose better physicians [4,13,31-33]. In this analysis of a large sample of quantitative



online reviews, we found that physician-based subdimensions were very highly correlated with one another, demonstrating that consumers rarely differentiate between the commonly rated subdimensions of physician care. Office-based subdimensions of care were also found to be highly correlated with one another. However, there was more heterogeneity observed when comparing physician-based subdimensions with office-based subdimensions, suggesting that patients are better at parsing between the perceived quality of the physician versus their office staff. Factor analysis objectively supports this contention, clearly identifying two discrete factors predicting overall satisfaction, one clustered around physician-based care and one around office-based aspects of care. In multivariable regression analysis, the composite factors measuring physician- and office-based care far eclipsed other demographics in prediction of overall satisfaction in terms of magnitude. We believe this data suggests that physician ratings should be simplified to two simple metrics: one evaluating physician-based care and one evaluating office-based services.

The principal finding of our study is that commonly rated subdimensions of physician-based care are highly correlated with one another, which calls into question their utility over a single measure of satisfaction with the physician. Either the vast majority of physicians are consistently all good, average, or bad across all categories of care or consumers are unable to discriminate between the measured characteristics of their physicians. Since the former explanation does not seem likely, we favor the latter explanation. Kadry et al [34] also found a high correlation between various subdimensions of care across multiple rating websites in their analysis of 4999 total ratings and argued for a single rating for physician-based care. Indeed, based on our more comprehensive data analysis, a single measure of satisfaction with the physician and a single measure satisfaction with the office staff would suffice. Reducing the number of ratings could improve the understandability of these reviews and increase response rates [35,36].

An additional explanation for why the physician-based subdimension scores in our study may be highly correlated is the fact that they are measuring a similar construct: bedside manner. While the office-based subdimensions measure discrete, quantifiable characteristics such as office environment and ease of scheduling appointments, the physician-based subdimensions measure aspects of relationship building between doctor and patient [37,38]. It is difficult to conceive of a physician who would be superb at one aspect of relationship building (eg, listening and answering questions) and abysmal at another (eg, building trust). To our knowledge, none of the subcomponent scores of online ratings have undergone rigorous psychometric validation to determine if they are measuring distinct constructs. Although our study is not a psychometric assessment, it does suggest deficiency in discrimination between the subdimension scores by their overwhelming correlation with each other. As an extension of this line of thought, an alternative to reducing physician ratings to single measures of physician- and office-based care would be to identify components of care that patients can differentiate between using rigorous psychometric techniques.

While online ratings may be flawed, they are clearly an important source of direct consumer feedback, and we believe that these ratings have the potential to give physicians important quality improvement feedback [4,6,14,39,40]. While composite measures of physician- and office-based care were the predominant predictors of overall satisfaction in multivariable linear regression, there were some other notable characteristics that are worth mentioning. In univariate analysis, incremental increases in wait time predicted significantly worse ratings. On just a 5-point Likert scale, physicians with wait times over 45 minutes had an average of a 1.9-point lower rating compared with physicians with wait times under 10 minutes. Even physicians with wait times of just 10 to 15 minutes had nearly a half point decrease in ratings. As physician ratings do not fall under a normal distribution, these decreases can have a significant impact in the online perception of a physician when compared with his or her peers [18]. Interestingly, physician age and experience did not seem to affect their ratings with the exception of physicians who had graduated between 2005 and 2016. While surveys have shown that patients generally prefer physicians toward the middle of their career, this younger group actually had higher ratings despite less clinical experience. Gao et al [2] also had similar findings from another physician review website [41]. Younger physicians may have a better understanding of their online presence and focus more time identifying ways to improve their rating. Nonetheless, as mentioned above, physician-based metrics and office-based metrics far outstripped these demographic predictors of overall satisfaction in our multivariable model.

## Limitations

There are several limitations to our study. We aggregated data only from one website. However, this is the most frequented physician rating website with a large, heterogeneous mixture of physicians from across the United States [34]. Aggregation of data from multiple websites may be impractical given the large number of websites and various rating methods. In addition, these reviews naturally have an implicit selection bias and may not always be authentic. To minimize this bias, we only included physicians with more than the median number of reviews (4).

## Conclusions

In this analysis of the ratings of 212,933 providers, we found that consumers do not often differentiate between commonly assessed physician-based subdimensions of care. Physicians were most often scored in a monochrome fashion: scores all good, average, or bad. Office-based subdimensions of care were also highly correlated and were scored in a similarly monochrome fashion. In multivariable analysis, composite latent factors identifying physician-based metrics and office-based metrics were both independently associated with overall satisfaction scores, eclipsing all other physician demographic predictors in terms of magnitude. Based on this, we question the utility of commonly used subdimension scores and instead recommend a single measure of satisfaction for the physician and a single measure of satisfaction for the office staff. Alternatively, further research should be conducted to identify

qualities of physicians and office staff that consumers are well positioned to evaluate and are meaningful to patient experience.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Healthgrades.com homepage and sample physician review.

[DOCX File, 2057 KB - [jmir\\_v22i10e11258\\_app1.docx](#)]

## References

1. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. *JAMA* 2017 Dec 21;317(7):766-768 [FREE Full text] [doi: [10.1001/jama.2016.18553](#)] [Medline: [28241346](#)]
2. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. *J Med Internet Res* 2012;14(1):e38 [FREE Full text] [doi: [10.2196/jmir.2003](#)] [Medline: [22366336](#)]
3. Hong YA, Liang C, Radcliff TA, Wigfall LT, Street RL. What do patients say about doctors online? A systematic review of studies on patient online reviews. *J Med Internet Res* 2019 Apr 08;21(4):e12521 [FREE Full text] [doi: [10.2196/12521](#)] [Medline: [30958276](#)]
4. Daskivich TJ, Houman J, Fuller G, Black JT, Kim HL, Spiegel B. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. *J Am Med Inform Assoc* 2018 Apr;25(4):401-407. [doi: [10.1093/jamia/ocx083](#)] [Medline: [29025145](#)]
5. McGrath RJ, Priestley JL, Zhou Y, Culligan PJ. The validity of online patient ratings of physicians: analysis of physician peer reviews and patient ratings. *Interact J Med Res* 2018 Apr 09;7(1):e8 [FREE Full text] [doi: [10.2196/ijmr.9350](#)] [Medline: [29631992](#)]
6. Rothenfluh F, Schulz PJ. Content, quality, and assessment tools of physician-rating websites in 12 countries: quantitative analysis. *J Med Internet Res* 2018 Jun 14;20(6):e212 [FREE Full text] [doi: [10.2196/jmir.9105](#)] [Medline: [29903704](#)]
7. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. *JAMA* 2014 Feb 19;311(7):734-735. [doi: [10.1001/jama.2013.283194](#)] [Medline: [24549555](#)]
8. Burkle C, Keegan M. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. *BMC Health Serv Res* 2015 Sep 26;15:416 [FREE Full text] [doi: [10.1186/s12913-015-1099-2](#)] [Medline: [26410383](#)]
9. Blumenthal D, Abrams M, Nuzum R. The Affordable Care Act at 5 years. *N Engl J Med* 2015 Jun 18;372(25):2451-2458. [doi: [10.1056/NEJMp1503614](#)] [Medline: [25946142](#)]
10. Physician performance-based compensation. Minnetonka: UnitedHealthcare Services, Inc; 2018. URL: <https://www.uhcprovider.com/en/reports-quality-programs/physician-perf-based-comp.html?rfid=UHCOCntRD> [accessed 2020-09-30]
11. Lagu T, Norton CM, Russo LM, Priya A, Goff SL, Lindenauer PK. Reporting of patient experience data on health systems' websites and commercial physician-rating websites: mixed-methods analysis. *J Med Internet Res* 2019 Mar 27;21(3):e12007 [FREE Full text] [doi: [10.2196/12007](#)] [Medline: [30916654](#)]
12. Harder B. 2020-21 best hospitals honor roll and medical specialties rankings. 2020 Jul 28. URL: <https://health.usnews.com/health-care/best-hospitals/articles/best-hospitals-honor-roll-and-overview> [accessed 2020-09-30]
13. Daskivich T, Houman J, Fuller G, Black J, Kim HL, Spiegel BS. Online consumer ratings of physicians and quality of care, value-based care, and peer-assessed performance [in press]. *J Am Med Inform Assoc* 2017:1.
14. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. *J Med Internet Res* 2016 Dec 13;18(12):e324 [FREE Full text] [doi: [10.2196/jmir.6612](#)] [Medline: [27965191](#)]
15. zocdoc. URL: <https://www.zocdoc.com/> [accessed 2020-09-30]
16. WebMD. URL: <https://doctor.webmd.com/> [accessed 2020-09-30]
17. Vitals. URL: <https://www.vitals.com/> [accessed 2020-09-30]
18. Daskivich T, Luu M, Noah B, Fuller G, Anger J, Spiegel B. Differences in online consumer ratings of health care providers across medical, surgical, and allied health specialties: observational study of 212,933 providers. *J Med Internet Res* 2018 May 09;20(5):e176 [FREE Full text] [doi: [10.2196/jmir.9160](#)] [Medline: [29743150](#)]
19. Comrey A, Lee H. *A First Course in Factor Analysis*. Cambridge: Academic Press; 2013.
20. Williams B, Onsman A, Brown T. *Exploratory factor analysis: a five-step guide for novices*. Australas J Paramed 2010 Aug 02;8(3):1. [doi: [10.33151/ajp.8.3.93](#)]
21. Hair JA, Tatham R, Black W. *Multivariate Data Analysis: With Readings*. 4th Edition. Upper Saddle River: Prentice-Hall, Inc; 1995.

22. Bartlett M. The effect of standardization on a  $\chi^2$  approximation in factor analysis. *Biometrika* 1951 Dec;38(3-4):337-344. [doi: [10.1093/biomet/38.3-4.337](https://doi.org/10.1093/biomet/38.3-4.337)]
23. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965 Jun;30:179-185. [doi: [10.1007/BF02289447](https://doi.org/10.1007/BF02289447)] [Medline: [14306381](https://pubmed.ncbi.nlm.nih.gov/14306381/)]
24. Nunnally J. *Psychometric Theory*. 2nd Edition. Hillsdale: McGraw-Hill Publishing; 1978.
25. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res* 1981 Feb;18(1):39-50. [doi: [10.2307/3151312](https://doi.org/10.2307/3151312)]
26. Weisberg S. *Applied Linear Regression*. Hoboken: John Wiley & Sons; 2005.
27. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2018.
28. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. *J Med Internet Res* 2013;15(2):e24 [FREE Full text] [doi: [10.2196/jmir.2360](https://doi.org/10.2196/jmir.2360)] [Medline: [23372115](https://pubmed.ncbi.nlm.nih.gov/23372115/)]
29. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. *J Med Internet Res* 2013;15(8):e187 [FREE Full text] [doi: [10.2196/jmir.2702](https://doi.org/10.2196/jmir.2702)] [Medline: [23985220](https://pubmed.ncbi.nlm.nih.gov/23985220/)]
30. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. *J Gen Intern Med* 2010 Sep;25(9):942-946 [FREE Full text] [doi: [10.1007/s11606-010-1383-0](https://doi.org/10.1007/s11606-010-1383-0)] [Medline: [20464523](https://pubmed.ncbi.nlm.nih.gov/20464523/)]
31. Segal J. The role of the Internet in doctor performance rating. *Pain Physician* 2009;12(3):659-664. [Medline: [19461833](https://pubmed.ncbi.nlm.nih.gov/19461833/)]
32. Yaraghi N, Wang W, Gao GG, Agarwal R. How online quality ratings influence patients' choice of medical providers: controlled experimental survey study. *J Med Internet Res* 2018 Mar 26;20(3):e99 [FREE Full text] [doi: [10.2196/jmir.8986](https://doi.org/10.2196/jmir.8986)] [Medline: [29581091](https://pubmed.ncbi.nlm.nih.gov/29581091/)]
33. Liu JJ, Matelski JJ, Bell CM. Scope, breadth, and differences in online physician ratings related to geography, specialty, and year: observational retrospective study. *J Med Internet Res* 2018 Mar 07;20(3):e76 [FREE Full text] [doi: [10.2196/jmir.7475](https://doi.org/10.2196/jmir.7475)] [Medline: [29514775](https://pubmed.ncbi.nlm.nih.gov/29514775/)]
34. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *J Med Internet Res* 2011;13(4):e95 [FREE Full text] [doi: [10.2196/jmir.1960](https://doi.org/10.2196/jmir.1960)] [Medline: [22088924](https://pubmed.ncbi.nlm.nih.gov/22088924/)]
35. Reichheld FF. The one number you need to grow. *Harv Bus Rev* 2003 Dec;81(12):46-54. [Medline: [14712543](https://pubmed.ncbi.nlm.nih.gov/14712543/)]
36. Jepson C, Asch DA, Hershey JC, Ubel PA. In a mailed physician survey, questionnaire length had a threshold effect on response rate. *J Clin Epidemiol* 2005 Jan;58(1):103-105. [doi: [10.1016/j.jclinepi.2004.06.004](https://doi.org/10.1016/j.jclinepi.2004.06.004)] [Medline: [15649678](https://pubmed.ncbi.nlm.nih.gov/15649678/)]
37. Rothenfluh F, Germei E, Schulz PJ. Consumer decision-making based on review websites: are there differences between choosing a hotel and choosing a physician? *J Med Internet Res* 2016 Dec 16;18(6):e129 [FREE Full text] [doi: [10.2196/jmir.5580](https://doi.org/10.2196/jmir.5580)] [Medline: [27311623](https://pubmed.ncbi.nlm.nih.gov/27311623/)]
38. McLennan S. The content and nature of narrative comments on Swiss physician rating websites: analysis of 849 comments. *J Med Internet Res* 2019 Sep 30;21(9):e14336 [FREE Full text] [doi: [10.2196/14336](https://doi.org/10.2196/14336)] [Medline: [31573918](https://pubmed.ncbi.nlm.nih.gov/31573918/)]
39. Chen J, Presson A, Zhang C, Ray D, Finlayson S, Glasgow R. Online physician review websites poorly correlate to a validated metric of patient satisfaction. *J Surg Res* 2018 Jul;227:1-6. [doi: [10.1016/j.jss.2018.01.037](https://doi.org/10.1016/j.jss.2018.01.037)] [Medline: [29804840](https://pubmed.ncbi.nlm.nih.gov/29804840/)]
40. Liu C, Uffenheimer M, Nasser Y, Cohen J, Ellenhorn J. "But his Yelp reviews are awful!": analysis of general surgeons' yelp reviews. *J Med Internet Res* 2019 Apr 30;21(4):e11646 [FREE Full text] [doi: [10.2196/11646](https://doi.org/10.2196/11646)] [Medline: [31038463](https://pubmed.ncbi.nlm.nih.gov/31038463/)]
41. RateMDs.com. URL: <https://www.ratemds.com> [accessed 2020-09-30]

*Edited by G Eysenbach; submitted 12.06.18; peer-reviewed by D Newman, S Bidmon, M Tkáč; comments to author 17.08.18; revised version received 12.01.19; accepted 27.04.20; published 27.10.20.*

**Please cite as:**

Zhao HH, Luu M, Spiegel B, Daskivich TJ

*Correlation of Online Physician Rating Subscores and Association With Overall Satisfaction: Observational Study of 212,933 Providers*  
*J Med Internet Res* 2020;22(10):e11258

URL: <http://www.jmir.org/2020/10/e11258/>

doi: [10.2196/11258](https://doi.org/10.2196/11258)

PMID: [33107826](https://pubmed.ncbi.nlm.nih.gov/33107826/)

©Hanson Hanqing Zhao, Michael Luu, Brennan Spiegel, Timothy John Daskivich. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 27.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly

cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Effect of Smartphone-Based Lifestyle Coaching App on Community-Dwelling Population With Moderate Metabolic Abnormalities: Randomized Controlled Trial

So Mi Jemma Cho<sup>1\*</sup>, BA; Jung Hyun Lee<sup>2\*</sup>, MD; Jee-Seon Shim<sup>3\*</sup>, PhD; Hyungseon Yeom<sup>2</sup>, MD; Su Jin Lee<sup>4</sup>, MD; Yong Woo Jeon<sup>1</sup>, MD; Hyeon Chang Kim<sup>1,3</sup>, MD, PhD, FAHA

<sup>1</sup>Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

<sup>2</sup>Department of Medicine, the Graduate School of Yonsei University, Seoul, Republic of Korea

<sup>3</sup>Cardiovascular and Metabolic Diseases Etiology Research Center, Yonsei University College of Medicine, Seoul, Republic of Korea

<sup>4</sup>Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

\*these authors contributed equally

**Corresponding Author:**

Hyeon Chang Kim, MD, PhD, FAHA

Department of Preventive Medicine

Yonsei University College of Medicine

50-1 Yonsei-ro, Seodaemun-gu

Seoul, 03722

Republic of Korea

Phone: 82 2 2228 1873

Email: [hckim@yuhs.ac](mailto:hckim@yuhs.ac)

## Abstract

**Background:** Metabolic disorders are established precursors to cardiovascular diseases, yet they can be readily prevented with sustained lifestyle modifications.

**Objective:** We assessed the effectiveness of a smartphone-based weight management app on metabolic parameters in adults at high-risk, yet without physician diagnosis nor pharmacological treatment for metabolic syndrome, in a community setting.

**Methods:** In this 3-arm parallel-group, single-blind, randomized controlled trial, we recruited participants aged 30 to 59 years with at least 2 conditions defined by the Third Report of the National Cholesterol Education Program expert panel (abdominal obesity, high blood pressure, high triglycerides, low high-density lipoprotein cholesterol, and high fasting glucose level). Participants were randomly assigned (1:1:1) by block randomization to either the nonuser group (control), the app-based diet and exercise self-logging group (app only), or the app-based self-logging and personalized coaching from professional dietitians and exercise coordinators group (app with personalized coaching). Assessments were performed at baseline, week 6, week 12, and week 24. The primary outcome was change in systolic blood pressure (between baseline and follow-up assessments). Secondary outcomes were changes in diastolic blood pressure, body weight, body fat mass, waist circumference, homeostatic model of assessment of insulin resistance, triglyceride level, and high-density lipoprotein cholesterol level between baseline and follow-up assessments. Analysis was performed using intention-to-treat.

**Results:** Between October 28, 2017 and May 28, 2018, 160 participants participated in the baseline screening examination. Participants (129/160, 80.6%) who satisfied the eligibility criteria were assigned to control (n=41), app only (n=45), or app with personalized coaching (n=43) group. In each group, systolic blood pressure showed decreasing trends from baseline (control: mean  $-10.95$ , SD  $2.09$  mmHg; app only: mean  $-7.29$ , SD  $1.83$  mmHg; app with personalized coaching: mean  $-7.19$ , SD  $1.66$  mmHg), yet without significant difference among the groups (app only:  $P=.19$ ; app with personalized coaching:  $P=.16$ ). Instead, those in the app with personalized coaching group had greater body weight reductions (control: mean  $-0.12$ , SD  $0.30$  kg; app only: mean  $-0.35$ , SD  $0.36$  kg,  $P=.67$ ; app with personalized coaching: mean  $-0.96$ , SD  $0.37$  kg;  $P=.08$ ), specifically by body fat mass reduction (control: mean  $-0.13$ , SD  $0.34$  kg; app only: mean  $-0.64$ , SD  $0.38$  kg,  $P=.22$ ; app with personalized coaching: mean  $-0.79$ , SD  $0.38$  kg;  $P=.08$ ).



**Conclusions:** Simultaneous diet and exercise self-logging and persistent lifestyle modification coaching were ineffective in lowering systolic blood pressure but effective in losing weight and reducing body fat mass. These results warrant future implementation studies of similar models of care on a broader scale in the context of primary prevention.

**Trial Registration:** ClinicalTrials.gov NCT03300271; <http://clinicaltrials.gov/ct2/show/NCT03300271>

(*J Med Internet Res* 2020;22(10):e17435) doi:[10.2196/17435](https://doi.org/10.2196/17435)

## KEYWORDS

metabolic health; health behavior; lifestyle modification; mobile health

## Introduction

Metabolic disorders are established precursors to cardiovascular disease [1]. Previous studies have demonstrated that even at subclinical stages, persons with elevated blood pressure, blood glucose level, cholesterol level, and adiposity are at significantly higher risk of adverse cardiometabolic outcomes [2]. Despite recent health care policy changes that have expanded healthy lifestyle advocacy initiatives, a considerable proportion do not achieve the guideline-recommended metabolic profile [3-5]. Therefore, timely and persistent management of metabolic abnormalities are crucial in preventing adverse health outcomes at an individual level and conserving substantial health care costs at a national level.

Alongside the advent of new medicines, technological innovations have aided in easing accessibility to and enriching the quality of health care. They have eliminated practical barriers, thereby allowing the distribution and improvement of health care via nonconventional routes at unprecedented speeds [6]. Several features include the transmission of medical records, social media forums for open discussion, web-based interactive education programs, higher precision diagnostics, real-time status tracking, digitalized clinics (ie, telemonitoring), and prescription dispensation [7,8].

In particular, the ubiquity of mobile phone technology has incentivized industries to create smartphone-based apps for health monitoring [9]. Previous trials [10,11] have evaluated the efficacy and effectiveness of such tools in the context of secondary prevention. For example, among patients receiving cardiac rehabilitation after hospitalization for myocardial infarction, daily text message reminders led to greater medication adherence and exercise capacity compared with patients receiving usual care [10]. The utility extends to general populations; a Finnish trial [11] demonstrated both short- and long-term weight loss among people who were overweight and who logged weight daily and received dietary management instructions over a period of 1 year. By collecting data in real time, these mobile-based apps enable researchers to assess multiple behaviors and to prompt change at low cost and with high ease.

Nonetheless, there are important limitations in the current literature. Previous studies [12-14] have primarily recruited clinic patients who were already using or were exceptionally motivated to use health management tools. Methodologically, many studies [7] estimated the effect of these mobile interventions based on per protocol analysis. It has been established that physical activity and diet affect blood pressure

level not only in hypertension patients but also in people with prehypertension or elevated blood pressure [15]. Recent blood pressure guidelines [15,16] emphasized early lifestyle modification for people whose blood pressure is above the normal range. However, the effects of smartphone-based apps in lowering blood pressure have not been properly assessed in people whose blood pressure is above the normal range in real-world settings.

In this context, the objective of the study was to evaluate the longitudinal effect of smartphone-based health care app on metabolic parameters in a sample of the general population with moderate metabolic abnormalities yet without clinical diagnosis nor pharmacological treatment. We hypothesized that the participants receiving both real-time personalized coaching and self-logging diet and physical activity would yield greater improvements in blood pressure and other metabolic parameters than those who were only self-logging or who did not use the app.

## Methods

### Study Design

The study was designed as a single-blind 3-arm parallel-design randomized controlled trial delivering a 6-month primary prevention program via mobile app to a population with moderate metabolic abnormalities with neither diagnosis nor treatment for metabolic disorders. The main objective was to evaluate the effectiveness of health management app on metabolic parameters over 3 follow-up examinations. Community-dwelling adults residing in Seoul and nearby capital regions were recruited. The study protocol was approved by the institutional review boards of Severance Hospital and Yonsei University Health System (4-2017-0666), and the protocol of the study is registered at ClinicalTrials.gov (NCT03300271).

### Participants

Candidate participants were identified based on objectively measured metabolic profiles from a previous Cardiovascular and Metabolic Diseases Etiology Research Center (CMERC) observational cohort study [17]. Briefly, the CMERC study [17] aimed to identify novel risk factors and to investigate the distribution and effects of known cardiovascular and metabolic diseases risk factors.

We recruited smartphone users aged 30-59 years with at least 2 metabolic abnormalities defined by the modified version of the Third Report of the National Cholesterol Education Program expert panel on detection, evaluation, and treatment of high blood cholesterol in adults [18] using the criteria for Asian

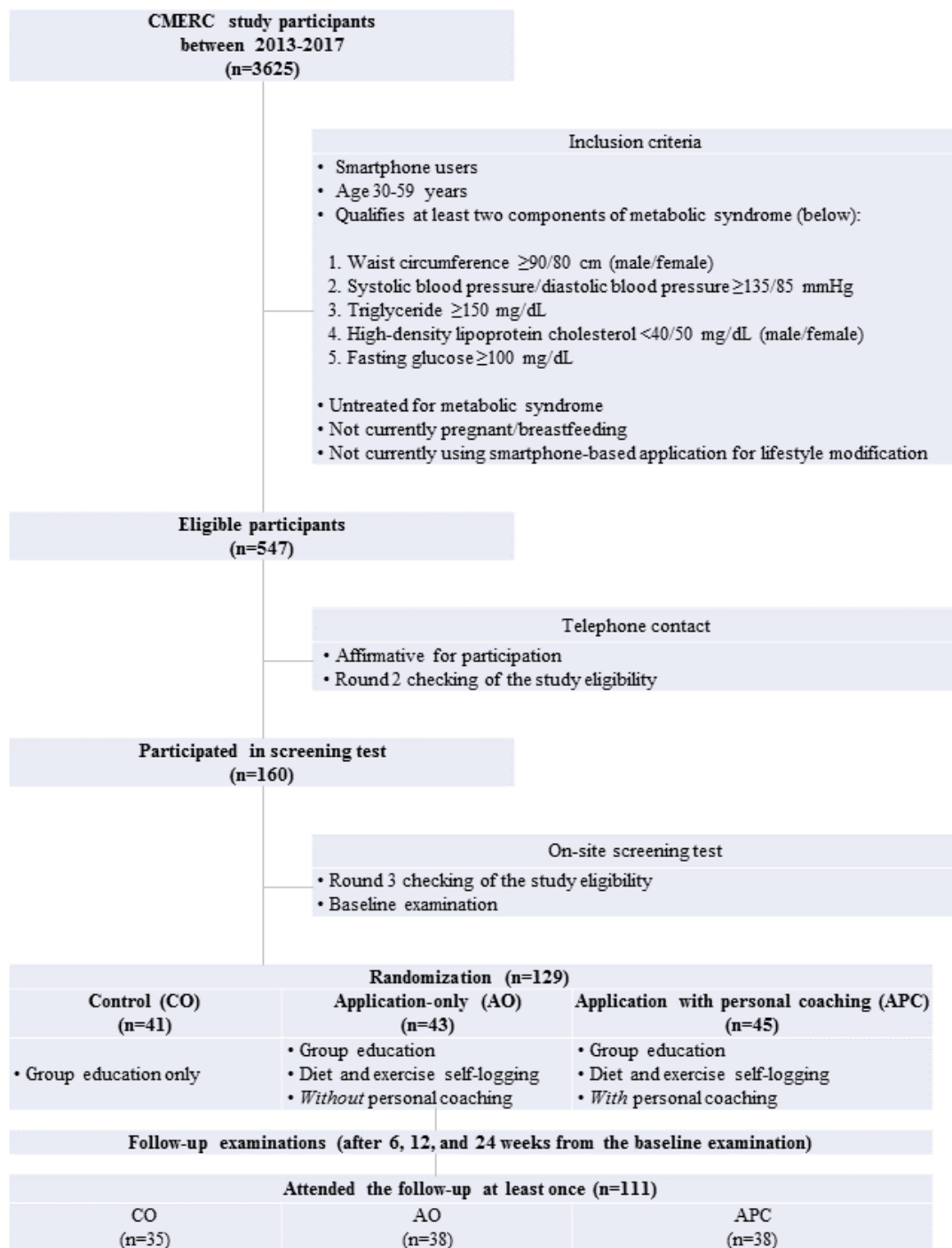
populations: waist circumference (male:  $\geq 90$  cm; female:  $\geq 80$  cm), blood pressure (systolic:  $\geq 135$  mmHg; diastolic:  $\geq 85$  mmHg), triglyceride  $\geq 150$  mg/dL, high-density lipoprotein (HDL) cholesterol (male:  $< 40$  mg/dL; female:  $< 50$  mg/dL), and fasting glucose level  $\geq 100$  mg/dL. Exclusion criteria included users of smartphone-based health care apps for lifestyle modification, individuals with a previous diagnosis of cardiovascular disease, malignant cancer, or metabolic syndrome, individuals who were taking antihypertensives, lipid-

or glucose-lowering drugs, and women who were pregnant or breastfeeding at the time of the study.

Among 3625 CMERC cohort participants [17], 546 people qualified. We contacted these individuals via telephone and mail for recruitment. Of the 160 (29.1%) who expressed affirmative for screening test, a total of 129 participants (23.6%) attended the baseline examination (Figure 1).

Participants were asked to give their written consent without knowledge of the intervention assignments.

**Figure 1.** Flow diagram of the study participants. CMERC: Cardiovascular and Metabolic Diseases Etiology Research Center.



## Randomization and Masking

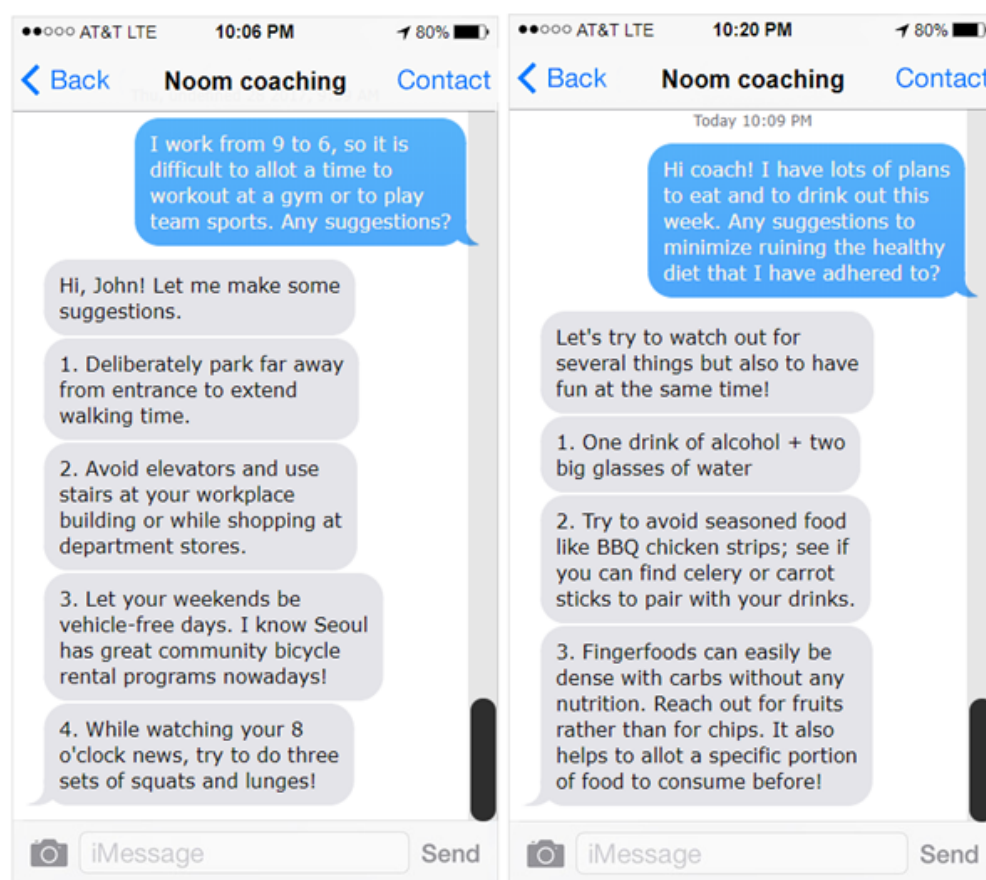
We randomly allocated participants into 1 of the 3 intervention arms via sex- and age-stratified block randomization. Specifically, we set the block size equal to 6 from sex (male or female) by age group in decile (30, 40, and 50) using R (version 3.4.4; R Foundation for Statistical Computing). We adopted a single-blind approach; thus, the effectiveness would be assessed by masked researchers unaware of the randomization results. The allocation concealment was achieved via individualized texting, which instructed all participants to adhere to their respective intervention, to avoid using other health management apps, and to refrain from sharing of their intervention instructions with each other for the entire study duration. To minimize crossover or contamination, we (1) retrospectively ensured no unallocated feature had been installed to the app only group from the app's user-specific metadata and (2) asked the participants whether they had engaged in other health-related trials, programs, or apps at every follow-up assessment.

## Procedures

At the baseline assessment, all participants received on-site education that entailed information on metabolic syndrome and preventive strategies, including validated exercise regimens and nutritious cooking recipes. Then, we randomly allocated

participants into 1 of the 3 intervention arms: the control group (control;  $n=41$ ) received only the aforementioned baseline education and was asked to refrain from concurrently engaging in any smartphone-based lifestyle modification app or programs; the app only (app only;  $n=45$ ) and the app with personalized coaching (app with personalized coaching;  $n=43$ ) groups also received the baseline education and were additionally asked to use a smartphone-based weight management app called Noom (Noom Inc). Specifically, Noom allows users to log details regarding daily food intake and physical activities. For instance, a user can record a specific menu item (ie, fastfood chain M's cheeseburger), consumed portion (ie, single serving, half of cheeseburger), and time of consumption (ie, breakfast, 8 AM). From Noom's nutrition metadata, the app readily calculates total calories and macronutrients consumed for each log entry. Likewise, a wide array of physical activities (ie, indoor treadmill walking) are available for selection from the predefined list. After entering the intensity (ie, pace) and duration (ie, 60 minutes) of physical activity, the app yields energy expenditure in kilocalories, accounting for the user's sex, age, height, and weight. As an additional feature, it delivers structured health-related curricula and personalized feedback from certified exercise regimen coordinators and clinical dieticians based on their reviews of the user's logs (Figure 2).

**Figure 2.** Example of personalized coaching.



The default frequency of personalized coaching was set to 3 times per week; however, the actual frequency varied depending

on the user's participation rate. Only the app with personalized coaching group received the personalized coaching. All users'

activation statuses were evaluated based on the number of weight, meal, and physical activity logs. Weekly, we defined each user as active if the user had recorded at least 1 aforementioned parameter.

The universal baseline examination included anthropometric measurements, blood tests, and face-to-face interviews on demographics, disease history, and health behavior. Then, participants received text messages notifying them of the study initiation. The app only and app with personalized coaching groups were given additional instructions regarding the Noom app installation procedures; each user received a unique identification code.

## Outcomes

The primary outcome was change in systolic blood pressure between the baseline and follow-up at 6, 12, and 24 weeks. The secondary outcomes were changes in diastolic blood pressure, body weight, body fat mass, waist circumference, homeostatic model of assessment of insulin resistance (HOMA-IR), and lipid profile (triglyceride and HDL cholesterol) between the baseline and follow-up at 6, 12, and 24 weeks.

Anthropometric measurements were performed with strict adherence to standardized protocols and using calibrated equipment. Blood pressure was consecutively measured using both single- and double-arm automated oscillometric devices (HEM-7080, Omron Health; WatchBP Office Central, Microlife) at a single sitting. The mean of second and third measurements were used for analysis. Participants underwent blood tests after overnight fasting for a minimum of 8 hours. Fasting plasma glucose and insulin levels were assessed using colorimetry method (ADVIA1800 Auto Analyzer, Siemens Medical Solutions). HOMA-IR was calculated as the product of fasting glucose and insulin levels divided by 405 in mg/dL. Weight was measured to the nearest 0.1 kg on a digital scale (DB-150, CAS). To minimize measurement variability, a zero-point adjustment was routinely conducted using weight blocks (20, 40, and 60 kg). Bioelectrical impedance analysis delineated body composition (BSM-330, INBODY). Waist circumference was measured to the nearest 0.1 cm using a plastic tape (SECA 201, SECA), while maintaining the level of the measuring tape.

Physical activity was assessed by the Korean version of the International Physical Activity Questionnaire [19]. Using validated alcohol consumption and cigarette smoking ratings, participants were divided into non-, previous-, and current categories for alcohol consumption and smoking.

## Statistical Analysis

The target sample size was 150 participants, chosen to provide precise estimates of the intended effect of mobile app usage. Specifically, the sample size calculation was conducted based on the expected difference in systolic blood pressure after 24 weeks across the 3 intervention arms [20]. We assumed a statistical power of 80% and a significance level of  $P < .05$ . Based on previous literature [20,21], we expected a mean systolic blood pressure difference of 6 mmHg with a standard deviation of 10 mmHg after 24 weeks. When considering 10% attrition

rate and 3-arm design, the study required a minimum recruitment of 150 participants.

We employed analysis of variance to assess differences in demographic and health-related behavior. Then, we compared the extent of changes in each metabolic parameter across participants randomly assigned to control versus app only and app with personalized coaching groups, separately, using an intention-to-treat approach; all participants who participated in at least 1 of the 3 follow-up assessment were included in the analysis. We used independent  $t$  tests to evaluate the primary and secondary outcomes at each time point. To account for repeated measurements over multiple follow-ups, we employed a linear mixed model to determine the effect of mobile health care apps on the prespecified outcomes. Specifically, the unstructured linear mixed model incorporates time and group $\times$ time interaction terms to assume no homogeneity across the 3 groups at the baseline. From the random intercept model, the estimated beta coefficient of the group $\times$ time interaction term was regarded as the effect of the app usage. The changes are presented as estimated beta coefficient ( $\beta$ ) and standard error (SE). All statistical tests were 2-sided and the statistical significance was set at a  $P < .05$ . All analyses were performed using R and SAS (version 9.4; SAS Institute Inc).

All data were collected, registered, and managed on ClinicalTrials.gov (NCT03300271). Data were deidentified and accessible only by designated researchers. All researchers strictly adhered to data security protocols. For unbiased data monitoring and trial safety overseeing, the research director delegated Dae Ryong Kang, a professor of biomedical data science at Yonsei University, Wonju College of Medicine, Wonju, Korea, who remained independent of the research execution.

## Results

### Participants

All participants were enrolled on October 28, 2017, and the last participant completed the week 24 follow-up on June 2, 2018. Of the 129 enrolled individuals, 41 (32%) were randomly assigned to the control group, 45 (35%) to the app only group, and 43 (33%) to the app with personalized coaching group (Figure 1). In the end, 111 participants attended at least 1 follow-up examination (week 6: 107; week 12: 100; week 24: 105), yielding a 14.0% attrition rate overall.

### Baseline Characteristics

Table 1 shows the general characteristics of the study participants at the baseline screening. Overall, the participants were similarly distributed in terms of age, metabolic parameters, and health behaviors across the 3 groups. At baseline, systolic blood pressure was comparable across the 3 groups (control: mean 131.8 mmHg; app only: mean 130.8 mmHg; app with personalized coaching: mean 133.3 mmHg). Such comparability across the groups ensured the baseline differences did not affect the changes in metabolic parameter over the study period.



**Table 1.** General characteristics of the study participants at the baseline screening.

Variables	Control (n=41)	App only (n=45)	App+personalized coaching (n=43)
Age (years), mean (SD)	49.5 (7.9)	49.2 (7.5)	48.9 (7.8)
<b>Sex, n (%)</b>			
Male	19 (46.3)	23 (51.1)	21 (48.8)
Female	22 (53.7)	22 (49.9)	22 (51.2)
Number of metabolic abnormalities, mean (SD)	2.9 (0.8)	2.9 (0.9)	2.8 (0.9)
Systolic blood pressure (mmHg), mean (SD)	131.8 (15.8)	130.8 (15.2)	133.3 (14.9)
Diastolic blood pressure (mmHg) , mean (SD)	87.4 (9.9)	86.6 (10.7)	89.0 (11.7)
Height (cm), mean (SD)	166.4 (8.8)	165.3 (10.4)	164.0 (8.9)
Weight (kg, mean (SD)	71.8 (13.2)	72.6 (12.4)	71.9 (11.6)
Body mass index (kg/m <sup>2</sup> ), mean (SD)	25.8 (2.8)	26.5 (3.2)	26.6 (2.9)
Waist circumference (cm) , mean (SD)	89.1 (8.2)	90.8 (8.2)	90.4 (7.0)
Percent body fat, mean (SD)	30.3 (6.8)	31.9 (7.1)	31.6 (5.7)
Fat mass (kg), mean (SD)	21.5 (5.2)	23.1 (6.4)	22.5 (4.5)
Skeletal muscle mass (kg) , mean (SD)	28.1 (7.1)	27.6 (6.2)	27.5 (6.1)
Total cholesterol (mg/dL), mean (SD)	212.9 (31.6)	197.4 (34.1)	205.1 (31.8)
Triglyceride (mg/dL), mean (SD)	212.9 (138.0)	176.6 (107.7)	169.2 (72.7)
HDL <sup>a</sup> cholesterol (mg/dL), mean (SD)	46.1 (8.4)	47.1 (11.2)	46.3 (10.5)
LDL <sup>b</sup> cholesterol (mg/dL), mean (SD)	139.1 (28.8)	127.5 (35.0)	137.1 (30.7)
Fasting blood glucose (mg/dL), mean (SD)	98.1 (13.8)	105.9 (39.4)	98.9 (19.0)
Insulin (mIU/mL), mean (SD)	11.4 (4.8)	11.5 (4.8)	11.8 (3.9)
Hemoglobin A <sub>1c</sub> (%), mean (SD)	5.9 (0.7)	6.1 (1.2)	5.8 (0.7)
HOMA-IR <sup>c</sup> , mean (SD)	2.8 (1.2)	3.0 (1.5)	2.9 (1.5)
Hypertension, mean (SD)	2 (4.9)	2 (4.4)	1 (2.3)
Diabetes, mean (SD)	0 (0.0)	0 (0.0)	0 (0.0)
Dyslipidemia, mean (SD)	3 (7.3)	1 (2.2)	1 (2.3)
<b>Smoking status, n (%)</b>			
Never	19 (46.3)	28 (62.2)	25 (58.1)
Former	8 (19.5)	11 (24.4)	11 (25.6)
Current	14 (34.2)	6 (13.3)	7 (16.3)
<b>Alcohol consumption, n (%)</b>			
Never	5 (12.2)	8 (17.8)	5 (11.6)
Former	2 (4.9)	0 (0.0)	2 (4.7)
Current	34 (82.9)	37 (82.2)	36 (83.7)
Alcohol consumption (g/day), mean (SD)	22.1 (38.8)	10.4 (15.5)	11.5 (18.1)
<b>MVPA<sup>d</sup> engagement, n (%)</b>			
Yes	2 (4.9)	8 (17.8)	9 (20.9)
No	39 (95.1)	37 (82.2)	34 (79.1)
Sedentary time (hours/day), mean (SD)	7.2 (2.7)	6.6 (2.8)	7.1 (3.9)

<sup>a</sup>HDL: high-density lipoprotein.<sup>b</sup>LDL: low-density lipoprotein.



<sup>c</sup>HOMA-IR: homeostatic model assessment of insulin resistance.

<sup>d</sup>MVPA: moderate-vigorous physical activity.

To ensure homogeneity between those who attended and those who did not attend follow-up at least once, we have compared the baseline characteristics as illustrated in [Multimedia Appendix 1](#). In regard to age, sex, anthropometry, and glycemic and lipid profiles, no statistically significant differences (age:  $P=.10$ ; sex:  $P=.09$ ; systolic blood pressure:  $P=.36$ ; diastolic blood pressure:  $P=.93$ ; BMI:  $P=.98$ ; total cholesterol:  $P=.95$ ; fasting glucose level:  $P=.25$ ) were detected between the 2 groups. The only notable difference was the proportion of current smokers; those who did not attend any follow-up examinations had a higher proportion of current smokers than those who attended at least once (44.4% versus 17.1%;  $P=.04$ ; [Multimedia Appendix 2](#)).

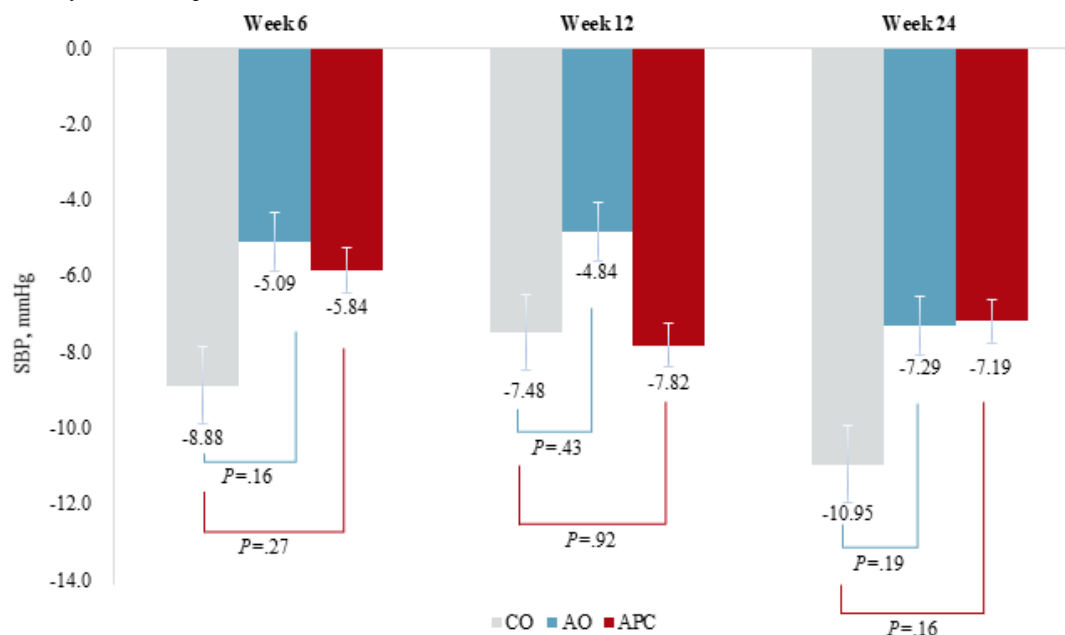
### Comparisons at Each Time Point

All 3 groups showed varying amounts of systolic blood pressure reduction overall. In reference to the baseline examination, the control and app only groups showed the most dramatic reduction at week 24 (control: mean  $-10.95$ , SD 11.98 mmHg; app only: mean  $-7.19$ , SD 9.98 mmHg), whereas the app with personalized

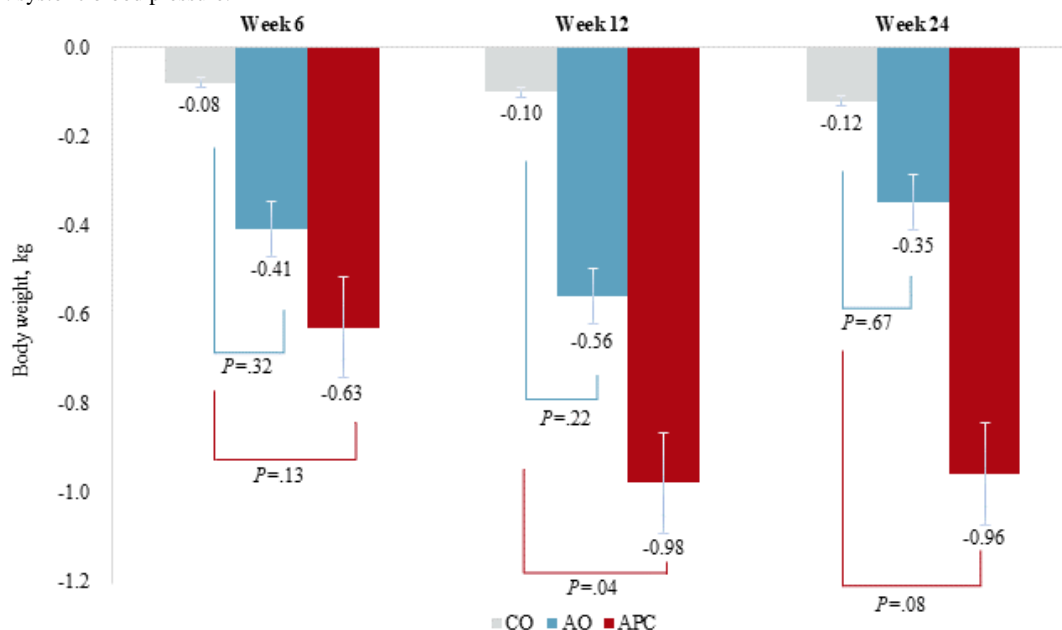
coaching group showed the most dramatic systolic blood pressure reduction at week 12 (mean  $-7.82$ , SD 11.98 mmHg). However, compared to the control group, neither the app only nor the app with personalized coaching group had significantly different systolic blood pressure change at any follow-up examination ([Figure 3](#)).

For secondary outcomes, participants in the app with personalized coaching group generally lost more body weight (mean  $-0.96$  kg versus  $-0.12$  kg; [Figure 4](#)), lost more body fat mass (mean  $-0.79$  kg versus  $0.13$  kg; [Figure 5](#)), and had smaller waist circumferences (mean  $-1.86$  cm versus  $-0.08$  cm; [Multimedia Appendix 2](#)) by the end of the study period. At week 24, diastolic blood pressure ( $P=.42$ ; [Multimedia Appendix 3](#)) and other glycemic (fasting glucose level:  $P=.99$ ; HOMA-IR:  $P=.63$ ; [Multimedia Appendix 4](#) and [Multimedia Appendix 5](#), respectively), and lipid indices (triglyceride level:  $P=.93$ ; HDL cholesterol level:  $P=.46$ ; [Multimedia Appendix 6](#) and [Multimedia Appendix 7](#), respectively) had not changed significantly.

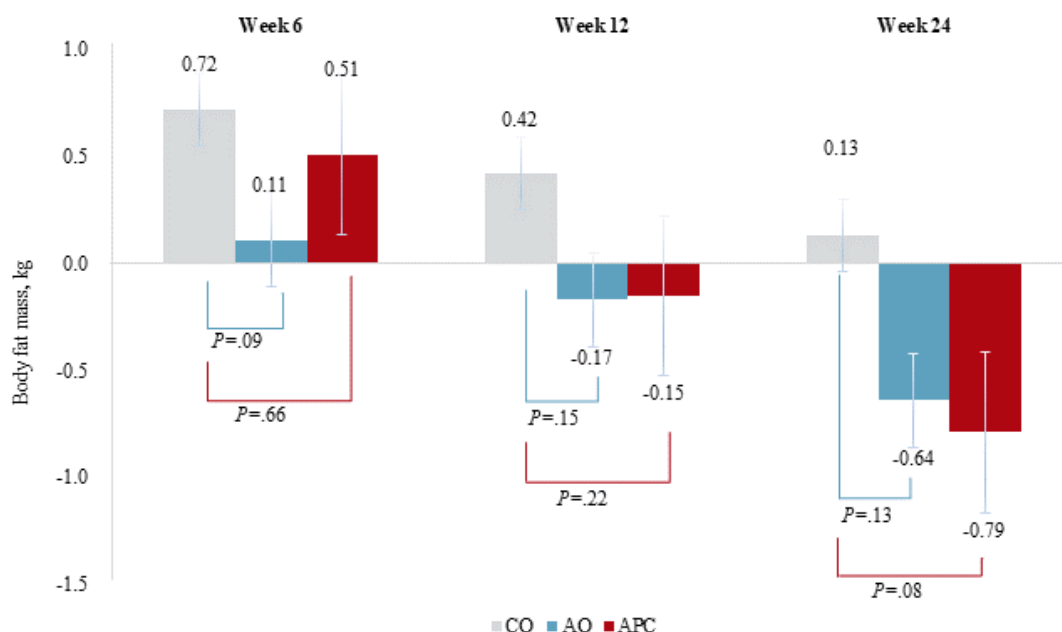
**Figure 3.** Systolic blood pressure changes from the baseline across the 3 groups at each time point. AO: app-only; APC: app with personalized coaching; CO: control, SBP: systolic blood pressure.



**Figure 4.** Body weight changes from the baseline across the 3 groups at each time point. AO: app-only; APC: app with personalized coaching; CO: control, SBP: systolic blood pressure.



**Figure 5.** Body fat mass changes from the baseline across the 3 groups at each time point. AO: app-only; APC: app with personalized coaching; CO: control.



### Linear Mixed Model

Neither the app only (control versus app only at week 24:  $\beta=2.93$ , SE 2.42,  $P=.23$ ) nor the app and additional personalized feedback feature (control versus app with personalized coaching at week 24:  $\beta=3.18$ , SE 2.42,  $P=.19$ ) significantly contributed to systolic blood pressure reduction at any of the follow-up examinations (Multimedia Appendix 1). In regard to secondary outcomes, statistically significant effects were observed for body weight and body fat. Compared to the participants in the control group, those in the app with personalized coaching group lost more body weight at week 12 ( $\beta=-0.93$ , SE 0.43;  $P=.03$ ) and at week 24 ( $\beta=-0.87$ , SE 0.42;  $P=.04$ ). This translated to an average difference of 0.14 kg of body weight loss per month

between participants in the control group and participants in the app with personalized coaching group. In parallel, the app with personalized coaching group lost more body fat than the control group did at week 24 ( $\beta=-0.95$ , SE 0.46,  $P=.04$ ). This was equivalent to participants in the app with personalized coaching group losing an average of 0.19 kg more body fat per month than those in the control group. The app with personalized coaching group showed significant HOMA-IR level reductions at week 12 ( $\beta=-0.55$ , SE 0.27,  $P=.04$ ). However, separate examination by fasting glucose level did not show significance ( $P=.80$ ) and the HOMA-IR reductions diminished by week 24 ( $\beta=-0.22$ , SE 0.26,  $P=.40$ ). Otherwise, there were no significant effects of the app on diastolic blood pressure (app only:  $P=.83$ ; app with personalized coaching group:  $P=.47$ ), waist

circumference (app only:  $P=.32$ ; app with personalized coaching:  $P=.12$ ), triglyceride (app only:  $P=.70$ ; app with personalized coaching:  $P=.23$ ), HDL cholesterol (app only:  $P=.14$ ; app with personalized coaching:  $P=.91$ ), or fasting blood glucose level (app only:  $P=.38$ ; app with personalized coaching:  $P=.84$ ; [Multimedia Appendix 1](#)).

## Discussion

### Principal Findings

Mobile-based behavioral health interventions that permit real-time data collection and sharing are increasingly commonplace, enabling researchers to assess multiple health behaviors in various contexts. At the same time, they prompt users to be more self-aware and to modify their own health behavior. In this context, the goal behind this study was to assess the real-world effectiveness of a smartphone-based self-monitoring health management app in a community-dwelling population of individuals without diagnosis or treatment of metabolic disorders. App usage showed differential effects on metabolic parameters at different time points. Overall, the primary outcome of systolic blood pressure followed a decreasing trend from the baseline yet did not change notably between the 3 groups at any follow-up examinations. Instead, the simultaneous diet/exercise logging and lifestyle coaching yielded relatively greater body weight reduction, specifically via body fat mass reduction. These effects were attenuated yet sustained at 6 month.

Despite no significant systolic blood pressure improvement in this study, previous studies [12-14,22-24] have identified the utility of mobile apps in lowering blood pressure through self-monitoring and information delivery services. In several randomized controlled trials [12-14,22-24] conducted in obesity clinic settings, text messages or emails for antihypertensive medication adherence, tailored guidance on salt intake, smoking cessation, and physical activity interventions reported significant differences in blood pressure reduction compared to usual care. However, unlike our study in which community-dwelling individuals without overt metabolic syndrome participated, similar studies [7,21,25] have been primarily performed on patients based on completers' analysis rather than intention-to-treat. Considering varied levels of initiative for health management often used as a proxy for adherence and attrition, our results indicate that effectiveness of these self-care apps may considerably differ by study population and analytical methods.

Moreover, Liu and colleagues' [25] examination of internet-based counseling interventions among patients with elevated blood pressure indicated that e-counseling interventions significantly reduced daytime systolic blood pressure by 3.8 mmHg (95% CI  $-5.63$  to  $-2.06$ ), with greater reductions from more sustained interventions. Since our study relied on a single-occasion blood pressure measurement in the examination setting, the results may potentially be distorted by whitecoat or masked hypertension.

Furthermore, a significant limitation of accumulated evidence is that many studies [7] were conducted for durations less than

6 months. Given that hypertension is a chronic condition that requires long-term pharmacological treatment and lifestyle modification, prolonged observation should be allotted for more accurate assessment of intervention adherence and subsequent blood pressure changes.

In our study, body weight reduction, via body fat mass loss, was the most evident improvement from app usage, as indicated by similar trials [26,27]. Among overweight adults, daily transmitted personalized multimedia message services providing weight control materials have proven to induce greater weight loss ( $-1.97$  kg difference, 95% CI  $-0.34$  to  $-3.60$  kg) than the non-receiving group [26]. A systematic review [25] consistently identified the association between self-monitoring or online obesity treatment programs and body size reduction via targeted advice on reduced energy intake, increased physical activity, and social support. Often, the studies incorporated the use of structured regimen, regular self-monitoring, circumstantially appropriate feedback, prompt communication, and social support [27]. Abraham et al [28] classified such intervention content, including information-motivation-behavior skills model (ie, providing information about behavior-health link), social-cognitive theory (ie, prompting barrier identification, general encouragement), control therapy (ie, self-monitoring, interactive feedback), relapse prevention and behavior sustenance, and more. Although the extent and nature of the interventions varied across studies, these elements commonly found in mobile technology interventions, altogether, appear crucial in driving greater changes.

Our study results indicated that the 6-month usage of the mobile app did not substantially improve insulin sensitivity. Yet, an analogous meta-analysis [29] demonstrated that comprehensive lifestyle modification delivered through a mobile app lowered hemoglobin A<sub>1c</sub> level by an average of 0.5% over 6 months of follow-up. Interestingly, this decrease was not accompanied by concurrent improvements in other diabetes risk factors, such as blood pressure, cholesterol levels, or adiposity [29]. However, considering that hemoglobin A<sub>1c</sub> reflects long-term fluctuations of glycemic control (unlike fasting glucose level used in our study), Liang et al [29] raised concerns regarding insufficient follow-up period. Moreover, since these trials were conducted on persons diagnosed with diabetes of varying subtypes and adherence to antihyperglycemic treatments, it would have been challenging to isolate the effectiveness of the mobile app intervention from changes in treatment behavior. Therefore, the different representation of glycemic state and participant characteristics warrant caution in assuming or denying the attribution of these technology-based intervention to other nonspecific benefits.

Over the span of 6 months, our participants did not show significantly different changes in triglyceride and HDL cholesterol levels across the 3 groups. Yet, Park and Kim [30] showed that, among gynecology and family medicine outpatients, the use of web-based diet and exercise diaries and the specialists' weekly lifestyle modification recommendations improved total ( $-12.9$  mg/dL) and low-density lipoprotein ( $-11.3$  mg/dL) cholesterol levels after 12 weeks. Again, it is hard to extrapolate from clinic setting population in which individuals

may already be utilizing health management app prior to trial entrance or may be taking medications (ie, statins) that distort the true effects of the interventions.

The greatest novelty of our study lies in its evaluation of the mobile app's effectiveness via intention-to-treat approach in nonpatient population without the concern of confounding by pharmacological treatment. Considering that the general population consists of individuals with wide-ranging levels of digital literacy and utility, willingness for lifestyle modification, and physiological and socioeconomic backgrounds, our study portrays the effectiveness of mobile-based health management app in real-world setting. Furthermore, the assessor-blinding enabled true compliance assessment in the sample regardless of individual preference for or competency with the app, thereby evaluating the app's real-world potential. In addition, because we assessed the changes in metabolic parameters repeatedly over a considerable time horizon, our results reflect the long-term trajectory of the metabolic indices, empirically. Lastly, as a 3-arm parallel design with concurrent control, we were able to discriminate whether the self-logging of diet and exercise was sufficient to induce improvements in metabolic parameter or whether additional personalized coaching was the critical element in shaping healthy behavior.

Nonetheless, several limitations warrant cautious interpretation of our study findings. During the screening test, we faced an unexpectedly greater number of participants who did not satisfy the inclusion criteria, and thus, we were unable to reach the goal sample size. During the study period, because the participants had varying attendance to each of the follow-up examination, there may have been residual heterogeneity among participants who attended follow-up examinations none, once, twice, or all follow-up examinations. Yet, those lost to follow-up were demographically similar to those who participated in the study with the exception of current smoking status. If any effects were present, we expect such attrition to have an effect on the study results toward the null. Moreover, because the participants were

selected from an already-established cohort, higher self-efficacy and proficiency may have been present. Therefore, the results may not be entirely generalizable to the general population. However, the original cohort study was purely observational; thus, our findings still represent individuals typically seeking health management without referral to physician nor pharmacological treatment. Furthermore, similar literature has suggested that baseline self-efficacy appraisals may not be entirely pertinent to practical skills or opportunities to sustain life changes [15]. In this context, we expect comparable initiatives for lifestyle modification between our study and the general population. Lastly, considering that blood pressure, body size, glycemic index, and lipid indices are reflective of chronic states, our 6-month study period may have been insufficient to witness changes. Future studies with a larger sample size and a prolonged study period may better assess the effectiveness of lifestyle modification mobile apps on the long-term trajectory of metabolic indices.

## Conclusions

Among the community-dwelling adults with moderate metabolic abnormalities without diagnosis or treatment for disorders, we examined the effect of a smartphone-based app on changes in metabolic parameters. By the end of the 6-month study period, no changes to systolic blood pressure were significant for participants who utilized both diet/exercise logging and personalized coaching compared to logging-only or nonusing groups. Instead, the self-logging and lifestyle coaching yielded greater body weight reduction via body fat mass loss. The research investigating mobile health management interventions that confer accessible and cost-effective benefits remains in its infancy, especially in general populations. Future studies focusing on comparative effectiveness using alternative study designs and on populations of various health, socioeconomic, and cultural backgrounds are needed to integrate these apps in everyday lives and clinic practice.

## Acknowledgments

This study was supported by a grant from the Korean Health Technology R & D Project, Ministry of Health & Welfare, Korea (HI13C0715). The Ministry of Health & Welfare did not have any role in the writing of the manuscript nor in the decision to submit the manuscript for publication. The authors have not been paid to write this article by any pharmaceutical company or any other agency. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## Authors' Contributions

JHL, J-SS, HY, and HCK conceived and designed the trial. J-SS and HCK developed the education program. SMJC, JHL, J-SS, HY, YWJ, and HCK helped perform the study. SMJC, JHL, J-SS, HY, YWJ, and HCK acquired the data. SMJC, JHL, and HCK analyzed and interpreted the data. SMJC and JHL drafted the manuscripts. SMJC, JHL, and HCK critically revised the manuscript. SMJC and JHL performed the statistical analysis. J-SS and HCK supervised the study. All authors approved the final version of the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Supplementary tables.

[[DOCX File, 44 KB](#) - [jmir\\_v22i10e17435\\_app1.docx](#)]

#### Multimedia Appendix 2

Diastolic blood pressure changes from the baseline across the 3 groups at each time point. AO: app-only; APC: app with personalized coaching; CO: control.

[[PNG File, 12 KB](#) - [jmir\\_v22i10e17435\\_app2.png](#)]

#### Multimedia Appendix 3

Waist circumference changes from the baseline across the 3 groups at each time point. AO: app-only; APC: app with personalized coaching; CO: control.

[[PNG File, 10 KB](#) - [jmir\\_v22i10e17435\\_app3.png](#)]

#### Multimedia Appendix 4

Fasting glucose level changes from the baseline across the 3 groups at each time point. AO: app-only; APC: app with personalized coaching; CO: control.

[[PNG File, 11 KB](#) - [jmir\\_v22i10e17435\\_app4.png](#)]

#### Multimedia Appendix 5

Homeostatic model of assessment of insulin resistance changes from the baseline across the 3 groups at each time point. AO: app-only; APC: app with personalized coaching; CO: control.

[[PNG File, 11 KB](#) - [jmir\\_v22i10e17435\\_app5.png](#)]

#### Multimedia Appendix 6

Triglyceride level changes from the baseline across the 3 groups at each time point. AO: app-only; APC: app with personalized coaching; CO: control.

[[PNG File, 10 KB](#) - [jmir\\_v22i10e17435\\_app6.png](#)]

#### Multimedia Appendix 7

High-density lipoprotein cholesterol level changes from the baseline across the 3 groups at each time point. AO: app-only; APC: app with personalized coaching; CO: control.

[[PNG File, 15 KB](#) - [jmir\\_v22i10e17435\\_app7.png](#)]

#### Multimedia Appendix 8

CONSORT-eHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 392 KB](#) - [jmir\\_v22i10e17435\\_app8.pdf](#)]

## References

1. Wilson PWF, D'Agostino RB, Parise H, Sullivan L, Meigs JB. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. *Circulation* 2005 Nov 15;112(20):3066-3072 [[FREE Full text](#)] [doi: [10.1161/CIRCULATIONAHA.105.539528](#)] [Medline: [16275870](#)]
2. Mottillo S, Filion KB, Genest J, Joseph L, Pilote L, Poirier P, et al. The Metabolic Syndrome and Cardiovascular Risk. *Journal of the American College of Cardiology* 2010 Sep;56(14):1113-1132. [doi: [10.1016/j.jacc.2010.05.034](#)]
3. Kang S, Kim S, Cho JH, Yoon C, Hwang S, Lee H, et al. Prevalence, Awareness, Treatment, and Control of Hypertension in Korea. *Sci Rep* 2019 Jul 29;9(1):10970 [[FREE Full text](#)] [doi: [10.1038/s41598-019-46965-4](#)] [Medline: [31358791](#)]
4. Kim B, Won JC, Lee JH, Kim H, Park JH, Ha KH, et al. Diabetes Fact Sheets in Korea, 2018: An Appraisal of Current Status. *Diabetes Metab J* 2019;43(4):487. [doi: [10.4093/dmj.2019.0067](#)]
5. Jeong JS, Kwon HS. Prevalence and Clinical Characteristics of Dyslipidemia in Koreans. *Endocrinol Metab (Seoul)* 2017 Mar;32(1):30-35 [[FREE Full text](#)] [doi: [10.3803/EnM.2017.32.1.30](#)] [Medline: [28345314](#)]
6. Garrett P, Brown CA, Hart-Hester S, Hamadain E, Dixon C, Pierce W, et al. Identifying barriers to the adoption of new technology in rural hospitals: a case report. *Perspect Health Inf Manag* 2006 Oct 25;3:9 [[FREE Full text](#)] [Medline: [18066367](#)]
7. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. *BMC Med Inform Decis Mak* 2012;12:67 [[FREE Full text](#)] [doi: [10.1186/1472-6947-12-67](#)] [Medline: [22781312](#)]
8. Boulous MNK, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online J Public Health Inform* 2014;5(3):229 [[FREE Full text](#)] [doi: [10.5210/ojphi.v5i3.4814](#)] [Medline: [24683442](#)]



9. Burke LE, Ma J, Azar KMJ, Bennett GG, Peterson ED, Zheng Y, American. Current Science on Consumer Use of Mobile Health for Cardiovascular Disease Prevention: A Scientific Statement From the American Heart Association. *Circulation* 2015 Sep 22;132(12):1157-1213. [doi: [10.1161/CIR.000000000000232](https://doi.org/10.1161/CIR.000000000000232)] [Medline: [26271892](https://pubmed.ncbi.nlm.nih.gov/26271892/)]
10. Pandey A, Krumme A, Patel T, Choudhry N. The Impact of Text Messaging on Medication Adherence and Exercise Among Postmyocardial Infarction Patients: Randomized Controlled Pilot Trial. *JMIR Mhealth Uhealth* 2017 Aug 03;5(8):e110 [FREE Full text] [doi: [10.2196/mhealth.7144](https://doi.org/10.2196/mhealth.7144)] [Medline: [28778843](https://pubmed.ncbi.nlm.nih.gov/28778843/)]
11. Haapala I, Barengo NC, Biggs S, Surakka L, Manninen P. Weight loss by mobile phone: a 1-year effectiveness study. *Public Health Nutr* 2009 Dec;12(12):2382-2391. [doi: [10.1017/S1368980009005230](https://doi.org/10.1017/S1368980009005230)] [Medline: [19323865](https://pubmed.ncbi.nlm.nih.gov/19323865/)]
12. Cottrell E, Chambers R, O'Connell P. Using simple telehealth in primary care to reduce blood pressure: a service evaluation. *BMJ Open* 2012;2(6) [FREE Full text] [doi: [10.1136/bmjopen-2012-001391](https://doi.org/10.1136/bmjopen-2012-001391)] [Medline: [23117563](https://pubmed.ncbi.nlm.nih.gov/23117563/)]
13. Kang S, Kim S, Cho JH, Yoon C, Hwang S, Lee H, et al. Prevalence, Awareness, Treatment, and Control of Hypertension in Korea. *Sci Rep* 2019 Jul 29;9(1):10970 [FREE Full text] [doi: [10.1038/s41598-019-46965-4](https://doi.org/10.1038/s41598-019-46965-4)] [Medline: [31358791](https://pubmed.ncbi.nlm.nih.gov/31358791/)]
14. Logan AG, Irvine MJ, McIsaac WJ, Tisler A, Rossos PG, Easty A, et al. Effect of home blood pressure telemonitoring with self-care support on uncontrolled systolic hypertension in diabetics. *Hypertension* 2012 Jul;60(1):51-57 [FREE Full text] [doi: [10.1161/HYPERTENSIONAHA.111.188409](https://doi.org/10.1161/HYPERTENSIONAHA.111.188409)] [Medline: [22615116](https://pubmed.ncbi.nlm.nih.gov/22615116/)]
15. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison HC, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension* 2017 Nov 13:e13-e115. [doi: [10.1161/HYP.0000000000000065](https://doi.org/10.1161/HYP.0000000000000065)] [Medline: [29133356](https://pubmed.ncbi.nlm.nih.gov/29133356/)]
16. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, ESC Scientific Document Group. 2018 ESC/ESH Guidelines for the management of arterial hypertension. *Eur Heart J* 2018 Sep 01;39(33):3021-3104. [doi: [10.1093/eurheartj/ehy339](https://doi.org/10.1093/eurheartj/ehy339)] [Medline: [30165516](https://pubmed.ncbi.nlm.nih.gov/30165516/)]
17. Shim J, Song BM, Lee JH, Lee SW, Park JH, Choi DP, et al. Cardiovascular and Metabolic Diseases Etiology Research Center (CMERC) cohort: study protocol and results of the first 3 years of enrollment. *Epidemiol Health* 2017;39:e2017016 [FREE Full text] [doi: [10.4178/epih.e2017016](https://doi.org/10.4178/epih.e2017016)] [Medline: [28395401](https://pubmed.ncbi.nlm.nih.gov/28395401/)]
18. National Cholesterol Education Program Expert Panel. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report. *Circulation* 2002 Dec 17;106(25):3143-3143. [doi: [10.1161/circ.106.25.3143](https://doi.org/10.1161/circ.106.25.3143)]
19. Chun MY. Validity and Reliability of Korean Version of International Physical Activity Questionnaire Short Form in the Elderly. *Korean J Fam Med* 2012;33(3):144. [doi: [10.4082/kjfm.2012.33.3.144](https://doi.org/10.4082/kjfm.2012.33.3.144)]
20. Twisk J, Bosman L, Hoekstra T, Rijnhart J, Welten M, Heymans M. Different ways to estimate treatment effects in randomised controlled trials. *Contemp Clin Trials Commun* 2018 Jun;10:80-85 [FREE Full text] [doi: [10.1016/j.conctc.2018.03.008](https://doi.org/10.1016/j.conctc.2018.03.008)] [Medline: [29696162](https://pubmed.ncbi.nlm.nih.gov/29696162/)]
21. Schippers M, Adam PCG, Smolenski DJ, Wong HTH, de Wit JBF. A meta-analysis of overall effects of weight loss interventions delivered via mobile phones and effect size differences according to delivery mode, personal contact, and intervention intensity and duration. *Obes Rev* 2017 Apr;18(4):450-459. [doi: [10.1111/obr.12492](https://doi.org/10.1111/obr.12492)] [Medline: [28187246](https://pubmed.ncbi.nlm.nih.gov/28187246/)]
22. Madsen LB, Kirkegaard P, Pedersen EB. Blood pressure control during telemonitoring of home blood pressure. A randomized controlled trial during 6 months. *Blood Press* 2008;17(2):78-86. [doi: [10.1080/08037050801915468](https://doi.org/10.1080/08037050801915468)] [Medline: [18568696](https://pubmed.ncbi.nlm.nih.gov/18568696/)]
23. Nolan RP, Liu S, Shoemaker JK, Hachinski V, Lynn H, Mikulis DJ, et al. Therapeutic benefit of internet-based lifestyle counselling for hypertension. *Can J Cardiol* 2012 May;28(3):390-396. [doi: [10.1016/j.cjca.2012.02.012](https://doi.org/10.1016/j.cjca.2012.02.012)] [Medline: [22498181](https://pubmed.ncbi.nlm.nih.gov/22498181/)]
24. Piette JD, Datwani H, Gaudioso S, Foster SM, Westphal J, Perry W, et al. Hypertension management using mobile technology and home blood pressure monitoring: results of a randomized trial in two low/middle-income countries. *Telemed J E Health* 2012 Oct;18(8):613-620 [FREE Full text] [doi: [10.1089/tmj.2011.0271](https://doi.org/10.1089/tmj.2011.0271)] [Medline: [23061642](https://pubmed.ncbi.nlm.nih.gov/23061642/)]
25. Liu S, Dunford SD, Leung YW, Brooks D, Thomas SG, Eysenbach G, et al. Reducing blood pressure with Internet-based interventions: a meta-analysis. *Can J Cardiol* 2013 May;29(5):613-621. [doi: [10.1016/j.cjca.2013.02.007](https://doi.org/10.1016/j.cjca.2013.02.007)] [Medline: [23618507](https://pubmed.ncbi.nlm.nih.gov/23618507/)]
26. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. *J Med Internet Res* 2009;11(1):e1 [FREE Full text] [doi: [10.2196/jmir.1100](https://doi.org/10.2196/jmir.1100)] [Medline: [19141433](https://pubmed.ncbi.nlm.nih.gov/19141433/)]
27. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. *Telemed J E Health* 2010 Nov;16(9):931-938 [FREE Full text] [doi: [10.1089/tmj.2010.0065](https://doi.org/10.1089/tmj.2010.0065)] [Medline: [21091286](https://pubmed.ncbi.nlm.nih.gov/21091286/)]
28. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. *Health Psychol* 2008 May;27(3):379-387. [doi: [10.1037/0278-6133.27.3.379](https://doi.org/10.1037/0278-6133.27.3.379)] [Medline: [18624603](https://pubmed.ncbi.nlm.nih.gov/18624603/)]
29. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. *Diabet Med* 2011 Apr;28(4):455-463. [doi: [10.1111/j.1464-5491.2010.03180.x](https://doi.org/10.1111/j.1464-5491.2010.03180.x)] [Medline: [21392066](https://pubmed.ncbi.nlm.nih.gov/21392066/)]

30. Park M, Kim H. Evaluation of mobile phone and Internet intervention on waist circumference and blood pressure in post-menopausal women with abdominal obesity. *Int J Med Inform* 2012 Jun;81(6):388-394. [doi: [10.1016/j.ijmedinf.2011.12.011](https://doi.org/10.1016/j.ijmedinf.2011.12.011)] [Medline: [22265810](https://pubmed.ncbi.nlm.nih.gov/22265810/)]

## Abbreviations

**CMERC:** Cardiovascular and Metabolic Diseases Etiology Research Center

**HDL:** high-density lipoprotein

**HOMA-IR:** homeostatic model of assessment of insulin resistance

*Edited by G Eysenbach; submitted 12.12.19; peer-reviewed by J Zhou, J Ismail, E Neter; comments to author 08.05.20; revised version received 05.06.20; accepted 25.06.20; published 09.10.20.*

*Please cite as:*

*Cho SMJ, Lee JH, Shim JS, Yeom H, Lee SJ, Jeon YW, Kim HC*

*Effect of Smartphone-Based Lifestyle Coaching App on Community-Dwelling Population With Moderate Metabolic Abnormalities: Randomized Controlled Trial*

*J Med Internet Res* 2020;22(10):e17435

URL: <https://www.jmir.org/2020/10/e17435>

doi: [10.2196/17435](https://doi.org/10.2196/17435)

PMID: [33034564](https://pubmed.ncbi.nlm.nih.gov/33034564/)

©So Mi Jemma Cho, Jung Hyun Lee, Jee-Seon Shim, Hyungseon Yeom, Su Jin Lee, Yong Woo Jeon, Hyeon Chang Kim. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 09.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Viewpoint

# Digital Micro Interventions for Behavioral and Mental Health Gains: Core Components and Conceptualization of Digital Micro Intervention Care

Amit Baumel<sup>1</sup>, PhD; Theresa Fleming<sup>2</sup>, PhD; Stephen M Schueller<sup>3</sup>, PhD

<sup>1</sup>University of Haifa, Haifa, Israel

<sup>2</sup>Victoria University of Wellington, Wellington, New Zealand

<sup>3</sup>University of California, Irvine, Irvine, CA, United States

**Corresponding Author:**

Amit Baumel, PhD

University of Haifa

Abba Khoushy Ave 199

Haifa, 3498838

Israel

Phone: 972 48240111

Email: [abaumel@univ.haifa.ac.il](mailto:abaumel@univ.haifa.ac.il)

## Abstract

Although many people access publicly available digital behavioral and mental health interventions, most do not invest as much effort in these interventions as hoped or intended by intervention developers, and ongoing engagement is often low. Thus, the impact of such interventions is minimized by a misalignment between intervention design and user behavior. Digital micro interventions are highly focused interventions delivered in the context of a person's daily life with little burden on the individual. We propose that these interventions have the potential to disruptively expand the reach of beneficial therapeutics by lowering the bar for entry to an intervention and the effort needed for purposeful engagement. This paper provides a conceptualization of digital micro interventions, their component parts, and principles guiding their use as building blocks of a larger therapeutic process (ie, digital micro intervention care). The model represented provides a structure that could improve the design, delivery, and research on digital micro interventions and ultimately improve behavioral and mental health care and care delivery.

(*J Med Internet Res* 2020;22(10):e20631) doi:[10.2196/20631](https://doi.org/10.2196/20631)

**KEYWORDS**

micro intervention; mental health; mhealth; eHealth; engagement; intervention; adherence; behavior change; behavioral health

## Introduction

Mental illnesses and substance use disorders constitute a major public health challenge, with large personal social and economic costs [1-3]. Evidence-based interventions exist; however, most people with mental or behavioral health issues receive no treatment, and for those who do receive care, the average duration of untreated illness is excessive [3-6]. This treatment gap is linked to psychological and social barriers, such as feelings of embarrassment, stigma, and shame [7], as well as structural barriers, such as pricing and inconvenient time or location of services [2,3,8-10].

Technology-enabled intervention delivery models have grown on the promise to increase access to care. Behavioral and mental health apps can often be accessed without requiring others to be aware of personal behavioral or mental health issues, without

physically attending appointments, and with considerably lower cost than that of face-to-face services. Self-help apps and programs, in particular, can often be accessed with no referral and be available 24/7. The reduction of barriers associated with traditional service access is important for improving access to care; however, it also brings some challenges. Those who make it to traditional services have often invested considerable effort to seek and engage with services and hence might often have high levels of motivation. For example, one study found that most outpatients surveyed prior to entering psychotherapy tended to believe they will be highly involved, invest large efforts, and make large changes in their life [11].

Models of care delivery that eliminate key barriers might reach an audience with very different characteristics, interests, and motivation than those reached by traditional services. Nowhere is this more evident than in the real-world usage of self-guided

digital behavioral and mental health interventions situated outside of traditional treatment settings. A recent study examined the real-world use of popular mental health apps and reported that after 15 days of use, the median percentage of daily active users (open rate) was 4.0%, with a median app retention rate of 3.9% [12]. These findings are congruent with researchers' own reports of poor real-world program use [13]. Subsequently, a recent systematic comparison of published reports and real-world usage of the same programs reported that users who participated in trials had four times higher eHealth program usage in comparison with real-world users of the same programs [14]. These findings imply that a high portion of those interested in digital behavioral and mental health interventions in the real world are not investing as much effort in these interventions as intended by intervention developers.

Many efforts have attempted to overcome this challenge, including efforts to address content packaging and personalization of treatment [15-17] and efforts to investigate the impact of intervention design features on engagement [18-21]. Another way of thinking, however, would be to accept that people might enter digital behavioral or mental health interventions with reduced levels of interest and motivation and to design interventions appropriate for such people. One potential modification to interventions would be to make interventions shorter and much more focused. These shorter and more focused interventions have been called "micro interventions" [22,23], which are highly focused interventions delivered in the context of a person's daily life in order to help them reach desired proximal targets. These small units of beneficial therapeutics can be delivered with little burden on the individual. Therefore, micro interventions have the potential to not only lower the bar for entry to an intervention but also the commitment and effort needed for purposeful engagement.

People commonly seek and receive micro interventions all the time. Receiving an answer to a question posted in a Facebook community about parenting could be viewed as a micro intervention. Playing an uplifting song may assist individuals with emotion regulation and could also be viewed as a micro intervention. Micro interventions are not unique to digital interventions, however, new technological affordances, such as the widespread penetration of smartphones [24,25] and the global increase in internet access [26,27], present opportunities to easily offer digital micro interventions in people's natural context. These digital micro interventions could substantially increase access to effective behavioral and mental health care by lowering the amount of effort required to reach beneficial gains.

Despite the accessibility of these affording technologies and the frequency of digital micro interventions being deployed daily, a conceptualization of digital micro interventions and how to effectively integrate them within larger therapeutic processes has not been offered. In this paper, we aim to address this gap. We begin by defining a digital micro intervention and its component parts. We then describe principles guiding the use of digital micro interventions as building blocks of a larger therapeutic process (ie, digital micro intervention care). We close by clarifying the differences between digital micro intervention care and a stepped-care approach. [Table 1](#) provides an explanatory overview of key terms described in the body of this work. Through this paper, we hope to help create a language and conceptualization that enables scholars to clarify the importance of specific digital micro interventions within larger processes and to support the creation of digital micro intervention care and the publication of studies around them.

**Table 1.** Key terms described within the paper.

Term	Definition	Example
<b>Digital micro intervention</b>	An intervention intended to achieve a highly focused objective using in-the-moment elements. These elements are not necessarily linked directly to the achievement of a larger clinical aim.	Guiding parents in several small steps to help them increase positive attention toward desired behaviors of their child.
Events	The elemental (smallest) components of digital micro interventions. Each event is an in-the-moment attempt for change or impact toward the overall target of the intervention.	An app-based gratitude exercise encouraging parents to focus on their child's positive attributes.
Decision rules	Guiding which events are deployed and when.	Including a short educational event before deploying a series of repeating in-the-moment gratitude exercises.
Proximal assessments	Assessing the impact of the event.	Short educational event might correspond to the proximal outcomes of increasing knowledge and motivation.
Overall micro intervention outcome	The target of the micro intervention which is not likely the same as the overall clinical goals.	Reaching parental sustainable positive attention toward the child (and not reduction in symptoms of the child's behavior problems).
<b>Digital micro intervention care</b>	Using digital micro interventions as building blocks of a larger therapeutic process aimed toward a target outcome.	Providing different digital micro interventions within a long process of helping parents develop emotional and social competences in their child.
Micro interventions	The building blocks of this model of care. See definition of micro intervention above.	See example for a digital micro intervention in the first row.
Conceptual model of the therapeutic process	Defines how digital micro interventions can address steps within the therapeutic process; utilized to identify the relevant digital micro interventions and the context in which they should be used.	Providing a rationale as to how emotional/social competences develop and what issues should be prioritized based on established concepts.
Therapeutic narrative as a linking bridge between interventions	A narrative presented to the user serving as a linking bridge enabling to move from one digital micro intervention to another in a way that consolidates the experience of the different interventions.	An automated prewritten text sent to parents that acknowledges their success in completing two past micro interventions; explains the rationale for the current intervention and in what ways this new intervention is meaningful.
Hub	Centralizes and links between the separate micro interventions; delivers micro interventions as well as the therapeutic narrative based on the conceptual model and an evaluation of user context/needs.	See the "One Hub to Enable Proper Digital Micro Intervention Care" section.

## Defining a Digital Micro Intervention and Its Component Parts

### Overview

Micro interventions differ from standard (ie, more extended) and brief interventions in their breadth, goals, and time frame. Standard and brief interventions are designed to achieve a broad clinical aim (eg, overcome depression and gain desired weight loss) by providing consumers with a complete package of elements [28-30]. Micro interventions, by comparison, are narrower in scope and intended to achieve a highly focused objective [22,23], using in-the-moment elements to promote emotional, cognitive, or behavioral change. These in-the-moment changes are not necessarily linked directly to the achievement of a larger clinical aim, although they may have implications or contributions to such an aim. For example, while a parent training program may aim to target reducing a child's disruptive behaviors as the overall goal (standard intervention), a digital micro intervention will target a much more focused aim such as guiding the parents in several small

steps to help them increase their positive attention toward desired behaviors of their child. Because these interventions are highly focused, in most cases, their time frame will be shorter than other interventions.

We propose that digital micro interventions are based on the following three core components: events (the individual outreaches to the target of the digital micro intervention), decision rules (guiding which events should be deployed and when), and assessments (determining the impact of the event and the intervention). In what follows, we will describe these components in detail.

### Events

Events are the elemental components of digital micro interventions. Each event is an in-the-moment attempt for change or impact toward the overall target of the intervention. A digital micro intervention may be based on one event [31,32] or multiple events [23,33]. Table 2 displays common goals of events along with associated definitions and examples. Events can be informational, such as providing didactic material in the



moment; interventional, such as attempting to spur some form of change or action in the moment like calm breathing; or supportive, such as providing some encouragement, understanding, or social support. The critical aspect is that the

end point for benefit tends to be near term. For example, an informational event should be something that can be taught quickly.

**Table 2.** Common goals of digital micro intervention events, descriptions, and related examples.

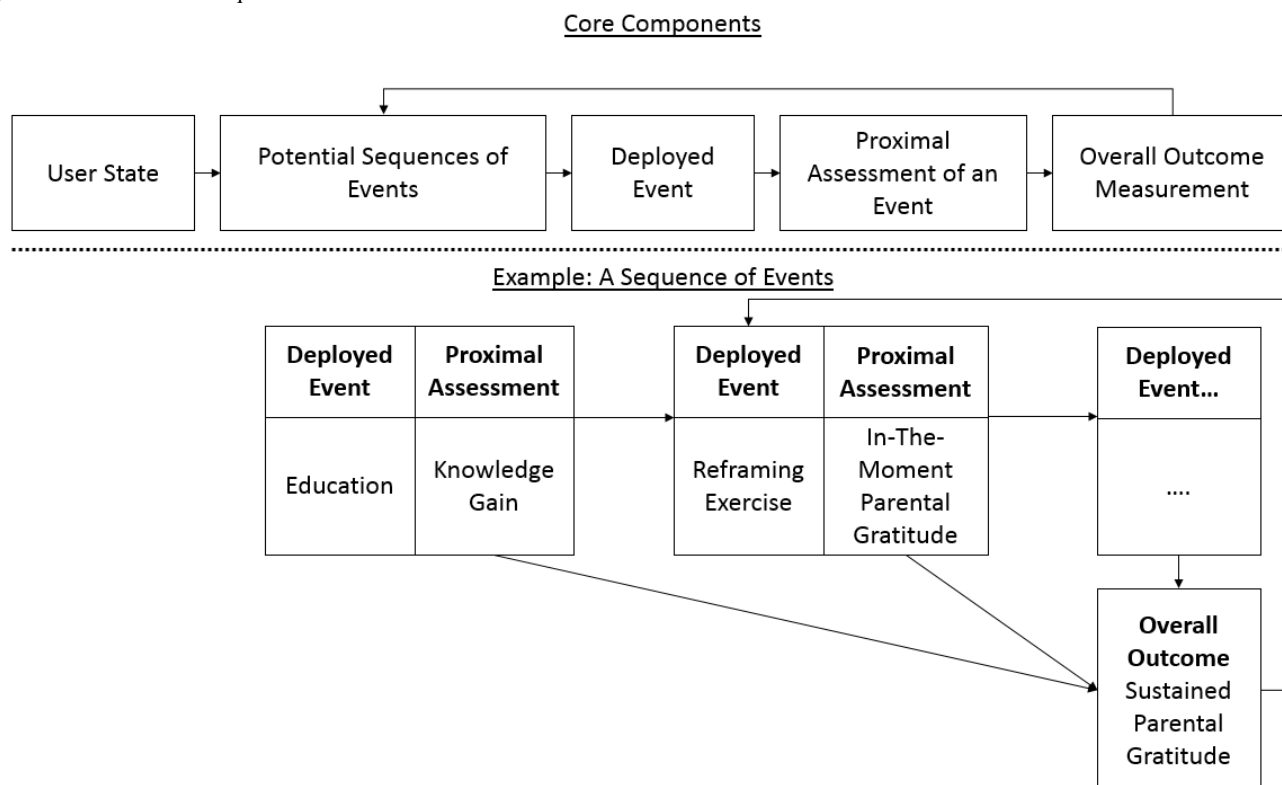
Goals/targets	Description	Example
Educational	Didactic material or psychoeducation intended to teach someone something in the moment.	A video teaching a parent how to discuss the importance of a certain behavior with a child. A timely text message motivating a person to conduct a physical activity.
Feedback	Providing information in the moment to reflect an individual's current state in a beneficial way.	Feedback about how one's current activity compares to activity at another time, such as you walked 15,000 steps today, which is 5000 more than average.
Change of perspective (eg, reframing)	Tunnel individual's focus or provoke thought processes intended to engage a person with a different perspective.	An app-based gratitude exercise encouraging an individual to focus on one's positive attributes, to increase in-the-moment satisfaction [22].
Trigger desired action	Reminder or incentive to get people to engage in a concrete action in the moment.	A notification from a wearable device suggesting that an individual stand [34].
Skill acquisition	Providing support in skill acquisition.	In-the-moment guidance on breathing exercise.
Load reduction of therapeutic-related activity	Enabling the manifestation of a beneficial activity through simplification.	In-the-moment report on calorie intake using a smartphone camera to capture a photo of food.
Symptom relief	Providing relief of negative symptoms.	An app helping a person to identify and perform competing activities to reduce in-the-moment desire for binge drinking.
Social/emotional support	Providing or indicating support from an actual or perceived other.	A text message indicating support or understanding.

## Decision Rules

Decision rules refer to specifications regarding which events should be deployed and when. They sequence and combine events in meaningful ways to create the micro intervention. When decision rules deploy events according to time-based, user-based, or environment-based information, the digital micro interventions can also be referred to as either ecological momentary interventions [35,36] or just-in-time adaptive interventions [37]. The sequence of events that follow from decision rules is closely tied to the intervention developer's understanding of the mechanistic underpinnings of how the micro intervention is perceived to operate (see Figure 1 for a suggested illustration). For example, if an intervention developer designs an intervention with the goal of achieving sustained parental gratitude, including a short educational event before a series of repeating in-the-moment exercises, the developer assumes that knowledge or understanding of the therapeutic rational is necessary prior to the interactive exercises.

This notion of a digital micro intervention as an effective sequence of events has conceptual overlap with the notion of evidence-based kernels of Embry and Bilgan [38]. In their

definition, kernels are influential elements, which constitute “fundamental units of behavior influence in the sense that deleting any component of a kernel would render it inert.” Similarly, the right combination of events is responsible for producing the digital micro intervention's outcome. As such, both the proximal assessments of events and the general outcome of the micro intervention can feed into decision rules to guide future events. Decision rules may also incorporate contextual elements, including a user's state, environmental variables, or previously identified user characteristics, to allow dynamic tailoring of the micro intervention's content [39,40]. For example, if parental gratitude is not sustained in individuals who do not achieve knowledge gain before a successful exercise event routine, it might mean that for individuals who fail in the first event (knowledge gain), there is no need to continue with the planned sequence. Alternatively, it might be that some individuals do not require a gain in knowledge prior to effective exercise events, and therefore, identifying them earlier based on their prior characteristics would enable to offer alternative paths within the same micro intervention. Indeed, contextual variables appear to be important predictors of the likelihood a person will respond to a prompt and engage with the intervention [41].

**Figure 1.** Illustration of a sequence of events based on decision rules.

## Assessments

Assessing a digital micro intervention's impact can be divided into assessing the overall outcome of the intervention and assessing the more proximal outcomes of deployed events. The overall outcome for digital micro interventions can be quite varied, although it is important to differentiate between the intervention outcome and the overall clinical goals, which are likely not the same. A digital micro intervention for individuals with serious mental illness, for example, might focus on medication adherence [42] in the hopes of reducing clinical symptoms. In this case, each event included within the intervention would focus on steps and proximal outcomes related to medication adherence. For example, supporting reasons for taking medication might correspond to the proximal outcome of increasing motivation, and providing tailored reminders for dosing periods or schedules might correspond to the proximal outcome of decreasing the cognitive effort needed to remember to initiate the desired activity. Assessing the overall benefit of this digital micro intervention, therefore, should focus not on how the intervention results in an overall clinical goal (eg, reduced symptoms of serious mental illness), but rather on how it impacts medication adherence. Importantly, if we examine the impact of events, we can explore how much each event contributes to the proximal outcome as well as the overall digital micro intervention goal.

Another aspect that relates to assessments is the user's engagement with the micro intervention. In digital micro interventions, user engagement can be documented continuously based on program usage and experience of engagement (ie, measuring the quality of attention, involvement, and immersion during program usage) [43-46]. These data could be key in

identifying not only those who properly engage with the intervention, but also those who disengage without getting all they need from it, in order to feed intervention decision rules to re-engage the users.

In this section, we have defined a digital micro intervention and presented its component parts. Subsequently, we address aspects that relate to the provision of digital micro intervention care. These aspects require us to take into consideration principles guiding the use of digital micro interventions as building blocks of a larger therapeutic process.

## Digital Micro Intervention Care: Digital Micro Interventions as Building Blocks of a Larger Therapeutic Process

### Overview

As described above, typically, individual micro interventions have relatively specific targets, rather than being full standalone treatments. Therefore, when developers think of a digital micro intervention within a larger therapeutic process, they have to determine what makes the intervention relevant or helpful, which additional digital micro interventions are expected or may be required over time, and how to integrate between the separate interventions within one therapeutic framework. In this section, we aim to address principles that support such a successful integration.

### A Conceptual Model to Guide Which Digital Micro Interventions are Required and in What Context

In order to identify the relevant digital micro interventions and the context in which they should be used, developers need a

conceptual model of the therapeutic process and how digital micro interventions can address steps within this process. As an illustration, consider the example of providing different digital micro interventions within a long process of helping parents develop emotional and social competences in their child. In order to identify the required interventions, it is first necessary to build a rationale as to how these competences develop and are maintained, and what issues should be prioritized. [Table 3](#) presents such an example. The table is based on social learning and behavioral theories [47,48], which build on concepts such

as the parenting pyramid offered in the Incredible Years program [49], and parental monitoring and prevention of child behavior problems [47]. According to the model portrayed in the table, there are levels of parenting behaviors (setting prioritization between basic and more advanced behaviors) and benefits for the child that occur when parents exercise these behaviors. The desired parenting behaviors are then used to identify the relevant digital micro interventions for each level (see examples in [Table 3](#)).

**Table 3.** An example for a model defining the capabilities parents are required to present in order to prevent behavior problems and promote social and emotional competence.

Priority level	Goal	Parent behavior	Examples for relevant digital micro interventions	Child gains
First level	Parental availability	Presence; attention; monitoring	Notifying the parents to put their smartphone away when doing activities with their child; triggering to leave the child a positive note on the kitchen table to increase presence during working days.	Attachment; self-esteem; cooperation
First level	Positive parenting practices	Positive involvement; positive modelling; conversations; play	Teaching how to play with a 4-year-old child through scenario-based learning.	Attachment; self-esteem; cooperation
Second level	Problem solving/prevention	Social coaching; proactive identification of relevant guidelines; consultations	Brief online video guidance on how to coach a 3-year-old child during play on using their mouth instead of their hands.	Social skills; meeting their potential; motivation; accountability
Third level	Dealing with acute negative symptoms	Neglecting unhelpful routines and parenting styles; embracing beneficial practices	Connecting the parent with a peer through an online community in order to find the right consequence for a child's misbehavior.	Back to normative developmental cycle; illness prevention

The prioritization portrayed in the levels also presents the context in which each intervention is mostly relevant. If a digital micro intervention care developer builds on this specific conceptualization, it would not make sense from their view to train parents to set clear limits before they are, for example, present in their children's life. As such, the incorporation of a model helps developers avoid designing and presenting people with a specific micro intervention that is not contextually relevant with the larger perspective in mind.

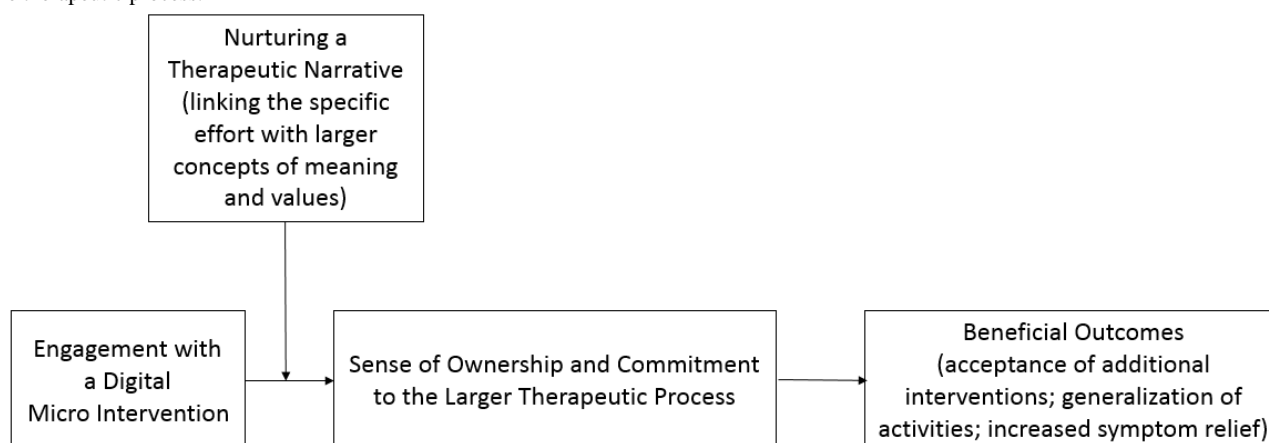
### Nurturing a Therapeutic Narrative to Link Between Independent Interventions

The literature suggests the importance of a clear therapeutic pathway and rationale to the user's therapeutic process and that relatability factors embedded within a digital intervention support a working alliance [50-53]. A theme can serve as a linking bridge enabling to move from one digital micro intervention to another in a way that consolidates between the different interventions. It is also essential because the right narrative would make the intervention more contextually relevant to the specific process the consumer is going through. For example, a parent who learned and conducted a gratitude exercise with the family around the dinner table may benefit if

it were connected to the theme of overcoming a child's noncompliance, when this digital micro intervention is part of the general aim of treating a child's disruptive behaviors. A different narrative may be in place for the same gratitude-invoking intervention, when the larger goal is different, for example, increasing parenting satisfaction among veterans with posttraumatic stress disorder [54,55]. Another simple example would be a digital micro intervention teaching individuals to change or reframe negative and unhelpful thoughts, which could be used for a broad range of clinical problems [56-58]. Naturally, the binding therapeutic narrative between different digital micro interventions would be different based on the targeted clinical problem.

Because the general process is the result of several digital micro interventions combined together, a therapeutic narrative may also have an important advantage as a moderator of the connection between each intervention and the sense of ownership and commitment toward the general process (see [Figure 2](#) for a suggested illustration). This sense of ownership and commitment may, in turn, lead to beneficial outcomes such as an increase in motivation to invest in subsequent digital micro interventions in the same domain.

**Figure 2.** Nurturing a therapeutic narrative as a moderator to the connection between a digital micro intervention and sense of ownership and commitment to the therapeutic process.



### One Hub to Enable Proper Digital Micro Intervention Care

The setting of multiple digital micro interventions within one therapeutic process requires a hub component in order to centralize and link between the separate interventions based on the principles discussed above. The hub is aimed at instilling a sense of accountability in the client [59] and at creating an integrated experience that relates to the therapeutic process, while taking an individual's history into account. The hub meets these aims by (1) recognizing an individual's state and context, (2) recommending interventions that are relevant, and (3) helping create and maintain the right narrative, meaning, and values that derive from each intervention and linking it to the larger therapeutic process.

The hub function could be performed by any entity that is set to meet these aims (a digital application, a human technology coach, a consumer who self-manages his or her state, or a psychotherapist). For example, in their study of the IntelliCare suite of apps aimed at treating depression and anxiety, Mohr and colleagues used a digital app as a hub that consolidated recommendations for the use of new apps [60]. Recommendations were based on a user's current usage data of different apps within the IntelliCare suite to identify apps that the person will most likely use and find useful. Eventually, 95% of participants downloaded five or more IntelliCare apps as part of their therapeutic process, demonstrating both the technical feasibility and acceptability of this approach [60]. In another study, health technology coaches met occasionally with patients diagnosed with schizophrenia and suggested different digital tools based on individuals' rehabilitation states and needs, and the impact of the tools they were using over time [61]. We could eventually see a new form of care, where a clinician (eg, a psychotherapist and psychiatrist) acts as a facilitator and meets with patients to suggest personalized digital micro interventions, including apps and online community support.

Figure 3 presents a visualization of the relationship between the core components of digital micro intervention care and aims to represent how the facilitator (digital app, health technology coach, clinician, etc) integrates different interventions to curate one whole therapeutic process. The key components are as follows:

(1) *The conceptual model*: Defines which digital micro interventions are required, the relevant therapeutic narratives that link between different interventions, and the need states that also determine the evaluation scheme.

(2) *Digital micro intervention planner/identifier*: Encompasses the process of identifying the relevant interventions that are currently available and planning new interventions. The required interventions are defined based on the targeted need states determined within the conceptual model (see Table 3 for an example).

(3) *Digital micro intervention suite*: Includes all the potential interventions that are available. Once a need state is identified, the relevant intervention will be offered to the individual.

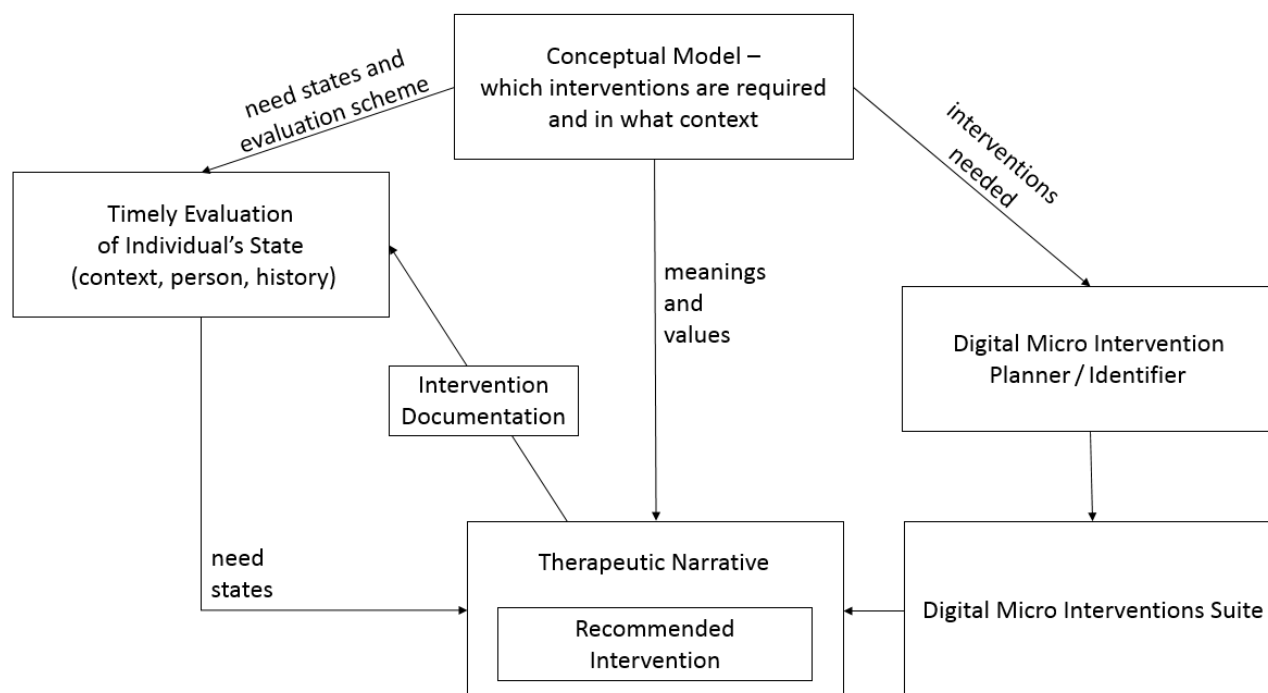
(4) *Timely evaluation of an individual's state*: The evaluation scheme is based on data that relates to the individual (eg, symptoms and demographics), context (eg, changes in the environment that may change an individual's needs), and details about past interventions and their outcomes. Need states can be alerted in two different ways. First, predefined rules can be made to trigger an evaluation based on a fixed time (eg, every 3 months) or context (eg, on a weekend when an individual is presumed to be more free or weekday when an individual is presumed to be working). Second, an individual can decide to report on his/her condition or the needs encountered, which will activate an evaluation.

(5) *Intervention phase*: The intervention phase is based on two different components. The basic component is the digital micro intervention itself. The second component represents the therapeutic narrative. The latter involves the following: (1) explaining the rationale for recommending the current micro intervention to the consumer, (2) clarifying how it relates to the general therapeutic process (if relevant based on the conceptual model), (3) integrating the current and past experiences (eg, success/failure in a previous intervention), and (4) relating to the future (eg, at the end of the intervention [what is going to happen next and why]) in relevant cases. Multimedia Appendix 1 presents an example of an automated therapeutic narrative that could be formed based on these subject matters.

(6) *Intervention documentation*: Summary data relating to engagement and outcomes are documented and used to feed the

timely evaluation of an individual's state. For example, if an individual disengages from an intervention without a clear change in symptoms and context, a different micro intervention can be offered.

**Figure 3.** Visualization of the relationship between core components of digital micro intervention care.



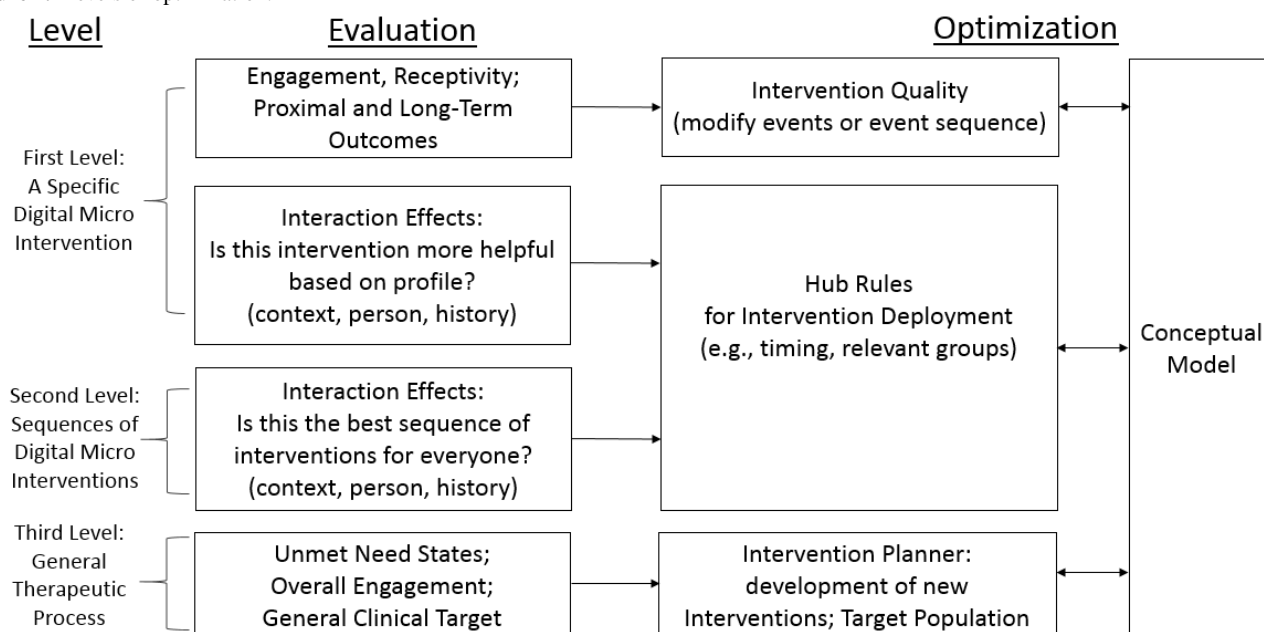
## Ongoing Optimization

The success of the therapeutic process is limited by each of the units (micro interventions) within it and the ways they are linked and interact with each other. An iterative approach may enable the identification of those interventions that are missing or less effective, or a need to refine the conceptual model, for ongoing development of the intervention ecosystem. This paper's focus is not on the methodological aspects of intervention optimization strategies; however, we refer the reader to the Multiphase Optimization Strategy as a good example for a methodological framework aimed at intervention optimization [62]. We find this framework to be relevant to digital micro intervention optimization because it is aimed at identifying the core sequence of elements that are efficacious, while providing a feasible method to account for potential intervening variables.

Figure 4 presents the different areas of evaluation that could be expected in order to build on past knowledge to optimize the system of care based on several levels of examination. The first level of optimization focuses on a specific digital micro intervention. The receptivity and impact of the intervention is measured by user engagement, proximal events, and long-term outcomes. These results lead to optimization of the

intervention's quality, and may raise subsequent questions about the intervention's impact that will require to add additional or alternative events or assessments. A subsequent evaluation involves the identification of specific populations and past experiences for whom/which the intervention was more or less successful (an interaction effect). The second level of optimization refers to the intervention sequence. Similar to the interaction effect examined in one digital micro intervention's context, it involves identifying the best sequence of *interventions* that may differ based on the client's context. The third and last level refers to the therapeutic process. It aims at understanding whether there are new need states that are not yet covered within the system, whether the clinical target is achieved, and in which cases. For example, if there are individuals who are properly engaged with the recommended sequence of micro interventions but do not reach the clinical target or individuals not engaged with any micro interventions that were offered. This will lead to optimizing the intervention planning and characterizing the user population for whom this system of care is relevant, and to the development of new interventions. The conceptual model is relevant in all levels of optimization because developers' understanding of what works and for whom enables refinement of the model itself, which is then used to optimize the interventions and their sequence.

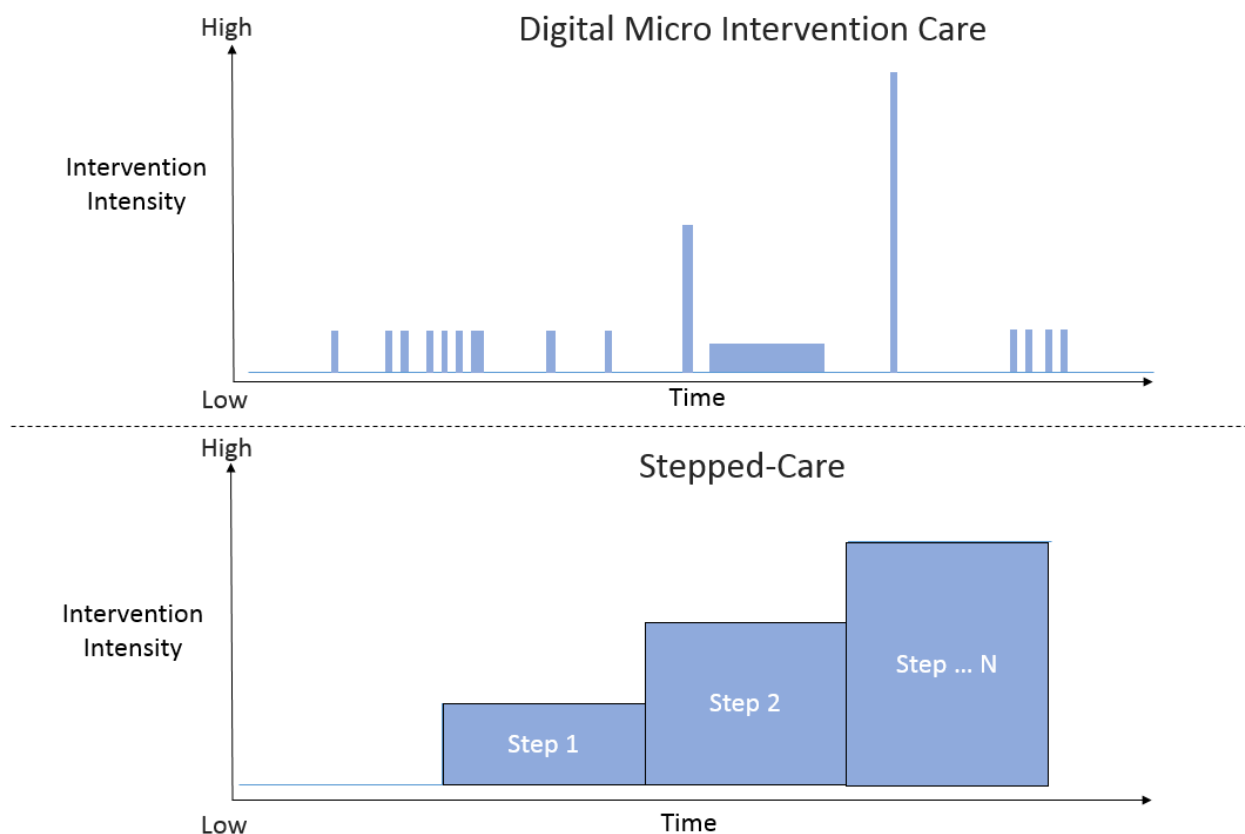


**Figure 4.** Levels of optimization.

### Example: Digital Micro Intervention Care Versus a Stepped-Care Approach

In this final section of the paper, we compare the concept of digital micro intervention care to a stepped-care approach. Stepped care is a common conceptualization that combines a set of interventions within one therapeutic process [63-65]. A stepped care approach is defined by two main principles. First, the least intensive/restrictive treatment that is expected to provide the desired health gain is offered to the patient. Second, the model is adaptive in the sense that those individuals who do not reach a desired health gain in a predetermined time period will be assigned to a more intensive/restrictive treatment option. Therefore, a stepped-care approach is based on treatments designed to reach the desired health target, where these treatments differ in intensity. In contrast, a single digital micro intervention is not expected to support a full therapeutic target (eg, overcoming depression) but rather be a step along the way. This is reasonable given that when conceptualizing the use of multiple digital micro interventions, we do not assume that people are willing to invest as much effort as ideally needed to complete a whole target of therapeutic processes.

Figure 5 presents illustrations of the two frameworks, based on these underlying differences in delivery of care. In comparison to a stepped-care model, the following aspects are noted in digital micro intervention care: (1) Each intervention unit is highly focused and is generally short, reducing the effort needed for completion; (2) The setting of care does not prioritize intervention sequence based on treatment intensity, but rather based on a complex understanding of people's needs and preferences (as described in the conceptual model section). Therefore, we do not see a graded intervention sequence in terms of intensity; (3) The therapeutic sequence does not focus on a given time window in which an intervention occurs continuously until completion, as would be expected in traditional models of proactive treatment delivery. This means that the therapeutic sequence may be very long and that individuals are sometimes engaged and sometimes not engaged with any intervention. In some cases, people do not receive any intervention most of the time. Maintaining ongoing contact with users may be required to allow triggering of the right intervention at the right time; (4) There might be a large number of different digital micro interventions to achieve the target of the treatment process, in comparison to stepped care models that are mostly based on few interventions (although each standard intervention includes multiple components).

**Figure 5.** Illustrations of stepped care and digital micro intervention care models.

## Conclusions

Digital micro interventions may provide new ways to deliver behavioral and mental health interventions while meeting people's capacity to invest a small effort in beneficial therapeutics. The development of digital micro interventions, however, demands new thinking. We cannot merely take previous interventions based on other models and assume they will have the focused and immediate impacts required of micro interventions. As such, we have proposed a novel model of understanding digital micro interventions by identifying their component parts (events), how these parts work together (decision rules), and how digital micro interventions might be incorporated into a larger therapeutic process. This model is helpful to assist researchers in optimizing digital micro interventions, developing a conceptual understanding of each intervention's role within a larger process, and leaning on principles to best integrate between these interventions over time.

It is worth noting, however, that multiple issues related to digital micro interventions still need to be resolved. First, research methodologies that can facilitate the evaluation of micro interventions are required. Micro randomized trials are one such methodology that could be of use here, especially when research questions are related to dynamic processes of change over time [66]. Second, methods for optimizing digital micro interventions

and improving micro intervention suites need to be explored. Third, we need to understand the people and targets for which the delivery of digital micro intervention care is more or less suitable.

Much of the current innovation in micro interventions has made use of technological affordances as technology makes it more feasible to deliver digital micro interventions in various ways. However, again, it does not necessarily mean that this sort of intervention has to be digital by nature. Because technology is highly embedded in our lives, it is possible that digital capabilities such as automated monitoring, triggering, and instant messaging tools will be commonly integrated into care, making the differentiation between digital and nondigital tools less relevant.

Expanding the portfolio of behavioral and mental health interventions to include interventions that allow people to interact in various modalities, intensities, and styles is likely necessary to make such interventions available to the vast number of people who could stand to benefit from them. Digital micro interventions represent one example of such an expansion, as they are brief, focused, and potentially more appropriate for people who enter with various levels of motivation. We hope the model represented here will structure and speed the work in this area and improve the research and delivery of digital micro intervention care.

## Conflicts of Interest

SMS has received payment for consulting from Otsuka Pharmaceutical. TF is a co-developer of SPARX computerized cognitive behavioral therapy for adolescent depression. The IP for SPARX is owned by Uniservices at The University of Auckland. Developers can benefit financially from licensing or sales of SPARX outside of New Zealand. AB declares no conflict of interest.

## Multimedia Appendix 1

An example for an automated therapeutic narrative.

[[PDF File \(Adobe PDF File\), 177 KB - jmir\\_v22i10e20631\\_app1.pdf](#)]

## References

1. The World Health Report 2001 - Mental Health: New Understanding, New Hope. Geneva, Switzerland: World Health Organization; 2001.
2. Neuroscience of Psychoactive Substance Use and Dependence. Geneva, Switzerland: World Health Organization; 2004.
3. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, WHO World Mental Health Survey Consortium. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA* 2004 Jun 02;291(21):2581-2590. [doi: [10.1001/jama.291.21.2581](#)] [Medline: [15173149](#)]
4. Kessler RC, Berglund PA, Bruce ML, Koch JR, Laska EM, Leaf PJ, et al. The prevalence and correlates of untreated serious mental illness. *Health Serv Res* 2001 Dec;36(6 Pt 1):987-1007 [FREE Full text] [Medline: [11775672](#)]
5. Kane JM, Robinson DG, Schooler NR, Mueser KT, Penn DL, Rosenheck RA, et al. Comprehensive Versus Usual Community Care for First-Episode Psychosis: 2-Year Outcomes From the NIMH RAISE Early Treatment Program. *Am J Psychiatry* 2016 Apr 01;173(4):362-372 [FREE Full text] [doi: [10.1176/appi.ajp.2015.15050632](#)] [Medline: [26481174](#)]
6. Wang PS, Berglund P, Olfson M, Pincus HA, Wells KB, Kessler RC. Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005 Jun 01;62(6):603-613. [doi: [10.1001/archpsyc.62.6.603](#)] [Medline: [15939838](#)]
7. Clement S, Brohan E, Jeffery D, Henderson C, Hatch SL, Thornicroft G. Development and psychometric properties the Barriers to Access to Care Evaluation scale (BACE) related to people with mental ill health. *BMC Psychiatry* 2012 Jun 20;12(1). [doi: [10.1186/1471-244x-12-36](#)]
8. Mohr D, Ho J, Duffecy J, Baron K, Lehman K, Jin L, et al. Perceived barriers to psychological treatments and their relationship to depression. *J Clin Psychol* 2010 Apr;66(4):394-409 [FREE Full text] [doi: [10.1002/jclp.20659](#)] [Medline: [20127795](#)]
9. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. *Ann Behav Med* 2006 Dec;32(3):254-258. [doi: [10.1207/s15324796abm3203\\_12](#)] [Medline: [17107299](#)]
10. Kazdin AE. Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. *Behav Res Ther* 2017 Jan;88:7-18. [doi: [10.1016/j.brat.2016.06.004](#)] [Medline: [28110678](#)]
11. Ogrodniczuk J, Kealy D, Laverdière O, Joyce A. So, who wants to be here? A survey of patients' motives for seeking psychotherapy services and their expected un-involvement in therapy. *Arch Psych Psych* 2018;20(3):26-30. [doi: [10.12740/app/92165](#)]
12. Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. *J Med Internet Res* 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: [10.2196/14567](#)] [Medline: [31573916](#)]
13. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety. *J Med Internet Res* 2018 Jun 06;20(6):e199 [FREE Full text] [doi: [10.2196/jmir.9275](#)] [Medline: [29875089](#)]
14. Baumel A, Edan S, Kane J. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. *Transl Behav Med* 2019 Nov 25;9(6):1020-1033. [doi: [10.1093/tbm/ibz147](#)] [Medline: [31689344](#)]
15. Fernando LM, Sim WH, Jorm AF, Rapee R, Lawrence KA, Yap MB. Parenting Resilient Kids (PaRK), an online parenting program to prevent anxiety and depression problems in primary school-aged children: Study protocol for a randomised controlled trial. *Trials* 2018 Apr 19;19(1):236 [FREE Full text] [doi: [10.1186/s13063-018-2605-8](#)] [Medline: [29673391](#)]
16. Norcross JC, Wampold BE. What works for whom: Tailoring psychotherapy to the person. *J Clin Psychol* 2011 Feb 24;67(2):127-132. [doi: [10.1002/jclp.20764](#)] [Medline: [21108312](#)]
17. Andrews G, Williams AD. Internet psychotherapy and the future of personalized treatment. *Depress Anxiety* 2014 Nov 18;31(11):912-915. [doi: [10.1002/da.22302](#)] [Medline: [25407580](#)]
18. Baumel A, Kane JM. Examining Predictors of Real-World User Engagement with Self-Guided eHealth Interventions: Analysis of Mobile Apps and Websites Using a Novel Dataset. *J Med Internet Res* 2018 Dec 14;20(12):e11491 [FREE Full text] [doi: [10.2196/11491](#)] [Medline: [30552077](#)]

19. Baumel A, Yom-Tov E. Predicting user adherence to behavioral eHealth interventions in the real world: examining which aspects of intervention design matter most. *Transl Behav Med* 2018 Sep 08;8(5):793-798. [doi: [10.1093/tbm/ibx037](https://doi.org/10.1093/tbm/ibx037)] [Medline: [29471424](https://pubmed.ncbi.nlm.nih.gov/29471424/)]
20. Fogg B. A behavior model for persuasive design. In: *Proceedings of the 4th international Conference on Persuasive Technology*. 2009 Presented at: 4th International Conference on Persuasive Technology; April 2009; California, USA p. 1-7. [doi: [10.1145/1541948.1541999](https://doi.org/10.1145/1541948.1541999)]
21. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. *J Med Internet Res* 2012 Nov 14;14(6):e152 [FREE Full text] [doi: [10.2196/jmir.2104](https://doi.org/10.2196/jmir.2104)] [Medline: [23151820](https://pubmed.ncbi.nlm.nih.gov/23151820/)]
22. Fuller-Tyszkiewicz M, Richardson B, Lewis V, Linardon J, Mills J, Juknaitis K, et al. A randomized trial exploring mindfulness and gratitude exercises as eHealth-based micro-interventions for improving body satisfaction. *Computers in Human Behavior* 2019 Jun;95:58-65. [doi: [10.1016/j.chb.2019.01.028](https://doi.org/10.1016/j.chb.2019.01.028)]
23. Meinschmidt G, Lee J, Stalujanis E, Belardi A, Oh M, Jung EK, et al. Smartphone-Based Psychotherapeutic Micro-Interventions to Improve Mood in a Real-World Setting. *Front Psychol* 2016 Jul 28;7:1112 [FREE Full text] [doi: [10.3389/fpsyg.2016.01112](https://doi.org/10.3389/fpsyg.2016.01112)] [Medline: [27516747](https://pubmed.ncbi.nlm.nih.gov/27516747/)]
24. Kakihara M. Grasping a Global View of Smartphone Diffusion: An Analysis from a Global Smartphone Study. 2014 Presented at: 2014 International Conference on Mobile Business; 2014; London, UK.
25. Mobile Fact Sheet. Pew Research Center. URL: <http://www.pewinternet.org/fact-sheet/mobile/> [accessed 2019-01-07]
26. Bezerra J, Bock W, Candelon F, Chai S, Choi E, Corwin J, et al. The mobile revolution: How mobile technologies drive a trillion-dollar impact. Boston, MA: Boston Consulting Group; 2015.
27. Rainie L. Internet, broadband, and cell phone statistics. Pew Research Center. 2010. URL: <https://www.pewresearch.org/internet/2010/01/05/internet-broadband-and-cell-phone-statistics/> [accessed 2020-10-16]
28. Hopko DR, Armento ME, Robertson SM, Ryba MM, Carvalho JP, Colman LK, et al. Brief behavioral activation and problem-solving therapy for depressed breast cancer patients: randomized trial. *J Consult Clin Psychol* 2011 Dec;79(6):834-849. [doi: [10.1037/a0025450](https://doi.org/10.1037/a0025450)] [Medline: [21988544](https://pubmed.ncbi.nlm.nih.gov/21988544/)]
29. Coolican J, Smith I, Bryson S. Brief parent training in pivotal response treatment for preschoolers with autism. *J Child Psychol Psychiatry* 2010 Dec;51(12):1321-1330. [doi: [10.1111/j.1469-7610.2010.02326.x](https://doi.org/10.1111/j.1469-7610.2010.02326.x)] [Medline: [21073457](https://pubmed.ncbi.nlm.nih.gov/21073457/)]
30. Hetrick S, Cox G, Witt K, Bir J, Merry S. Cognitive behavioural therapy (CBT), third-wave CBT and interpersonal therapy (IPT) based interventions for preventing depression in children and adolescents. *Cochrane Database Syst Rev* 2016 Aug 09(8):CD003380. [doi: [10.1002/14651858.CD003380.pub4](https://doi.org/10.1002/14651858.CD003380.pub4)] [Medline: [27501438](https://pubmed.ncbi.nlm.nih.gov/27501438/)]
31. Ayers S, Fitzgerald G, Thompson S. Brief Online Self-help Exercises for Postnatal Women to Improve Mood: A Pilot Study. *Matern Child Health J* 2015 Nov 18;19(11):2375-2383. [doi: [10.1007/s10995-015-1755-5](https://doi.org/10.1007/s10995-015-1755-5)] [Medline: [26084670](https://pubmed.ncbi.nlm.nih.gov/26084670/)]
32. Strauman TJ, Socolar Y, Kwapil L, Cornwell JF, Franks B, Sehnert S, et al. Microinterventions targeting regulatory focus and regulatory fit selectively reduce dysphoric and anxious mood. *Behav Res Ther* 2015 Sep;72:18-29 [FREE Full text] [doi: [10.1016/j.brat.2015.06.003](https://doi.org/10.1016/j.brat.2015.06.003)] [Medline: [26163353](https://pubmed.ncbi.nlm.nih.gov/26163353/)]
33. Elephant AB, Contreras O, Muñoz RF, Bunge EL, Leykin Y. Microinterventions produce immediate but not lasting benefits in mood and distress. *Internet Interv* 2017 Dec;10:17-22 [FREE Full text] [doi: [10.1016/j.invent.2017.08.004](https://doi.org/10.1016/j.invent.2017.08.004)] [Medline: [29270366](https://pubmed.ncbi.nlm.nih.gov/29270366/)]
34. Pellegrini CA, Hoffman SA, Daly ER, Murillo M, Iakovlev G, Spring B. Acceptability of smartphone technology to interrupt sedentary time in adults with diabetes. *Transl Behav Med* 2015 Sep 14;5(3):307-314 [FREE Full text] [doi: [10.1007/s13142-015-0314-3](https://doi.org/10.1007/s13142-015-0314-3)] [Medline: [26327936](https://pubmed.ncbi.nlm.nih.gov/26327936/)]
35. Heron K, Smyth J. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *Br J Health Psychol* 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: [10.1348/135910709X466063](https://doi.org/10.1348/135910709X466063)] [Medline: [19646331](https://pubmed.ncbi.nlm.nih.gov/19646331/)]
36. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. *Depress Anxiety* 2017 Jun 11;34(6):540-545. [doi: [10.1002/da.22649](https://doi.org/10.1002/da.22649)] [Medline: [28494123](https://pubmed.ncbi.nlm.nih.gov/28494123/)]
37. Nahum-Shani I, Smith S, Spring B, Collins L, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAI)s in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Ann Behav Med* 2018 May 18;52(6):446-462 [FREE Full text] [doi: [10.1007/s12160-016-9830-8](https://doi.org/10.1007/s12160-016-9830-8)] [Medline: [27663578](https://pubmed.ncbi.nlm.nih.gov/27663578/)]
38. Embry DD, Biglan A. Evidence-based kernels: fundamental units of behavioral influence. *Clin Child Fam Psychol Rev* 2008 Sep 20;11(3):75-113 [FREE Full text] [doi: [10.1007/s10567-008-0036-x](https://doi.org/10.1007/s10567-008-0036-x)] [Medline: [18712600](https://pubmed.ncbi.nlm.nih.gov/18712600/)]
39. Graham AL, Papandonatos GD, Jacobs MA, Amato MS, Cha S, Cohn AM, et al. Optimizing Text Messages to Promote Engagement With Internet Smoking Cessation Treatment: Results From a Factorial Screening Experiment. *J Med Internet Res* 2020 Apr 2;22(4):e17734. [doi: [10.2196/17734](https://doi.org/10.2196/17734)]
40. Muench F, van Stolk-Cooke K, Kuerbis A, Stadler G, Baumel A, Shao S, et al. A Randomized Controlled Pilot Trial of Different Mobile Messaging Interventions for Problem Drinking Compared to Weekly Drink Tracking. *PLoS One* 2017 Feb 1;12(2):e0167900 [FREE Full text] [doi: [10.1371/journal.pone.0167900](https://doi.org/10.1371/journal.pone.0167900)] [Medline: [28146560](https://pubmed.ncbi.nlm.nih.gov/28146560/)]



41. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To Prompt or Not to Prompt? A Microrandomized Trial of Time-Varying Push Notifications to Increase Proximal Engagement With a Mobile Health App. *JMIR Mhealth Uhealth* 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: [10.2196/10123](https://doi.org/10.2196/10123)] [Medline: [30497999](https://pubmed.ncbi.nlm.nih.gov/30497999/)]
42. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keenanasseril A, et al. Interventions for enhancing medication adherence. *Cochrane Database Syst Rev* 2014 Nov 20(11):CD000011 [FREE Full text] [doi: [10.1002/14651858.CD000011.pub4](https://doi.org/10.1002/14651858.CD000011.pub4)] [Medline: [25412402](https://pubmed.ncbi.nlm.nih.gov/25412402/)]
43. Glas N, Pelachaud C. Definitions of engagement in human-agent interaction. 2015 Presented at: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII); September 21-24, 2015; Xi'an, China. [doi: [10.1109/acii.2015.7344688](https://doi.org/10.1109/acii.2015.7344688)]
44. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Transl Behav Med* 2017 Jun 13;7(2):254-267 [FREE Full text] [doi: [10.1007/s13142-016-0453-1](https://doi.org/10.1007/s13142-016-0453-1)] [Medline: [27966189](https://pubmed.ncbi.nlm.nih.gov/27966189/)]
45. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". *Transl Behav Med* 2020 Feb 03;10(1):267-277. [doi: [10.1093/tbm/ibz039](https://doi.org/10.1093/tbm/ibz039)] [Medline: [30927357](https://pubmed.ncbi.nlm.nih.gov/30927357/)]
46. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. *Mhealth* 2018 Sep;4:37-37 [FREE Full text] [doi: [10.21037/mhealth.2018.08.04](https://doi.org/10.21037/mhealth.2018.08.04)] [Medline: [30363749](https://pubmed.ncbi.nlm.nih.gov/30363749/)]
47. Dishion T, McMahon R. Parental monitoring and the prevention of child and adolescent problem behavior: a conceptual and empirical formulation. *Clin Child Fam Psychol Rev* 1998 Mar;1(1):61-75. [doi: [10.1023/a:1021800432380](https://doi.org/10.1023/a:1021800432380)] [Medline: [11324078](https://pubmed.ncbi.nlm.nih.gov/11324078/)]
48. Sanders MR. Triple P-Positive Parenting Program: towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. *Clin Child Fam Psychol Rev* 1999 Jun;2(2):71-90. [doi: [10.1023/a:1021843613840](https://doi.org/10.1023/a:1021843613840)] [Medline: [11225933](https://pubmed.ncbi.nlm.nih.gov/11225933/)]
49. Borden LA, Schultz TR, Herman KC, Brooks CM. The Incredible Years Parent Training Program: Promoting resilience through evidence-based prevention groups. *Group Dynamics: Theory, Research, and Practice* 2010;14(3):230-241. [doi: [10.1037/a0020322](https://doi.org/10.1037/a0020322)]
50. Cavanagh K, Millings A. (Inter)personal Computing: The Role of the Therapeutic Relationship in E-mental Health. *J Contemp Psychother* 2013 Jul 17;43(4):197-206. [doi: [10.1007/s10879-013-9242-z](https://doi.org/10.1007/s10879-013-9242-z)]
51. Holter MT, Johansen A, Brendryen H. How a Fully Automated eHealth Program Simulates Three Therapeutic Processes: A Case Study. *J Med Internet Res* 2016 Jun 28;18(6):e176 [FREE Full text] [doi: [10.2196/jmir.5415](https://doi.org/10.2196/jmir.5415)] [Medline: [27354373](https://pubmed.ncbi.nlm.nih.gov/27354373/)]
52. Barazzone N, Cavanagh K, Richards DA. Computerized cognitive behavioural therapy and the therapeutic alliance: a qualitative enquiry. *Br J Clin Psychol* 2012 Nov;51(4):396-417. [doi: [10.1111/j.2044-8260.2012.02035.x](https://doi.org/10.1111/j.2044-8260.2012.02035.x)] [Medline: [23078210](https://pubmed.ncbi.nlm.nih.gov/23078210/)]
53. Baumel A, Birnbaum ML, Sucala M. A Systematic Review and Taxonomy of Published Quality Criteria Related to the Evaluation of User-Facing eHealth Programs. *J Med Syst* 2017 Aug 22;41(8):128. [doi: [10.1007/s10916-017-0776-6](https://doi.org/10.1007/s10916-017-0776-6)] [Medline: [28735372](https://pubmed.ncbi.nlm.nih.gov/28735372/)]
54. Cohen E, Zerach G, Solomon Z. The implication of combat-induced stress reaction, PTSD, and attachment in parenting among war veterans. *J Fam Psychol* 2011 Oct;25(5):688-698. [doi: [10.1037/a0024065](https://doi.org/10.1037/a0024065)] [Medline: [21639634](https://pubmed.ncbi.nlm.nih.gov/21639634/)]
55. Casselman RB, Pemberton JR. ACT-Based Parenting Group for Veterans With PTSD: Development and Preliminary Outcomes. *The American Journal of Family Therapy* 2014 Jul 30;43(1):57-66. [doi: [10.1080/01926187.2014.939003](https://doi.org/10.1080/01926187.2014.939003)]
56. Marks I, Lovell K, Noshirvani H, Livanou M, Thrasher S. Treatment of posttraumatic stress disorder by exposure and/or cognitive restructuring: a controlled study. *Arch Gen Psychiatry* 1998 Apr 01;55(4):317-325. [doi: [10.1001/archpsyc.55.4.317](https://doi.org/10.1001/archpsyc.55.4.317)] [Medline: [9554427](https://pubmed.ncbi.nlm.nih.gov/9554427/)]
57. Robson JJ, Troutman-Jordan M. A concept analysis of cognitive reframing. *Journal of Theory Construction and Testing* 2014;18(2):55-59.
58. Vernooij-Dassen M, Draskovic I, McCleery J, Downs M. Cognitive reframing for carers of people with dementia. *Cochrane Database Syst Rev* 2011 Nov 09(11):CD005318. [doi: [10.1002/14651858.CD005318.pub2](https://doi.org/10.1002/14651858.CD005318.pub2)] [Medline: [22071821](https://pubmed.ncbi.nlm.nih.gov/22071821/)]
59. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *J Med Internet Res* 2011 Mar 10;13(1):e30 [FREE Full text] [doi: [10.2196/jmir.1602](https://doi.org/10.2196/jmir.1602)] [Medline: [21393123](https://pubmed.ncbi.nlm.nih.gov/21393123/)]
60. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: An Eclectic, Skills-Based App Suite for the Treatment of Depression and Anxiety. *J Med Internet Res* 2017 Jan 05;19(1):e10 [FREE Full text] [doi: [10.2196/jmir.6645](https://doi.org/10.2196/jmir.6645)] [Medline: [28057609](https://pubmed.ncbi.nlm.nih.gov/28057609/)]
61. Baumel A, Correll CU, Hauser M, Brunette M, Rotondi A, Ben-Zeev D, et al. Health Technology Intervention After Hospitalization for Schizophrenia: Service Utilization and User Satisfaction. *Psychiatr Serv* 2016 Sep 01;67(9):1035-1038. [doi: [10.1176/appi.ps.201500317](https://doi.org/10.1176/appi.ps.201500317)] [Medline: [27247171](https://pubmed.ncbi.nlm.nih.gov/27247171/)]
62. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham, Switzerland: Springer; 2018.



63. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. *Br J Psychiatry* 2005 Jan 02;186(1):11-17. [doi: [10.1192/bjp.186.1.11](https://doi.org/10.1192/bjp.186.1.11)] [Medline: [15630118](https://pubmed.ncbi.nlm.nih.gov/15630118/)]
64. Haaga DA. Introduction to the special section on stepped care models in psychotherapy. *Journal of Consulting and Clinical Psychology* 2000;68(4):547-548. [doi: [10.1037/0022-006x.68.4.547](https://doi.org/10.1037/0022-006x.68.4.547)]
65. Sobell MB, Sobell LC. Stepped care as a heuristic approach to the treatment of alcohol problems. *Journal of Consulting and Clinical Psychology* 2000 Aug;68(4):573-579. [doi: [10.1037/0022-006x.68.4.573](https://doi.org/10.1037/0022-006x.68.4.573)]
66. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing Digital Integrated Care via Micro-Randomized Trials. *Clin Pharmacol Ther* 2018 Jul 19;104(1):53-58 [FREE Full text] [doi: [10.1002/cpt.1079](https://doi.org/10.1002/cpt.1079)] [Medline: [29604043](https://pubmed.ncbi.nlm.nih.gov/29604043/)]

*Edited by G Eysenbach; submitted 24.05.20; peer-reviewed by C Yeager, CY Lin; comments to author 07.07.20; revised version received 11.08.20; accepted 13.08.20; published 29.10.20.*

*Please cite as:*

Baumel A, Fleming T, Schueller SM

*Digital Micro Interventions for Behavioral and Mental Health Gains: Core Components and Conceptualization of Digital Micro Intervention Care*

*J Med Internet Res* 2020;22(10):e20631

URL: <http://www.jmir.org/2020/10/e20631/>

doi:[10.2196/20631](https://doi.org/10.2196/20631)

PMID:[33118946](https://pubmed.ncbi.nlm.nih.gov/33118946/)

©Amit Baumel, Theresa Fleming, Stephen M Schueller. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Development of an Item Bank to Measure Medication Adherence: Systematic Review

Yu Heng Kwan<sup>1,2\*</sup>, PhD; Livia Jia Yi Oo<sup>2\*</sup>, BSc; Dionne Hui Fang Loh<sup>3\*</sup>, BEng; Jie Kie Phang<sup>4</sup>, BSc; Si Dun Weng<sup>2</sup>, BSc; Dan V Blalock<sup>5,6</sup>, PhD; Eng Hui Chew<sup>2</sup>, PhD; Kai Zhen Yap<sup>2</sup>, PhD; Corrinne Yong Koon Tan<sup>7</sup>, MBA; Sungwon Yoon<sup>1,3</sup>, PhD; Warren Fong<sup>4,8,9</sup>, FAMS; Truls Østbye<sup>1\*</sup>, PhD; Lian Leng Low<sup>3,10,11\*</sup>, FCFP(S); Hayden Barry Bosworth<sup>5,6,12,13\*</sup>, PhD; Julian Thumboo<sup>1,4,9\*</sup>, FRCP

<sup>1</sup>Program in Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore

<sup>2</sup>Department of Pharmacy, National University of Singapore, Singapore, Singapore

<sup>3</sup>PULSES Centre Grant, SingHealth Regional Health System, Singapore, Singapore

<sup>4</sup>Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore

<sup>5</sup>Center of Innovation to Accelerate Discovery and Practice Transformation (ADAPT), Durham Veterans Affairs Health Care System, Durham, NC, United States

<sup>6</sup>Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, United States

<sup>7</sup>Pharmacy Transformation Office, National Healthcare Group Pharmacy, Singapore, Singapore

<sup>8</sup>Duke-NUS Medical School, Singapore, Singapore

<sup>9</sup>NUS Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

<sup>10</sup>Department of Family Medicine and Continuing Care, Singapore General Hospital, Singapore, Singapore

<sup>11</sup>Post Acute and Continuing Care, Outram Community Hospital, Singapore, Singapore

<sup>12</sup>School of Nursing, Duke University Medical Center, Durham, NC, United States

<sup>13</sup>Department of Population Health Sciences, Duke University Medical Center, Durham, NC, United States

\* these authors contributed equally

## Corresponding Author:

Yu Heng Kwan, PhD

Program in Health Services and Systems Research

Duke-NUS Medical School

8 College Road

Singapore, 169857

Singapore

Phone: 65 63778465

Email: [yuheng@u.duke.nus.edu](mailto:yuheng@u.duke.nus.edu)

## Abstract

**Background:** Medication adherence is important in managing the progression of chronic diseases. A promising approach to reduce cognitive burden when measuring medication adherence lies in the use of computer - adaptive tests (CATs) or in the development of shorter patient-reported outcome measures (PROMs). However, the lack of an item bank currently hampers this progress.

**Objective:** We aim to develop an item bank to measure general medication adherence.

**Methods:** Using the preferred reporting items for systematic review and meta-analysis (PRISMA), articles published before October 2019 were retrieved from PubMed, Embase, CINAHL, the Cochrane Library, and Web of Science. Items from existing PROMs were classified and selected ("binned" and "winnowed") according to standards published by the Patient-Reported Outcomes Measurement Information System (PROMIS) Cooperative Group.

**Results:** A total of 126 unique PROMs were identified from 213 studies in 48 countries. Items from the literature review (47 PROMs with 579 items for which permission has been obtained) underwent binning and winnowing. This resulted in 421 candidate items (77 extent of adherence and 344 reasons for adherence).

**Conclusions:** We developed an item bank for measuring general medication adherence using items from validated PROMs. This will allow researchers to create new PROMs from selected items and provide the foundation to develop CATs.

**KEYWORDS**

systematic review; patient-reported outcome measures; item bank; adherence

**Introduction**

Medication adherence is defined as the degree to which a patient's behavior corresponds with the agreed recommendations from a health care provider [1]. The average adherence rate ranges from 50% among patients suffering from chronic diseases in developed countries [2] to 79% among those receiving medical treatment prescribed by a nonpsychiatrist physician [3,4]. Nonetheless, medication nonadherence is recognized as a significant public health issue since it can result in poor health outcomes and increased health care costs [5]. Medication adherence, which is important in managing the progression of chronic diseases, may be assessed using patient-reported outcome measures (PROMs). Using PROMs to measure medication adherence may be susceptible to a social desirability bias [6]; however, PROMs are much more practical in daily clinical practice because of their relatively low cost and ease of administration as compared to pill counting or an electronic monitoring system such as the medication event monitoring system (MEMS).

PROMs are measures of the status of a patient's health condition that originate directly from the patient, without interpretation of the patient's response by a caregiver or physician [7]. Numerous PROMs have been developed and validated to measure medication adherence. These include instruments such as the Medication Adherence Report Scale (MARS) [8], the 4- and 8-item Morisky Medication Adherence Scale (MMAS-4 and MMAS-8) [9], the Hill-Bone Medication Adherence (HBMA) scale [10], and the Domains of Subjective Extent of Nonadherence (DOSE-Nonadherence) scale [11]. However, most self-report measures that were developed using classical test theory [12,13] are static and administered using a common item set regardless of the respondent's level of medication adherence [14]. Since patients are asked the same questions repeatedly, this approach results in significant cognitive burden [15], low precision [16], a waste of the patients' time, as well as a lack of additional, new information [17].

A novel approach to overcome this limitation lies in the use of a computer - adaptive test (CAT) to create new PROMs for measuring medication adherence. A CAT is a system for tailoring a test, whereby the next item administered to the respondent is determined by and adaptive to the patient's response to the previously-administered item [18]. The Patient-Reported Outcomes Measurement Information System (PROMIS) was developed through a CAT and item response theory (IRT). Instead of focusing on the entire test, IRT shifts the focus to the individual questions [19]. The use of IRT with a CAT allows for the identification, individualization, and administration of a feasible number of items that are likely to offer the highest precision [20]. In order to achieve higher precision, PROMIS investigators were required to identify and develop items covering the full range of experience in the domains the instrument was intended to measure (ie, content

validity) [14]. Thus, the first step to a CAT is an item bank consisting of questions from medication adherence PROMs. Therefore, this systematic literature review aims to identify and develop an item bank through a comprehensive summary of the questions from validated medication adherence PROMs.

**Methods**

This systematic review was guided by the preferred reporting items for systematic review and meta-analysis (PRISMA) statement [21]; standards published by the PROMIS committee [22] were adapted in the development of the item bank.

**Search Strategy**

Articles published before October 2019 were retrieved from PubMed, Embase, CINAHL, the Cochrane Library, and Web of Science. A search strategy (Multimedia Appendix 1) of 4 components was used as follows: construct of interest, population, instrument, and measurement properties. The searches focused on medication adherence PROMs. Where available, the sensitivity of the searches was enhanced using search filters developed by Terwee et al [23], which involves a combination of search terms designed to retrieve studies on measurement properties of measurement instruments. The search records were downloaded into Endnote X9 (Clarivate Analytics), and any duplicates were removed.

**Article Selection**

All titles and abstracts were screened independently by 2 reviewers (LJYO and SDW). A third reviewer (YHK) was consulted when a disagreement arose between the 2 reviewers. For articles that were potentially relevant, the full text of these articles was independently reviewed by the same 2 reviewers for inclusion or exclusion.

Articles were included if they were full-text original publications in English that validated PROMs for medication adherence. Articles were excluded if the PROMs were completed by proxy, or if they were conference abstracts, expert opinions, narrative reviews, or not peer-reviewed. Animal and case studies, as well as non-English language studies, were also excluded. These exclusions were not used to construct the search strategy to avoid the omission of relevant articles.

**Data Extraction**

Where available, 2 reviewers (LJYO and SDW) extracted study population characteristics (sample size, age, gender, and country) data from the articles.

**Identification of Existing PROMs for Inclusion**

The names of PROMs extracted from the previous step were consolidated. In order to optimize the number of relevant items for evaluation, the most recent and exhaustive version was included when multiple versions of the same PROM were found across the included studies.

For the assessment of PROMs for inclusion, reviewers obtained information regarding the PROM through internet searches. Copies of the shortlisted PROMs were retrieved either from sources available to the public (ie, official websites or research publications) or by requesting copies from the developers or study investigators of these PROMs. Permission was obtained from the study investigators for the inclusion of the PROM into the item bank. Where possible, permission from the PROM developers was sought in the case that the study investigators were not in a position to provide consent due to claims of intellectual property. After the initial contact, 2 follow-up reminder emails were sent to the unresponsive study investigators. This resulted in a final list of PROMs from which items were extracted and evaluated.

### Item Classification (Binning)

Item classification, or binning, refers to a systematic process for grouping items according to meaning and specific latent construct. This process aims to obtain a bin with the most exhaustive list of items, from which a smaller number of items may be chosen to adequately represent the bin. The number of items that would adequately represent a bin was not predetermined, as the purpose of this process was to identify sufficient items that encompass the meaning of the bin and to eliminate unnecessary redundancy in the pool of items [14].

Binning was done using terms in English, and each item was included in as many bins as was deemed fit. To ensure that the binning process was exhaustive, 2 independent reviewers (LJYO and DHFL) evaluated any one item for possible inclusion, and an item identified for inclusion to a bin by at least 1 reviewer was included in that bin.

A 2-stage process was carried out for binning. First-order binning was completed at the level of the domain: each item was evaluated for possible inclusion into (1) Extent of Adherence and/or (2) Reasons for Nonadherence. The domains were derived from the self-report measure developed by Voils et al [11].

For the Reasons for Nonadherence domain, second-order binning was completed at the level of the subdomain consisting of (1) social- and economic-related factors, (2) health care team and system-related factors, (3) therapy-related factors, (4) condition-related factors, and (5) patient-related factors. The subdomains were derived from the World Health Organisation (WHO) framework for medication adherence [2]. The

framework has been widely used in various literature [24-32]. Second-order binning was not deemed to be necessary for the Extent of Adherence domain.

### Item Selection (Winnowing)

The process of winnowing aims to narrow the large pool of items down to a representative set of items; this is done by identifying item characteristics that would either include or exclude items from the item bank based on the definition of the domains [14]. Winnowing was performed by 2 reviewers (LJYO and DHFL) independently assessing each bin, and items that best represented the respective domains were first selected. The process was carried out separately for each domain.

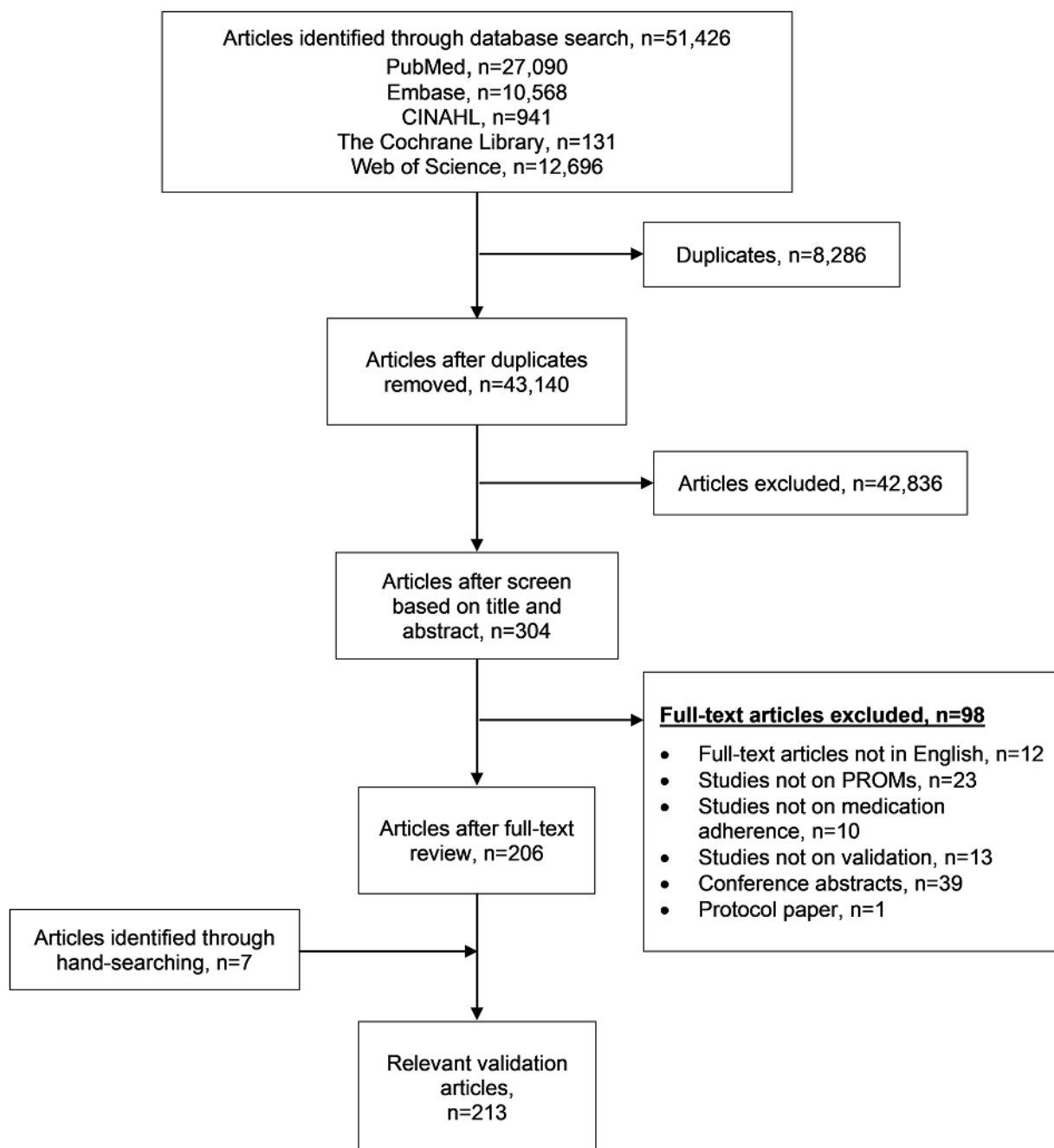
The following criteria were used to eliminate items from consideration [14]: (1) the content of the item was inconsistent with the definition of medication adherence, or with the scope of the extent of and reasons for medication adherence; (2) the item was semantically redundant with a previous item; (3) the content of the item was too narrow to be universally applicable; (4) the stem of the item was highly disease-specific, which reduces overall applicability and limits the adaptation of the item; (5) the item was confusing; and (6) the item was open-ended, which increases the difficulty of implementation.

After the pair of reviewers completed the item selection independently, a third reviewer (YHK), who is trained in measurement science and item banking and was not previously involved in the binning and winnowing process, was consulted to identify the items that best represented each domain, as well as the items for removal.

## Results

### Search Results and Characteristics of the Included Articles

A total of 51,426 studies were obtained from the database search, of which 8286 duplicates were excluded. A review of the titles and abstracts led to the exclusion of 42,836 studies. Subsequently, full-text review excluded 98 studies, with reasons provided in Figure 1. An additional 7 studies were identified through hand-searching of reference lists, resulting in 213 relevant validation studies for potential inclusion into the item bank. The characteristics of the relevant studies are presented in Table 1.

**Figure 1.** Flow chart of the systematic literature review.



**Table 1.** Characteristics of studies for potential inclusion into the item bank (n=213).

General Characteristics	Values
Number of unique countries involved	48
Number of unique PROMs <sup>a</sup> studied	126
<b>Sample size, n (%)</b>	
< 50	17 (8.0)
50 – 99	31 (14.6)
100 – 199	64 (30.0)
200 – 299	30 (14.1)
300 – 399	22 (10.3)
400 – 499	15 (7.0)
> 500	34 (16.0)
<b>Mean age in years, n (%)</b>	
0 < mean age ≤ 20	9 (4.2)
20 < mean age ≤ 40	28 (13.2)
40 < mean age ≤ 60	101 (47.4)
60 < mean age ≤ 80	54 (25.4)
≥ 80	2 (0.9)
Not reported <sup>b</sup>	19 (8.9)
<b>Proportion of males, n (%)</b>	
< 0.2	15 (7.0)
0.2 ≤ x < 0.4	46 (21.6)
0.4 ≤ x < 0.6	77 (36.2)
0.6 ≤ x < 0.8	40 (18.8)
≥ 0.8	23 (10.8)
Not reported	12 (5.6)

<sup>a</sup> PROMs: patient-reported outcome measures.

<sup>b</sup> Includes values in the form of median, range, or not reported.

### Identification of Existing PROMs for Inclusion

The review of the included articles identified 126 unique PROMs measuring medication adherence, which were validated in 48 countries. A majority of the identified PROMs were self-administered questionnaires.

After obtaining written permission from the study investigators and PROM developers for the use of the items as part of the item bank, 47 PROMs from 53 studies were included in the item bank, as presented in [Textbox 1](#). PROMs from the

remaining 160 studies ([Multimedia Appendix 2 \[33-189\]](#)) were excluded due to a lack of consent. Among the 53 included studies, the Modified Drug Adherence Work-Up (M-DRAW) tool, the Reduced Glaucoma Treatment Compliance Assessment Tool (GTCAT), and the 3-Item Self-Report Measure for Medication Adherence were each evaluated in 2 studies; the General Medication Adherence Scale (GMAS) was evaluated in 3 studies; and DOSE-Nonadherence was evaluated in 4 studies. PROMs included were mostly developed and validated in English. A total of 579 items were collated from these PROMs, including 71 non-English items.

**Textbox 1.** List of the patient-reported outcome measures (PROMs) that were included in the item bank. \*Instruments which were not named; \*\*instruments developed/validated in non-English but have an existing English-translated version for publication purposes; \*\*\*instruments developed and/or validated in non-English.

#### Generic PROMs

- Diagnostic Adherence to Medication Scale (DAMS) [190] (6 items)
- 7-Item Adherence to Refills and Medications Scale (ARMS-7) [191]\*\* (7 items)
- Adherence to Refills and Medications Scale (ARMS) [192] (12 items)
- Brief Medication Adherence Scale (BMAS) [193]\*\* (10 items)
- Domains of Subjective Extent of Nonadherence (DOSE-Nonadherence) Scale [11,194-196] (21 items)
- Every Visit Adherence Questionnaire [197] (1 item)
- General Medication Adherence Scale (GMAS) [198-200] (11 items)
- Medication Adherence Estimation and Differentiation Scale (MEDS) [201] (16 items)
- Modified Drug Adherence Work-Up (M-DRAW) Tool [202,203] (14 items)
- Self-Reported Adherence (SERAD) Questionnaire [204]\*\*\* (2 items)
- Simplified Medication Adherence Questionnaire (SMAQ) [205] (6 items)
- 3-Item Self-Report Measure for Medication Adherence [206,207]\* (3 items)

#### Disease-Specific PROMs

- 12-Item Medication Adherence Scale for Patients with Chronic Disease [208]\* (12 items)
- 5-Item Compliance Questionnaire for Rheumatology (CQR5) [209] (5 items)
- Adult Asthma Adherence Questionnaire (AAAQ) [210] (5 items)
- Antidepressant Adherence Scale (AAS) [211] (4 items)
- Antipsychotic Medication Beliefs and Attitudes Scale (AMBAS) [212] (12 items)
- Assessment Scale for Treatment Compliance in Type 2 Diabetes Mellitus [213]\* (30 items)
- Brief Evaluation of Medication Influences and Beliefs (BEMIB) [214] (8 items)
- Chinese and Western Medication Adherence Scale in Chronic Kidney Disease [215]\* (18 items)
- Chinese Diabetes Medication Self-Efficacy Scale (CDMSS) [216]\*\*\* (19 items)
- Combination Antiretroviral Therapy Adherence Questionnaire [217]\* (4 items)
- Diabetes Management Questionnaire (DMQ) [218] (21 items)
- Diabetes Medication Self-Efficacy Scale (DMSS) [219] (19 items)
- ICAMP Adherence Questionnaire [220] (10 items)
- Immunosuppressant Therapy Adherence Scale (ITAS) [221]\*\*\* (4 items)
- Iraqi Anti-Diabetic Medication Adherence Scale (IADMAS) [222] (8 items)
- IRT-30 [223] (10 items)
- Lasso-10 [223] (30 items)
- Measure for Intention to Adhere to HIV Treatment [224]\* (14 items)
- Medication Adherence Self-Reports in Adults with Type 2 Diabetes [225]\* (6 items)
- Medication Adherence Survey for Hemodialysis Patients [226]\* (23 items)
- Multiple Sclerosis Treatment Adherence Questionnaire (MS-TAQ) [227] (11 items)
- Outcome Expectations for Osteoporosis Medication Adherence Scale in Chinese Immigrants (OEOMA-C) [228]\*\* (5 items)
- Patient-Reported Measures Assessing Adherence Behaviors and Barriers in Patients Living with HIV [229]\* (7 items)
- Perceived Barriers to Antiretroviral Therapy Adherence (PEDIA) Scale [230]\*\* (18 items)
- Portuguese End-Stage Renal Disease Questionnaire (PESRD-AQ) [231]\*\*\* (46 items)
- Questionnaire for Adherence with Topical Treatments in Psoriasis (QATOP) [232] (9 items)
- Reduced Glaucoma Treatment Compliance Assessment Tool (GTCAT) [233,234] (28 items)

- Risk of Nonadherence to Antibiotic Treatment Questionnaire [235]\* (20 items)
- Self-Assessment Tool to Measure Imatinib Adherence in Patients with Chronic Myeloid Leukemia [236]\* (10 items)
- Self-Efficacy for Osteoporosis Medication Adherence Scale in Chinese Immigrants (SEOMA-C) [228]\*\* (14 items)
- Self-Report Measures in Assessing Antiretroviral Adherence [237]\* (3 items)
- Self-Reported Compliance to Metered-Dose Inhalers Questionnaire [238]\* (4 items)
- Self-Reported Questionnaire Assessing Adherence to Antiretroviral Medication [239]\* (5 items)
- Test of the Adherence to Inhalers (TAI) [240] (12 items)
- Treatment Adherence Survey – Patient Version (TAS-P) [241] (16 items)

## Item Evaluation

At the stage of winnowing, 319 items inconsistent with the definitions of the domains or subdomains were removed. The 2 independent reviewers achieved a 90.4% agreement on the items for removal. Most eliminated items were found to be highly disease-specific (eg, “I arrange my oral antidiabetic medication or insulin dose myself according to my food intake.”) or open-ended (eg, “How many doses did you miss?”).

In the Extent of Adherence, a total of 77 representative items were identified, while 344 items were included in the Reasons for Nonadherence, with an average of 70 items (SD 43.1; range 24–137) within each of the 5 dimensions. The breakdown of the number of items in each domain and subdomain is summarised in Table 2, and a representative table of the mapping of each item is presented in Multimedia Appendix 3. Binned and winnowed items that were granted approval by the study investigators or PROM developers to be openly listed in the item bank are presented in Multimedia Appendix 4.

**Table 2.** Summary of domains and the number of items in each bin.

Domain/subdomain	Number of items
Extent of adherence	77
<b>Reasons for adherence</b>	
Social and economic factors	32
Health care team and system-related factors	24
Condition-related factors	37
Therapy-related factors	114
Patient-related factors	137

## Discussion

This systematic review summarises the process of developing an item bank for the domain of general medication adherence. A multistep approach adapted from the PROMIS standards for the development of item banks [14] was used. In order to enhance the comprehensiveness of the item bank, items originating from both disease-specific and nondisease specific PROMs were considered and reviewed for inclusion. To the best of our knowledge, this is the first systematic review to collate items from various medication adherence PROMs into an item bank. As numerous PROMs have been developed and validated for use in the measurement of medication adherence, an item bank consisting of the most representative items may improve the relevance and precision of assessments [242].

The current item bank enables researchers to select items for the creation of new PROMs. The rare and highly disease-specific items were eliminated during the winnowing process; however, the items included in the item bank may be adapted for use in both specific patient groups and the general population. There have been conflicting results on the association between age and medical adherence, with some studies providing evidence

of an association between younger age and nonadherence [243] and others showing an association between older age and nonadherence [244]. A CAT that was developed using IRT based on our item bank can potentially examine the changing pattern of medical adherence during a lifetime more precisely than traditional PROMs that are developed using classical test theory [245,246]. In addition, previous studies have shown that adherence to different medications may vary within the same individual [247,248]. We can potentially overcome this issue by specifying the type of medication in the question stem in the CAT.

Towards the use of the item bank in CATs, the next steps are to revise and review the items through cognitive interviews, to calibrate the items using IRT, and to evaluate the validity of test scores when the items are administered adaptively. The items that underwent binning and winnowing originated from various PROMs. They were created in varying styles, phrasings, and sentence structures, with different response options for each item, including dichotomous, Likert, and semantic differential scales. Due to the discordances among the items, the item revision process would be prudent to facilitate the administration of the items as 1 coherent test. Once fully developed, the resulting instrument may potentially improve measurement

precision and allow for a reduction in assessment time [249] and, accordingly, patient cognitive burden. Furthermore, since the scores are directly comparable [250], such an instrument would allow health care providers to measure and compare the medication adherence of a patient from consult to consult. Alongside other clinical measures, such as symptom scores and quality of life scales, these measurements may enable physicians to detect issues in a patient's medication adherence, implement timely interventions, evaluate the effectiveness of the interventions, and make prompt modifications when necessary.

This study has several strengths; 5 databases, as well as sensitive search filters, were used to capture as many potentially relevant articles as possible. The rigor of this study was established using the PRISMA statement and PROMIS standards. The PRISMA statement was used because it enhances the transparency and clarity of systematic reviews [251]. The utilization of modern statistical methods to improve the functionality of PROMs has elevated the expectations of instruments beyond robust psychometric properties. As such, the PROMIS standards were adapted for use as they endorse the minimum standards for PROMs used in patient-centered health outcomes and facilitate the development of common metrics for accurate comparisons across conditions, healthcare systems, and geographical locations [252].

This systematic review has some limitations. Firstly, we only included full-text articles in English. Full-text articles were necessary as they are peer-reviewed and recommended by Terwee et al [253]. Nonetheless, the 12 foreign-language articles that would otherwise be eligible for full-text review made up only 3.9% of the included articles. Secondly, the item bank does not include items developed by the 160 authors (75.1% of the

articles assessed for inclusion) who either chose not to consent to the inclusion of their PROMs or were uncontactable. Of note, PROMs included in the item bank have adequately captured the concepts measured by the majority of items for which consent was not obtained. To ensure that no range of concepts has been overlooked from the exclusion of the studies, future studies can consider conducting interviews with patients, as well as conducting expert opinion reviews to elicit the conceptual model for medical adherence relevant to the local context. Thirdly, this systematic review only included validation studies of medication adherence PROMs. This was deemed appropriate given the importance of validation in justifying the use of the instrument [254]. Well-validated adherence scales have been strongly correlated with objective measures of adherence [255], allowing accurate and reliable assessment of medication adherence. In addition, expert review and patient feedback were not sought as part of the item evaluation process in this study. As the concept of medical adherence may differ between different sociocultural contexts [256,257], different countries should perform an expert review and gather patient feedback through cognitive interviews based on this item bank to ensure the development of a culturally sensitive instrument to measure medication adherence.

In conclusion, this study has identified and collated the items from 47 unique medication adherence PROMs into an item bank through a systematic review. Researchers are able to select appropriate items from the item bank for the creation of new PROMs. Future research may consider revising and reviewing the items through cognitive interviews, calibrating the items through IRT, and developing a CAT to measure medication adherence precisely.

---

## Acknowledgments

KYH, WSD, and DHFL contributed equally as co-first authors. TØ, LLL, HBB, and JT contributed equally as senior coauthors.

We would like to thank librarian Ms Wong Swei Nee at the National University of Singapore Medical Library for her advice on search strategies.

This work was supported by the SingHealth Regional Health System (RHS) Population-based, Unified Learning System for Enhanced and Sustainable Health (PULSES) Centre Grant – “Measuring Medication Adherence in Singapore” (CGDec19S03), and PULSES Fellowship (CGDec19F02).

---

## Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Search strategy.

[DOCX File, 32 KB - [jmir\\_v22i10e19089\\_app1.docx](#) ]

---

### Multimedia Appendix 2

Studies excluded from the item bank due to a lack of consent.

[DOCX File, 41 KB - [jmir\\_v22i10e19089\\_app2.docx](#) ]

---

### Multimedia Appendix 3

Item mapping of domains and subdomains.

[DOCX File, 57 KB - [jmir\\_v22i10e19089\\_app3.docx](#) ]

## Multimedia Appendix 4

Item bank to measure the extent of and reasons for medication adherence.

[DOCX File, 67 KB - [jmir\\_v22i10e19089\\_app4.docx](#)]

## References

1. Jimmy B, Jose J. Patient medication adherence: measures in daily practice. *Oman Med J* 2011 May;26(3):155-159 [FREE Full text] [doi: [10.5001/omj.2011.38](#)] [Medline: [22043406](#)]
2. World Health Organization. Adherence to Long-Term Therapies: Evidence For Action. Geneva, Switzerland: World Health Organization 2003:E.
3. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Med Care* 2004 Mar;42(3):200-209. [Medline: [15076819](#)]
4. Gellad WF, Thorpe CT, Steiner JF, Voils CI. The myths of medication adherence. *Pharmacoepidemiol Drug Saf* 2017 Dec;26(12):1437-1441. [doi: [10.1002/pds.4334](#)] [Medline: [28994158](#)]
5. Balkrishnan R. The importance of medication adherence in improving chronic-disease related outcomes: what we know and what we need to further know. *Med Care* 2005 Jun;43(6):517-520. [doi: [10.1097/01.mlr.0000166617.68751.5f](#)] [Medline: [15908845](#)]
6. Arnsten JH, Demas PA, Farzadegan H, Grant RW, Gourevitch MN, Chang CJ, et al. Antiretroviral therapy adherence and viral suppression in HIV-infected drug users: comparison of self-report and electronic monitoring. *Clin Infect Dis* 2001 Oct 15;33(8):1417-1423 [FREE Full text] [doi: [10.1086/323201](#)] [Medline: [11550118](#)]
7. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. *Med Care* 2013 Aug;51(8 Suppl 3):S73-S79 [FREE Full text] [doi: [10.1097/MLR.0b013e31829b1d84](#)] [Medline: [23774513](#)]
8. Mahler C, Hermann K, Horne R, Ludt S, Haefeli WE, Szecsenyi J, et al. Assessing reported adherence to pharmacological treatment recommendations. Translation and evaluation of the Medication Adherence Report Scale (MARS) in Germany. *J Eval Clin Pract* 2010 Jun;16(3):574-579. [doi: [10.1111/j.1365-2753.2009.01169.x](#)] [Medline: [20210821](#)]
9. Moon SJ, Lee W, Hwang JS, Hong YP, Morisky DE. Accuracy of a screening tool for medication adherence: A systematic review and meta-analysis of the Morisky Medication Adherence Scale-8. *PLoS One* 2017;12(11):e0187139 [FREE Full text] [doi: [10.1371/journal.pone.0187139](#)] [Medline: [29095870](#)]
10. Song Y, Han H, Song H, Nam S, Nguyen T, Kim MT. Psychometric evaluation of hill-bone medication adherence subscale. *Asian Nurs Res (Korean Soc Nurs Sci)* 2011 Sep;5(3):183-188 [FREE Full text] [doi: [10.1016/j.anr.2011.09.007](#)] [Medline: [25030368](#)]
11. Voils CI, Maciejewski ML, Hoyle RH, Reeve BB, Gallagher P, Bryson CL, et al. Initial validation of a self-report measure of the extent of and reasons for medication nonadherence. *Med Care* 2012 Dec;50(12):1013-1019 [FREE Full text] [doi: [10.1097/MLR.0b013e318269e121](#)] [Medline: [22922431](#)]
12. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther* 2014 May;36(5):648-662 [FREE Full text] [doi: [10.1016/j.clinthera.2014.04.006](#)] [Medline: [24811753](#)]
13. McKenna SP, Heaney A, Wilburn J. Measurement of patient-reported outcomes. 2: Are current measures failing us? *J Med Econ* 2019 Jun;22(6):523-530. [doi: [10.1080/13696998.2018.1560304](#)] [Medline: [30556787](#)]
14. DeWalt DA, Rothrock N, Yount S, Stone AA, PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care* 2007 May;45(5 Suppl 1):S12-S21 [FREE Full text] [doi: [10.1097/01.mlr.0000254567.79743.e2](#)] [Medline: [17443114](#)]
15. Luckett T, Butow PN, King MT, Oguchi M, Heading G, Hackl NA, et al. A review and recommendations for optimal outcome measures of anxiety, depression and general distress in studies evaluating psychosocial interventions for English-speaking adults with heterogeneous cancer diagnoses. *Support Care Cancer* 2010 Oct;18(10):1241-1262. [doi: [10.1007/s00520-010-0932-8](#)] [Medline: [20596731](#)]
16. Jabrayilov R, Emons WHM, Sijtsma K. Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Appl Psychol Meas* 2016 Nov;40(8):559-572 [FREE Full text] [doi: [10.1177/0146621616664046](#)] [Medline: [29881070](#)]
17. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights* 2013;6:61-68 [FREE Full text] [doi: [10.4137/HSL.S11093](#)] [Medline: [25114561](#)]
18. Kimura T. The impacts of computer adaptive testing from a variety of perspectives. *J Educ Eval Health Prof* 2017;14:12 [FREE Full text] [doi: [10.3352/jeehp.2017.14.12](#)] [Medline: [28552899](#)]
19. Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. *J Rheumatol* 2014 Jan;41(1):153-158. [doi: [10.3899/jrheum.130813](#)] [Medline: [24241485](#)]
20. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res* 1997 Aug;6(6):595-600. [doi: [10.1023/a:1018420418455](#)] [Medline: [9330558](#)]



21. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
22. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol* 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: [10.1016/j.jclinepi.2010.04.011](https://doi.org/10.1016/j.jclinepi.2010.04.011)] [Medline: [20685078](https://pubmed.ncbi.nlm.nih.gov/20685078/)]
23. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009 Oct;18(8):1115-1123 [FREE Full text] [doi: [10.1007/s11136-009-9528-5](https://doi.org/10.1007/s11136-009-9528-5)] [Medline: [19711195](https://pubmed.ncbi.nlm.nih.gov/19711195/)]
24. Yap AF, Thirumoorthy T, Kwan YH. Medication adherence in the elderly. *Journal of Clinical Gerontology and Geriatrics* 2016 Jun;7(2):64-67. [doi: [10.1016/j.jcgg.2015.05.001](https://doi.org/10.1016/j.jcgg.2015.05.001)]
25. Yap AF, Thirumoorthy T, Kwan YH. Systematic review of the barriers affecting medication adherence in older adults. *Geriatr Gerontol Int* 2016 Oct;16(10):1093-1101. [doi: [10.1111/ggi.12616](https://doi.org/10.1111/ggi.12616)] [Medline: [26482548](https://pubmed.ncbi.nlm.nih.gov/26482548/)]
26. Goh XTW, Tan YB, Thirumoorthy T, Kwan YH. A systematic review of factors that influence treatment adherence in paediatric oncology patients. *J Clin Pharm Ther* 2017 Feb;42(1):1-7. [doi: [10.1111/jcpt.12441](https://doi.org/10.1111/jcpt.12441)] [Medline: [28045208](https://pubmed.ncbi.nlm.nih.gov/28045208/)]
27. Goh H, Kwan YH, Seah Y, Low LL, Fong W, Thumboo J. A systematic review of the barriers affecting medication adherence in patients with rheumatic diseases. *Rheumatol Int* 2017 Oct;37(10):1619-1628. [doi: [10.1007/s00296-017-3763-9](https://doi.org/10.1007/s00296-017-3763-9)] [Medline: [28681249](https://pubmed.ncbi.nlm.nih.gov/28681249/)]
28. Yeam CT, Chia S, Tan HCC, Kwan YH, Fong W, Seng JJB. A systematic review of factors affecting medication adherence among patients with osteoporosis. *Osteoporos Int* 2018 Dec;29(12):2623-2637. [doi: [10.1007/s00198-018-4759-3](https://doi.org/10.1007/s00198-018-4759-3)] [Medline: [30417253](https://pubmed.ncbi.nlm.nih.gov/30417253/)]
29. Fernandez-Lazaro CI, García-González JM, Adams DP, Fernandez-Lazaro D, Mielgo-Ayuso J, Caballero-Garcia A, et al. Adherence to treatment and related factors among patients with chronic conditions in primary care: a cross-sectional study. *BMC Fam Pract* 2019 Sep 14;20(1):132 [FREE Full text] [doi: [10.1186/s12875-019-1019-3](https://doi.org/10.1186/s12875-019-1019-3)] [Medline: [31521114](https://pubmed.ncbi.nlm.nih.gov/31521114/)]
30. Herborg H, Haugbølle LS, Sørensen L, Rossing C, Dam P. Developing a generic, individualised adherence programme for chronic medication users. *Pharm Pract (Granada)* 2008 Jul;6(3):148-157 [FREE Full text] [doi: [10.4321/s1886-36552008000300006](https://doi.org/10.4321/s1886-36552008000300006)] [Medline: [25177406](https://pubmed.ncbi.nlm.nih.gov/25177406/)]
31. Jansà M, Hernández C, Vidal M, Nuñez M, Bertran MJ, Sanz S, et al. Multidimensional analysis of treatment adherence in patients with multiple chronic conditions. A cross-sectional study in a tertiary hospital. *Patient Educ Couns* 2010 Nov;81(2):161-168. [doi: [10.1016/j.pec.2009.12.012](https://doi.org/10.1016/j.pec.2009.12.012)] [Medline: [20167450](https://pubmed.ncbi.nlm.nih.gov/20167450/)]
32. McLoughlin A, Bennett K, Cahir C. Developing a model of the determinants of medication nonadherence in older community-dwelling patients. *Ann Behav Med* 2019 Oct 07;53(11):942-954. [doi: [10.1093/abm/kaz004](https://doi.org/10.1093/abm/kaz004)] [Medline: [30870558](https://pubmed.ncbi.nlm.nih.gov/30870558/)]
33. Irvine AA, Saunders JT, Blank MB, Carter WR. Validation of scale measuring environmental barriers to diabetes-regimen adherence. *Diabetes Care* 1990 Jul;13(7):705-711. [doi: [10.2337/diacare.13.7.705](https://doi.org/10.2337/diacare.13.7.705)] [Medline: [2387191](https://pubmed.ncbi.nlm.nih.gov/2387191/)]
34. Lewin AB, LaGreca AM, Geffken GR, Williams LB, Duke DC, Storch EA, et al. Validity and reliability of an adolescent and parent rating scale of type 1 diabetes adherence behaviors: the Self-Care Inventory (SCI). *J Pediatr Psychol* 2009 Oct;34(9):999-1007 [FREE Full text] [doi: [10.1093/jpepsy/jsp032](https://doi.org/10.1093/jpepsy/jsp032)] [Medline: [19423660](https://pubmed.ncbi.nlm.nih.gov/19423660/)]
35. Moss AC, Lillis Y, Edwards George JB, Choudhry NK, Berg AH, Cheifetz AS, et al. Attitudes to mesalamine questionnaire: a novel tool to predict mesalamine nonadherence in patients with IBD. *Am J Gastroenterol* 2014 Dec;109(12):1850-1855. [doi: [10.1038/ajg.2014.158](https://doi.org/10.1038/ajg.2014.158)] [Medline: [24913040](https://pubmed.ncbi.nlm.nih.gov/24913040/)]
36. de Oliveira-Filho AD, Morisky DE, Neves SJF, Costa FA, de Lyra DP. The 8-item Morisky Medication Adherence Scale: validation of a Brazilian-Portuguese version in hypertensive adults. *Res Social Adm Pharm* 2014;10(3):554-561. [doi: [10.1016/j.sapharm.2013.10.006](https://doi.org/10.1016/j.sapharm.2013.10.006)] [Medline: [24268603](https://pubmed.ncbi.nlm.nih.gov/24268603/)]
37. Jerant A, DiMatteo R, Arnsten J, Moore-Hill M, Franks P. Self-report adherence measures in chronic illness: retest reliability and predictive validity. *Med Care* 2008 Nov;46(11):1134-1139. [doi: [10.1097/MLR.0b013e31817924e4](https://doi.org/10.1097/MLR.0b013e31817924e4)] [Medline: [18953223](https://pubmed.ncbi.nlm.nih.gov/18953223/)]
38. Dima AL, Schweitzer A, Diaconu R, Remor E, Wanless RS. Adherence to ARV medication in Romanian young adults: self-reported behaviour and psychological barriers. *Psychol Health Med* 2013;18(3):343-354. [doi: [10.1080/13548506.2012.722648](https://doi.org/10.1080/13548506.2012.722648)] [Medline: [22985131](https://pubmed.ncbi.nlm.nih.gov/22985131/)]
39. Dima AL, van Ganse E, Laforest L, Texier N, de Bruin M, The Astro-Lab Group. Measuring medication adherence in asthma: Development of a novel self-report tool. *Psychol Health* 2017 Oct;32(10):1288-1307. [doi: [10.1080/08870446.2017.1290248](https://doi.org/10.1080/08870446.2017.1290248)] [Medline: [28276742](https://pubmed.ncbi.nlm.nih.gov/28276742/)]
40. Najimi A, Mostafavi F, Sharifirad G, Golshiri P. Development and study of self-efficacy scale in medication adherence among Iranian patients with hypertension. *J Educ Health Promot* 2017;6:83 [FREE Full text] [doi: [10.4103/jehp.jehp\\_64\\_16](https://doi.org/10.4103/jehp.jehp_64_16)] [Medline: [29114551](https://pubmed.ncbi.nlm.nih.gov/29114551/)]
41. Regnault A, Viala-Danten M, Gilet H, Berdeaux G. Scoring and psychometric properties of the Eye-Drop Satisfaction Questionnaire (EDSQ), an instrument to assess satisfaction and compliance with glaucoma treatment. *BMC Ophthalmol* 2010 Feb 01;10:1 [FREE Full text] [doi: [10.1186/1471-2415-10-1](https://doi.org/10.1186/1471-2415-10-1)] [Medline: [20122146](https://pubmed.ncbi.nlm.nih.gov/20122146/)]

42. Athavale AS, Bentley JP, Banahan BF, McCaffrey DJ, Pace PF. Preliminary development of the Medication Nonpersistence Scale. *J Am Pharm Assoc* (2003) 2017;57(3):389-394.e1. [doi: [10.1016/j.japh.2017.01.014](https://doi.org/10.1016/j.japh.2017.01.014)] [Medline: [28279599](https://pubmed.ncbi.nlm.nih.gov/28279599/)]
43. TURCU- TIOLICĂ A. DEVELOPING OF A NEW TOOL FOR EVALUATION OF THERAPEUTIC ADHERENCE IN COPD. *FARMACIA* 2018 Oct 13;66(5):920-924. [doi: [10.31925/farmacia.2018.5.25](https://doi.org/10.31925/farmacia.2018.5.25)]
44. Zeller A, Schroeder K, Peters TJ. An adherence self-report questionnaire facilitated the differentiation between nonadherence and nonresponse to antihypertensive treatment. *J Clin Epidemiol* 2008 Mar;61(3):282-288. [doi: [10.1016/j.jclinepi.2007.04.007](https://doi.org/10.1016/j.jclinepi.2007.04.007)] [Medline: [18226752](https://pubmed.ncbi.nlm.nih.gov/18226752/)]
45. Zongo A, Guénette L, Moisan J, Grégoire J. Predictive Validity of Self-Reported Measures of Adherence to Noninsulin Antidiabetes Medication against Control of Glycated Hemoglobin Levels. *Can J Diabetes* 2016 Feb;40(1):58-65. [doi: [10.1016/j.cjcd.2015.06.008](https://doi.org/10.1016/j.cjcd.2015.06.008)] [Medline: [26507401](https://pubmed.ncbi.nlm.nih.gov/26507401/)]
46. Zongo A, Guénette L, Moisan J, Guillaumie L, Lauzier S, Grégoire J. Revisiting the internal consistency and factorial validity of the 8-item Morisky Medication Adherence Scale. *SAGE Open Med* 2016;4:2050312116674850 [FREE Full text] [doi: [10.1177/2050312116674850](https://doi.org/10.1177/2050312116674850)] [Medline: [27895914](https://pubmed.ncbi.nlm.nih.gov/27895914/)]
47. Zweben A, Piepmeier ME, Fucito L, O'Malley SS. The clinical utility of the Medication Adherence Questionnaire (MAQ) in an alcohol pharmacotherapy trial. *J Subst Abuse Treat* 2017 Jun;77:72-78 [FREE Full text] [doi: [10.1016/j.jsat.2017.04.001](https://doi.org/10.1016/j.jsat.2017.04.001)] [Medline: [28476276](https://pubmed.ncbi.nlm.nih.gov/28476276/)]
48. Feldman BJ, Fredericksen RJ, Crane PK, Safren SA, Mugavero MJ, Willig JH, et al. Evaluation of the single-item self-rating adherence scale for use in routine clinical care of people living with HIV. *AIDS Behav* 2013 Jan;17(1):307-318 [FREE Full text] [doi: [10.1007/s10461-012-0326-7](https://doi.org/10.1007/s10461-012-0326-7)] [Medline: [23108721](https://pubmed.ncbi.nlm.nih.gov/23108721/)]
49. Jankowska-Polanska B, Uchmanowicz I, Chudiak A, Dudek K, Morisky DE, Szymanska-Chabowska A. Psychometric properties of the Polish version of the eight-item Morisky Medication Adherence Scale in hypertensive adults. *Patient Prefer Adherence* 2016;10:1759-1766 [FREE Full text] [doi: [10.2147/PPA.S101904](https://doi.org/10.2147/PPA.S101904)] [Medline: [27672314](https://pubmed.ncbi.nlm.nih.gov/27672314/)]
50. Svarstad BL, Chewning BA, Sleath BL, Claesson C. The Brief Medication Questionnaire: a tool for screening patient adherence and barriers to adherence. *Patient Educ Couns* 1999 Jun;37(2):113-124. [doi: [10.1016/s0738-3991\(98\)00107-4](https://doi.org/10.1016/s0738-3991(98)00107-4)] [Medline: [14528539](https://pubmed.ncbi.nlm.nih.gov/14528539/)]
51. Resnick B, Wehren L, Orwig D. Reliability and validity of the self-efficacy and outcome expectations for osteoporosis medication adherence scales. *Orthop Nurs* 2003;22(2):139-147. [doi: [10.1097/00006416-200303000-00012](https://doi.org/10.1097/00006416-200303000-00012)] [Medline: [12703398](https://pubmed.ncbi.nlm.nih.gov/12703398/)]
52. Spire B, Arnould B, Barbier F, Durant J, Gilquin J, Landman R, et al. Simplification and first validation of a short battery of patient questionnaires for clinical management of HIV-infected patients: The HIV-SQUAD (Symptom Quality of life Adherence) Questionnaire. *HIV Clin Trials* 2009;10(4):215-232. [doi: [10.1310/hct1004-215](https://doi.org/10.1310/hct1004-215)] [Medline: [19723610](https://pubmed.ncbi.nlm.nih.gov/19723610/)]
53. De Las Cuevas C, Peñate W. Psychometric properties of the eight-item Morisky Medication Adherence Scale (MMAS-8) in a psychiatric outpatient setting. *Int J Clin Health Psychol* 2015;15(2):121-129 [FREE Full text] [doi: [10.1016/j.ijchp.2014.11.003](https://doi.org/10.1016/j.ijchp.2014.11.003)] [Medline: [30487829](https://pubmed.ncbi.nlm.nih.gov/30487829/)]
54. Brooks CM, Richards JM, Kohler CL, Soong SJ, Martin B, Windsor RA, et al. Assessing adherence to asthma medication and inhaler regimens: a psychometric analysis of adult self-report scales. *Med Care* 1994 Mar;32(3):298-307. [doi: [10.1097/00005650-199403000-00008](https://doi.org/10.1097/00005650-199403000-00008)] [Medline: [8145604](https://pubmed.ncbi.nlm.nih.gov/8145604/)]
55. Tan C, Teng GG, Chong KJ, Cheung PP, Lim A, Wee HL, et al. Utility of the Morisky Medication Adherence Scale in gout: a prospective study. *Patient Prefer Adherence* 2016;10:2449-2457 [FREE Full text] [doi: [10.2147/PPA.S119719](https://doi.org/10.2147/PPA.S119719)] [Medline: [27980395](https://pubmed.ncbi.nlm.nih.gov/27980395/)]
56. Willey C, Redding C, Stafford J, Garfield F, Geletko S, Flanigan T, et al. Stages of change for adherence with medication regimens for chronic disease: development and validation of a measure. *Clin Ther* 2000 Jul;22(7):858-871. [doi: [10.1016/s0149-2918\(00\)80058-2](https://doi.org/10.1016/s0149-2918(00)80058-2)] [Medline: [10945512](https://pubmed.ncbi.nlm.nih.gov/10945512/)]
57. Lin C, Ou H, Nikoobakht M, Broström A, Årestedt K, Pakpour AH. Validation of the 5-Item Medication Adherence Report Scale in Older Stroke Patients in Iran. *J Cardiovasc Nurs* 2018;33(6):536-543. [doi: [10.1097/JCN.0000000000000488](https://doi.org/10.1097/JCN.0000000000000488)] [Medline: [29649015](https://pubmed.ncbi.nlm.nih.gov/29649015/)]
58. Osborn CY, Gonzalez JS. Measuring insulin adherence among adults with type 2 diabetes. *J Behav Med* 2016 Aug;39(4):633-641 [FREE Full text] [doi: [10.1007/s10865-016-9741-y](https://doi.org/10.1007/s10865-016-9741-y)] [Medline: [27062271](https://pubmed.ncbi.nlm.nih.gov/27062271/)]
59. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. *J Clin Hypertens (Greenwich)* 2008 May;10(5):348-354 [FREE Full text] [Medline: [18453793](https://pubmed.ncbi.nlm.nih.gov/18453793/)]
60. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care* 1986 Jan;24(1):67-74. [Medline: [3945130](https://pubmed.ncbi.nlm.nih.gov/3945130/)]
61. Rofail D, Abetz L, Viala M, Gait C, Baladi J, Payne K. Satisfaction and adherence in patients with iron overload receiving iron chelation therapy as assessed by a newly developed patient instrument. *Value Health* 2009;12(1):109-117 [FREE Full text] [doi: [10.1111/j.1524-4733.2008.00390.x](https://doi.org/10.1111/j.1524-4733.2008.00390.x)] [Medline: [18637142](https://pubmed.ncbi.nlm.nih.gov/18637142/)]
62. Shin D, Kim C. Psychometric evaluation of a Korean version of the 8-item Medication Adherence Scale in rural older adults with hypertension. *Aust J Rural Health* 2013 Dec;21(6):336-342. [doi: [10.1111/ajr.12070](https://doi.org/10.1111/ajr.12070)] [Medline: [24299439](https://pubmed.ncbi.nlm.nih.gov/24299439/)]
63. De GS, Abraham I, Gemoets H, Evers G. Development of the long-term medication behaviour self-efficacy scale: qualitative study for item development. *J Adv Nurs* 1994 Feb;19(2):233-238. [Medline: [8188953](https://pubmed.ncbi.nlm.nih.gov/8188953/)]

64. de Klerk E, van der Heijde D, van der Tempel H, van der Linden S. Development of a questionnaire to investigate patient compliance with antirheumatic drug therapy. *J Rheumatol* 1999 Dec;26(12):2635-2641. [Medline: [10606375](#)]
65. Blumberg EJ, Hovell MF, Kelley NJ, Vera AY, Sipan CL, Berg JP. Self-report INH adherence measures were reliable and valid in Latino adolescents with latent tuberculosis infection. *J Clin Epidemiol* 2005 Jun;58(6):645-648. [doi: [10.1016/j.jclinepi.2004.11.022](#)] [Medline: [15878479](#)]
66. Unni EJ, Farris KB. Development of a new scale to measure self-reported medication nonadherence. *Res Social Adm Pharm* 2015;11(3):e133-e143. [doi: [10.1016/j.sapharm.2009.06.005](#)] [Medline: [21272524](#)]
67. Unni EJ, Olson JL, Farris KB. Revision and validation of Medication Adherence Reasons Scale (MAR-Scale). *Curr Med Res Opin* 2014 Feb;30(2):211-221. [doi: [10.1185/03007995.2013.851075](#)] [Medline: [24102296](#)]
68. Unni EJ, Sternbach N, Goren A. Using the Medication Adherence Reasons Scale (MAR-Scale) to identify the reasons for non-adherence across multiple disease conditions. *Patient Prefer Adherence* 2019;13:993-1004 [FREE Full text] [doi: [10.2147/PPA.S205359](#)] [Medline: [31308635](#)]
69. Salt E, Hall L, Peden AR, Home R. Psychometric properties of three medication adherence scales in patients with rheumatoid arthritis. *J Nurs Meas* 2012;20(1):59-72. [doi: [10.1891/1061-3749.20.1.59](#)] [Medline: [22679710](#)]
70. Tommelein E, Mehuys E, Van Tongelen I, Brusselle G, Boussery K. Accuracy of the Medication Adherence Report Scale (MARS-5) as a quantitative measure of adherence to inhalation medication in patients with COPD. *Ann Pharmacother* 2014 May;48(5):589-595. [doi: [10.1177/1060028014522982](#)] [Medline: [24523393](#)]
71. Alhomoud F, Alhomoud F, Millar I. How effectively are your patients taking their medicines? A critical review of the Strathclyde Compliance Risk Assessment Tool in relation to the 'MMAS' and 'MARS'. *J Eval Clin Pract* 2016 Jun;22(3):411-420. [doi: [10.1111/jep.12501](#)] [Medline: [26696012](#)]
72. Vale FC, Santa-Helena ETD, Santos MA, Carvalho WMDES, Menezes PR, Basso CR, et al. Development and validation of the WebAd-Q Questionnaire to monitor adherence to HIV therapy. *Rev Saude Publica* 2018;52:62 [FREE Full text] [doi: [10.11606/s1518-8787.2018052000337](#)] [Medline: [29846437](#)]
73. Cinar FI, Cinar M, Yilmaz S, Acikel C, Erdem H, Pay S, et al. Cross-Cultural Adaptation, Reliability, and Validity of the Turkish Version of the Compliance Questionnaire on Rheumatology in Patients With Behçet's Disease. *J Transcult Nurs* 2016 Sep;27(5):480-486. [doi: [10.1177/1043659615577699](#)] [Medline: [25801762](#)]
74. Alsolami FJ, Hou X, Correa-Velez I. An Arabic instrument to Measure Medication Adherence in Saudi Hypertensive Patients. *ME-JFM* 2013 Sep;11(7):17-23. [doi: [10.5742/mefm.2014.92409](#)]
75. Ortega Suárez FJ, Sánchez Plumed J, Pérez Valentín MA, Pereira Palomo P, Muñoz Cepeda MA, Lorenzo Aguiar D, Grupo de Estudio Vatreñ. Validation on the simplified medication adherence questionnaire (SMAQ) in renal transplant patients on tacrolimus. *Nefrologia* 2011;31(6):690-696 [FREE Full text] [doi: [10.3265/Nefrologia.pre2011.Aug.10973](#)] [Medline: [22130285](#)]
76. Julian FS, Martin P, Erickson SR. Validation of the Special Projects of National Significance adherence tool in HIV/AIDS patients. *Ann Pharmacother* 2010 Jun;44(6):1003-1009. [doi: [10.1345/aph.1M690](#)] [Medline: [20442352](#)]
77. Fabbrini G, Abbruzzese G, Barone P, Antonini A, Tinazzi M, Castegnaro G, REASON study group. Adherence to anti-Parkinson drug therapy in the. *Neurol Sci* 2013 Nov;34(11):2015-2022. [doi: [10.1007/s10072-013-1438-1](#)] [Medline: [23728715](#)]
78. Fond G, Boyer L, Boucekin M, Aden LA, Schürhoff F, Tessier A, FACE-SZ (FondaMental Academic Centers of Expertise for Schizophrenia) group. Validation study of the Medication Adherence Rating Scale. Results from the FACE-SZ national dataset. *Schizophr Res* 2017 Apr;182:84-89. [doi: [10.1016/j.schres.2016.10.023](#)] [Medline: [27789187](#)]
79. Owie GO, Olotu SO, James BO. Reliability and validity of the Medication Adherence Rating Scale in a cohort of patients with schizophrenia from Nigeria. *Trends Psychiatry Psychother* 2018;40(2):85-92 [FREE Full text] [doi: [10.1590/2237-6089-2017-0077](#)] [Medline: [29768528](#)]
80. Ogedegbe G, Mancuso CA, Allegrante JP, Charlson ME. Development and evaluation of a medication adherence self-efficacy scale in hypertensive African-American patients. *J Clin Epidemiol* 2003 Jun;56(6):520-529. [Medline: [12873646](#)]
81. Cate H, Bhattacharya D, Clark A, Holland R, Broadway DC. A comparison of measures used to describe adherence to glaucoma medication in a randomised controlled trial. *Clin Trials* 2015 Dec;12(6):608-617. [doi: [10.1177/1740774515592636](#)] [Medline: [26183437](#)]
82. Jónsdóttir H, Opjordsmoen S, Birkenaes AB, Engh JA, Ringen PA, Vaskinn A, et al. Medication adherence in outpatients with severe mental disorders: relation between self-reports and serum level. *J Clin Psychopharmacol* 2010 Apr;30(2):169-175. [doi: [10.1097/JCP.0b013e3181d2191e](#)] [Medline: [20520290](#)]
83. Cohen HW, Shmukler C, Ullman R, Rivera CM, Walker EA. Measurements of medication adherence in diabetic patients with poorly controlled HbA(1c). *Diabet Med* 2010 Feb;27(2):210-216 [FREE Full text] [doi: [10.1111/j.1464-5491.2009.02898.x](#)] [Medline: [20546266](#)]
84. Arnet I, Metaxas C, Walter PN, Morisky DE, Hersberger KE. The 8-item Morisky Medication Adherence Scale translated in German and validated against objective and subjective polypharmacy adherence measures in cardiovascular patients. *J Eval Clin Pract* 2015 Apr;21(2):271-277. [doi: [10.1111/jep.12303](#)] [Medline: [25558796](#)]
85. Sadakathulla I, Mateti U, Kellarai A, Bhat K. Adhering to antihypertensive treatment is vitally important. *Salud(i)Ciencia* 2019;23(4):314-324 [FREE Full text]



86. Zschocke I, Mrowietz U, Lotzin A, Karakasili E, Reich K. Assessing adherence factors in patients under topical treatment: development of the Topical Therapy Adherence Questionnaire (TTAQ). *Arch Dermatol Res* 2014 Apr;306(3):287-297 [FREE Full text] [doi: [10.1007/s00403-014-1446-x](https://doi.org/10.1007/s00403-014-1446-x)] [Medline: [24509981](https://pubmed.ncbi.nlm.nih.gov/24509981/)]
87. Kennedy SB. Developing a self-administered tool to predict adherence to antiretroviral therapy: design, method, and objectives. *AIDS Patient Care STDS* 2000 Jun;14(6):309-316. [doi: [10.1089/10872910050046331](https://doi.org/10.1089/10872910050046331)] [Medline: [10897503](https://pubmed.ncbi.nlm.nih.gov/10897503/)]
88. Prado JC, Kupek E, Mion D. Validity of four indirect methods to measure adherence in primary care hypertensives. *J Hum Hypertens* 2007 Jul;21(7):579-584. [doi: [10.1038/sj.jhh.1002196](https://doi.org/10.1038/sj.jhh.1002196)] [Medline: [17443212](https://pubmed.ncbi.nlm.nih.gov/17443212/)]
89. George J, Vuong T, Bailey MJ, Kong DC, Marriott JL, Stewart K. Medication Regimen Complexity and Adherence in Patients at Risk of Medication Misadventure. *Journal of Pharmacy Practice and Research* 2015 Apr 13;36(2):99-102. [doi: [10.1002/j.2055-2335.2006.tb00580.x](https://doi.org/10.1002/j.2055-2335.2006.tb00580.x)]
90. Kim J, Lee W, Hong Y, Ryu W, Lee KJ, Lee W, et al. Psychometric properties of a short self-reported measure of medication adherence among patients with hypertension treated in a busy clinical setting in Korea. *J Epidemiol* 2014;24(2):132-140 [FREE Full text] [doi: [10.2188/jea.je20130064](https://doi.org/10.2188/jea.je20130064)] [Medline: [24463958](https://pubmed.ncbi.nlm.nih.gov/24463958/)]
91. Koschack J, Marx G, Schnakenberg J, Kochen MM, Himmel W. Comparison of two self-rating instruments for medication adherence assessment in hypertension revealed insufficient psychometric properties. *J Clin Epidemiol* 2010 Mar;63(3):299-306. [doi: [10.1016/j.jclinepi.2009.06.011](https://doi.org/10.1016/j.jclinepi.2009.06.011)] [Medline: [19762213](https://pubmed.ncbi.nlm.nih.gov/19762213/)]
92. Cohen JL, Mann DM, Wisnivesky JP, Home R, Leventhal H, Musumeci-Szabó TJ, et al. Assessing the validity of self-reported medication adherence among inner-city asthmatic adults: the Medication Adherence Report Scale for Asthma. *Ann Allergy Asthma Immunol* 2009 Oct;103(4):325-331. [Medline: [19852197](https://pubmed.ncbi.nlm.nih.gov/19852197/)]
93. Simoni JM, Huh D, Wang Y, Wilson IB, Reynolds NR, Remien RH, et al. The validity of self-reported medication adherence as an outcome in clinical trials of adherence-promotion interventions: Findings from the MACH14 study. *AIDS Behav* 2014 Dec;18(12):2285-2290 [FREE Full text] [doi: [10.1007/s10461-014-0905-x](https://doi.org/10.1007/s10461-014-0905-x)] [Medline: [25280447](https://pubmed.ncbi.nlm.nih.gov/25280447/)]
94. Nordmann J, Denis P, Vigneux M, Trudeau E, Guillemin I, Berdeaux G. Development of the conceptual framework for the Eye-Drop Satisfaction Questionnaire (EDSQ) in glaucoma using a qualitative study. *BMC Health Serv Res* 2007 Aug 06;7:124 [FREE Full text] [doi: [10.1186/1472-6963-7-124](https://doi.org/10.1186/1472-6963-7-124)] [Medline: [17683594](https://pubmed.ncbi.nlm.nih.gov/17683594/)]
95. Risser J, Jacobson TA, Kripalani S. Development and psychometric evaluation of the Self-efficacy for Appropriate Medication Use Scale (SEAMS) in low-literacy patients with chronic disease. *J Nurs Meas* 2007;15(3):203-219. [Medline: [18232619](https://pubmed.ncbi.nlm.nih.gov/18232619/)]
96. van den Boogaard J, Lyimo RA, Boeree MJ, Kibiki GS, Aarnoutse RE. Electronic monitoring of treatment adherence and validation of alternative adherence measures in tuberculosis patients: a pilot study. *Bull World Health Organ* 2011 Sep 01;89(9):632-639 [FREE Full text] [doi: [10.2471/BLT.11.086462](https://doi.org/10.2471/BLT.11.086462)] [Medline: [21897483](https://pubmed.ncbi.nlm.nih.gov/21897483/)]
97. Lee JY, Lee SY, Hahn HJ, Son IJ, Hahn SG, Lee EB. Cultural adaptation of a compliance questionnaire for patients with rheumatoid arthritis to a Korean version. *Korean J Intern Med* 2011 Mar;26(1):28-33 [FREE Full text] [doi: [10.3904/kjim.2011.26.1.28](https://doi.org/10.3904/kjim.2011.26.1.28)] [Medline: [21437159](https://pubmed.ncbi.nlm.nih.gov/21437159/)]
98. Jansà M, Vidal M, Giménez M, Conget I, Galindo M, Roca D, et al. Psychometric analysis of the Spanish and Catalan versions of the Diabetes Self-Care inventory-revised version questionnaire. *Patient Prefer Adherence* 2013;7:997-1005 [FREE Full text] [doi: [10.2147/PPA.S50271](https://doi.org/10.2147/PPA.S50271)] [Medline: [24124352](https://pubmed.ncbi.nlm.nih.gov/24124352/)]
99. Chaiyachati K, Hirschhorn LR, Tanser F, Newell M, Barnighausen T. Validating five questions of antiretroviral nonadherence in a public-sector treatment program in rural South Africa. *AIDS Patient Care STDS* 2011 Mar;25(3):163-170 [FREE Full text] [doi: [10.1089/apc.2010.0257](https://doi.org/10.1089/apc.2010.0257)] [Medline: [21269131](https://pubmed.ncbi.nlm.nih.gov/21269131/)]
100. Wickersham KE, Sereika SM, Kang H, Tamres LK, Erlen JA. Use of a Self-Report Medication Adherence Scale for Measuring Adherence to Antiretroviral Therapy in Patients With HIV/AIDS. *J Nurs Meas* 2018 Aug;26(2):E72-E88. [doi: [10.1891/1061-3749.26.2.E72](https://doi.org/10.1891/1061-3749.26.2.E72)] [Medline: [30567952](https://pubmed.ncbi.nlm.nih.gov/30567952/)]
101. Kosilov K, Loparev S, Kuzina I, Kosilova L, Ivanovskaya M, Prokofyeva A. Effectiveness of a new tool for self-evaluation of adherence to antimuscarinic drug treatment in older patients of both sexes with urge incontinence. *Geriatr Gerontol Int* 2018 Jan;18(1):115-122. [doi: [10.1111/ggi.13150](https://doi.org/10.1111/ggi.13150)] [Medline: [28921819](https://pubmed.ncbi.nlm.nih.gov/28921819/)]
102. Kosilov K, Loparev S, Kuzina I, Shakirova O, Zhuravskaya N, Lobodenko A. Self-assessment of treatment compliance with antimuscarinic drugs and lower urinary tract condition among women with urinary incontinence. *Int Urogynecol J* 2017 Nov;28(11):1663-1669. [doi: [10.1007/s00192-017-3333-4](https://doi.org/10.1007/s00192-017-3333-4)] [Medline: [28429053](https://pubmed.ncbi.nlm.nih.gov/28429053/)]
103. Kosilov K, Loparev S, Kuzina I, Shakirova O, Zhuravskaya N, Lobodenko A. The effective tool for self-assessment of adherence to treatment in patients with benign prostatic obstruction and overactive bladder symptoms. *Aging Male* 2017 Mar;20(1):39-44. [doi: [10.1080/13685538.2016.1247435](https://doi.org/10.1080/13685538.2016.1247435)] [Medline: [28074677](https://pubmed.ncbi.nlm.nih.gov/28074677/)]
104. Walewski KM, Cicutto L, D'Urzo AD, Heslegrave RJ, Chapman KR. Evaluation of a questionnaire to assess compliance with anti-asthma medications. *J Asthma* 2004 Feb;41(1):77-83. [doi: [10.1081/jas-120026064](https://doi.org/10.1081/jas-120026064)] [Medline: [15046381](https://pubmed.ncbi.nlm.nih.gov/15046381/)]
105. Reynolds K, Viswanathan HN, O'Malley CD, Muntner P, Harrison TN, Cheetham TC, et al. Psychometric properties of the Osteoporosis-specific Morisky Medication Adherence Scale in postmenopausal women with osteoporosis newly treated with bisphosphonates. *Ann Pharmacother* 2012 May;46(5):659-670. [doi: [10.1345/aph.1Q652](https://doi.org/10.1345/aph.1Q652)] [Medline: [22510666](https://pubmed.ncbi.nlm.nih.gov/22510666/)]
106. Reynolds K, Viswanathan HN, Muntner P, Harrison TN, Cheetham TC, Hsu JY, et al. Validation of the Osteoporosis-Specific Morisky Medication Adherence Scale in long-term users of bisphosphonates. *Qual Life Res* 2014 Sep;23(7):2109-2120. [doi: [10.1007/s11136-014-0662-3](https://doi.org/10.1007/s11136-014-0662-3)] [Medline: [24604077](https://pubmed.ncbi.nlm.nih.gov/24604077/)]

107. Thompson K, Kulkarni J, Sergejew AA. Reliability and validity of a new Medication Adherence Rating Scale (MARS) for the psychoses. *Schizophr Res* 2000 May 5;42(3):241-247. [Medline: [10785582](#)]
108. Kosilov KV, Loparev SA, Kuzina IG, Shakirova OV, Gainullina YI, Kosilova LV, et al. A new tool for self-evaluation of adherence to antimuscarinic drugs treatment in patients with urinary incontinence. *Arab J Urol* 2017 Dec;15(4):372-379 [FREE Full text] [doi: [10.1016/j.aju.2017.10.007](#)] [Medline: [29234543](#)]
109. Zemmour K, Tinland A, Boucekine M, Girard V, Loubière S, Resseguier N, French Housing First Study Group. Validation of the Medication Adherence Rating Scale in homeless patients with schizophrenia: Results from the French Housing First experience. *Sci Rep* 2016 Aug 18;6:31598 [FREE Full text] [doi: [10.1038/srep31598](#)] [Medline: [27534796](#)]
110. Kyngäs HA, Skaar-Chandler CA, Duffy ME. The development of an instrument to measure the compliance of adolescents with a chronic disease. *J Adv Nurs* 2000 Dec;32(6):1499-1506. [doi: [10.1046/j.1365-2648.2000.01611.x](#)] [Medline: [11136419](#)]
111. Fialko L, Garety PA, Kuipers E, Dunn G, Bebbington PE, Fowler D, et al. A large-scale validation study of the Medication Adherence Rating Scale (MARS). *Schizophr Res* 2008 Mar;100(1-3):53-59. [doi: [10.1016/j.schres.2007.10.029](#)] [Medline: [18083007](#)]
112. Guénette L, Moisan J, Préville M, Boyer R. Measures of adherence based on self-report exhibited poor agreement with those based on pharmacy records. *J Clin Epidemiol* 2005 Sep;58(9):924-933. [doi: [10.1016/j.jclinepi.2005.02.002](#)] [Medline: [16085196](#)]
113. Lee L, El-Den S, Horne R, Carter SR. Patient satisfaction with information, concerns, beliefs and adherence to topical corticosteroids. *Patient Educ Couns* 2019 Jun;102(6):1203-1209. [doi: [10.1016/j.pec.2019.01.019](#)] [Medline: [30928342](#)]
114. McDonald-Miszczak L, Maris P, Fitzgibbon T, Ritchie G. A pilot study examining older adults' beliefs related to medication adherence: the BERMA survey. *J Aging Health* 2004 Nov;16(5):591-614. [doi: [10.1177/0898264304265772](#)] [Medline: [15448274](#)]
115. Nogueira-Silva L, Sá-Sousa A, Lima MJ, Monteiro A, Dennison-Himmelfarb C, Fonseca JA. Translation and cultural adaptation of the Hill-Bone Compliance to High Blood Pressure Therapy Scale to Portuguese. *Rev Port Cardiol* 2016 Feb;35(2):93-97 [FREE Full text] [doi: [10.1016/j.repc.2015.07.013](#)] [Medline: [26852304](#)]
116. Matza LS, Park J, Coyne KS, Skinner EP, Malley KG, Wolever RQ. Derivation and validation of the ASK-12 adherence barrier survey. *Ann Pharmacother* 2009 Oct;43(10):1621-1630. [doi: [10.1345/aph.1M174](#)] [Medline: [19776298](#)]
117. Lubinga SJ, Millar I, Babigumira JB. Pilot evaluation of the psychometric properties of a self-medication Risk Assessment Tool among elderly patients in a community setting. *BMC Res Notes* 2011 Oct 11;4:398 [FREE Full text] [doi: [10.1186/1756-0500-4-398](#)] [Medline: [21989334](#)]
118. Chesney MA, Ickovics JR, Chambers DB, Gifford AL, Neidig J, Zwickl B, PATIENT CARE COMMITTEE & ADHERENCE. Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: The AACTG Adherence Instruments. *AIDS Care* 2010 May 27;12(3):255-266. [doi: [10.1080/09540120050042891](#)]
119. Chisholm MA, Lance CE, Williamson GM, Mulloy LL. Development and validation of an immunosuppressant therapy adherence barrier instrument. *Nephrol Dial Transplant* 2005 Jan;20(1):181-188. [doi: [10.1093/ndt/gfh576](#)] [Medline: [15572384](#)]
120. Ali MAS, Abou-Taleb DAE, Mohamed RR. Treatment adherence and beliefs about medicines among Egyptian vitiligo patients. *Dermatol Ther* 2016 Nov;29(6):413-418. [doi: [10.1111/dth.12397](#)] [Medline: [27594551](#)]
121. Alsous M, Alhalaiqa F, Abu Farha R, Abdel Jalil M, McElnay J, Horne R. Reliability and validity of Arabic translation of Medication Adherence Report Scale (MARS) and Beliefs about Medication Questionnaire (BMQ)-specific for use in children and their parents. *PLoS One* 2017;12(2):e0171863 [FREE Full text] [doi: [10.1371/journal.pone.0171863](#)] [Medline: [28192467](#)]
122. Axelsson M, Ekerljung L, Lundbäck B, Lötvall J. Personality and unachieved treatment goals related to poor adherence to asthma medication in a newly developed adherence questionnaire - a population-based study. *Multidiscip Respir Med* 2016;11:42 [FREE Full text] [doi: [10.1186/s40248-016-0078-8](#)] [Medline: [27980735](#)]
123. DiBonaventura M, Wintfeld N, Huang J, Goren A. The association between nonadherence and glycated hemoglobin among type 2 diabetes patients using basal insulin analogs. *Patient Prefer Adherence* 2014;8:873-882 [FREE Full text] [doi: [10.2147/PPA.S55550](#)] [Medline: [24971002](#)]
124. Fahey M, Abdulmajeed A, Sabra K. Measurement of Adherence to Anti-Hypertensive Medication as Perceived by Doctors and Patients. *Qatar Medical Journal* 2006 Jun;2006(1):16. [doi: [10.5339/qmj.2006.1.16](#)]
125. Kikkert MJ, Koeter MWJ, Dekker JJM, Burti L, Robson D, Puschner B, et al. The predictive validity of subjective adherence measures in patients with schizophrenia. *Int J Methods Psychiatr Res* 2011 Jun;20(2):73-81 [FREE Full text] [doi: [10.1002/mpr.335](#)] [Medline: [21557378](#)]
126. Kleppe M, Lacroix J, Ham J, Midden C. The development of the ProMAS: a Probabilistic Medication Adherence Scale. *Patient Prefer Adherence* 2015;9:355-367 [FREE Full text] [doi: [10.2147/PPA.S76749](#)] [Medline: [25784791](#)]
127. Krousel-Wood M, Joyce C, Holt EW, Levitan EB, Dornelles A, Webber LS, et al. Development and evaluation of a self-report tool to predict low pharmacy refill adherence in elderly patients with uncontrolled hypertension. *Pharmacotherapy* 2013 Aug;33(8):798-811 [FREE Full text] [doi: [10.1002/phar.1275](#)] [Medline: [23649849](#)]



128. Krousel-Wood M, Jannu A, Re RN, Muntner P, Desalvo K. Reliability of a Medication Adherence Measure in an Outpatient Setting. *The American Journal of the Medical Sciences* 2005 Sep;330(3):128-133. [doi: [10.1097/00000441-200509000-00006](https://doi.org/10.1097/00000441-200509000-00006)]
129. Lu M, Saffren SA, Skolnik PR, Rogers WH, Coady W, Hardy H, et al. Optimal recall period and response task for self-reported HIV medication adherence. *AIDS Behav* 2008 Jan;12(1):86-94. [doi: [10.1007/s10461-007-9261-4](https://doi.org/10.1007/s10461-007-9261-4)] [Medline: [17577653](https://pubmed.ncbi.nlm.nih.gov/17577653/)]
130. Nakhaeizadeh M, Khalooei A. Psychometric Properties of Persian Version of the 8-item Morisky Medication Adherence Scale in Type 2 Diabetes Patients. *JCDR* 2019;YC14-YC18. [doi: [10.7860/jcdr/2019/41408.13076](https://doi.org/10.7860/jcdr/2019/41408.13076)]
131. Saffari M, Zeidi IM, Fridlund B, Chen H, Pakpour AH. A Persian Adaptation of Medication Adherence Self-Efficacy Scale (MASES) in Hypertensive Patients: Psychometric Properties and Factor Structure. *High Blood Press Cardiovasc Prev* 2015 Sep;22(3):247-255. [doi: [10.1007/s40292-015-0101-8](https://doi.org/10.1007/s40292-015-0101-8)] [Medline: [25986076](https://pubmed.ncbi.nlm.nih.gov/25986076/)]
132. Kim MT, Hill MN, Bone LR, Levine DM. Development and Testing of the Hill-Bone Compliance to High Blood Pressure Therapy Scale. *Progress in Cardiovascular Nursing* 2000 Jun;15(3):90-96. [doi: [10.1111/j.1751-7117.2000.tb00211.x](https://doi.org/10.1111/j.1751-7117.2000.tb00211.x)]
133. Mannheimer SB, Mukherjee R, Hirschhorn LR, Dougherty J, Celano SA, Ciccarone D, et al. The CASE adherence index: A novel method for measuring adherence to antiretroviral therapy. *AIDS Care* 2006 Oct;18(7):853-861 [FREE Full text] [doi: [10.1080/09540120500465160](https://doi.org/10.1080/09540120500465160)] [Medline: [16971298](https://pubmed.ncbi.nlm.nih.gov/16971298/)]
134. Duncan NA, Kronenberger WG, Roberson CP, Shapiro AD. VERITAS-PRN: a new measure of adherence to episodic treatment regimens in haemophilia. *Haemophilia* 2010 Jan;16(1):47-53. [doi: [10.1111/j.1365-2516.2009.02094.x](https://doi.org/10.1111/j.1365-2516.2009.02094.x)] [Medline: [19754842](https://pubmed.ncbi.nlm.nih.gov/19754842/)]
135. Duncan N, Kronenberger W, Roberson C, Shapiro A. VERITAS-Pro: a new measure of adherence to prophylactic regimens in haemophilia. *Haemophilia* 2010 Mar;16(2):247-255. [doi: [10.1111/j.1365-2516.2009.02129.x](https://doi.org/10.1111/j.1365-2516.2009.02129.x)] [Medline: [19925631](https://pubmed.ncbi.nlm.nih.gov/19925631/)]
136. Reynolds NR, Sun J, Nagaraja HN, Gifford AL, Wu AW, Chesney MA. Optimizing measurement of self-reported adherence with the ACTG Adherence Questionnaire: a cross-protocol analysis. *J Acquir Immune Defic Syndr* 2007 Dec 01;46(4):402-409. [Medline: [18077832](https://pubmed.ncbi.nlm.nih.gov/18077832/)]
137. van de Steeg N, Sielk M, Pentzek M, Bakx C, Altiner A. Drug-adherence questionnaires not valid for patients taking blood-pressure-lowering drugs in a primary health care setting. *J Eval Clin Pract* 2009 Jun;15(3):468-472. [doi: [10.1111/j.1365-2753.2008.01038.x](https://doi.org/10.1111/j.1365-2753.2008.01038.x)] [Medline: [19366389](https://pubmed.ncbi.nlm.nih.gov/19366389/)]
138. Sowunmi OA, Onifade PO. Psychometric evaluation of medication adherence rating scale (MARS) among Nigerian patients with schizophrenia. *Niger J Clin Pract* 2019 Sep;22(9):1281-1285 [FREE Full text] [doi: [10.4103/njcp.njcp\\_325\\_18](https://doi.org/10.4103/njcp.njcp_325_18)] [Medline: [31489867](https://pubmed.ncbi.nlm.nih.gov/31489867/)]
139. Kampman O, Lehtinen K, Lassila V. The reliability of compliance assessments performed by doctors and patients during neuroleptic treatment: a comparison of compliance ratings. *Acta Psychiatr Scand* 2001 Oct;104(4):299-304. [doi: [10.1034/j.1600-0447.2001.00159.x](https://doi.org/10.1034/j.1600-0447.2001.00159.x)] [Medline: [11722305](https://pubmed.ncbi.nlm.nih.gov/11722305/)]
140. Marsicano EDO, Fernandes NDS, Colugnati F, Grincenkov FRDS, Fernandes NMDS, De GS, et al. Transcultural adaptation and initial validation of Brazilian-Portuguese version of the Basel assessment of adherence to immunosuppressive medications scale (BAASIS) in kidney transplants. *BMC Nephrol* 2013 May 21;14:108 [FREE Full text] [doi: [10.1186/1471-2369-14-108](https://doi.org/10.1186/1471-2369-14-108)] [Medline: [23692889](https://pubmed.ncbi.nlm.nih.gov/23692889/)]
141. Pandey A, Raza F, Velasco A, Brinker S, Ayers C, Das SR, et al. Comparison of Morisky Medication Adherence Scale with therapeutic drug monitoring in apparent treatment-resistant hypertension. *J Am Soc Hypertens* 2015 Jun;9(6):420-426.e2. [doi: [10.1016/j.jash.2015.04.004](https://doi.org/10.1016/j.jash.2015.04.004)] [Medline: [26051923](https://pubmed.ncbi.nlm.nih.gov/26051923/)]
142. Sakthong P, Chabunthom R, Charoenvisuthiwongs R. Psychometric properties of the Thai version of the 8-item Morisky Medication Adherence Scale in patients with type 2 diabetes. *Ann Pharmacother* 2009 May;43(5):950-957. [doi: [10.1345/aph.1L453](https://doi.org/10.1345/aph.1L453)] [Medline: [19366872](https://pubmed.ncbi.nlm.nih.gov/19366872/)]
143. Choo PW, Rand CS, Inui TS, Lee ML, Cain E, Cordeiro-Breault M, et al. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. *Med Care* 1999 Sep;37(9):846-857. [Medline: [10493464](https://pubmed.ncbi.nlm.nih.gov/10493464/)]
144. Garcia-Marcos PW, Brand PLP, Kaptein AA, Klok T. Is the MARS questionnaire a reliable measure of medication adherence in childhood asthma? *J Asthma* 2016 Dec;53(10):1085-1089. [doi: [10.1080/02770903.2016.1180699](https://doi.org/10.1080/02770903.2016.1180699)] [Medline: [27177241](https://pubmed.ncbi.nlm.nih.gov/27177241/)]
145. Atsuta R, To Y, Sakamoto S, Mukai I, Kobayashi A, Kinoshita A, et al. Assessing usability of the. *Allergol Int* 2017 Jul;66(3):411-417 [FREE Full text] [doi: [10.1016/j.alit.2016.09.001](https://doi.org/10.1016/j.alit.2016.09.001)] [Medline: [27712949](https://pubmed.ncbi.nlm.nih.gov/27712949/)]
146. Ayiesah R, Leonard JH, Chong CY. Development and validation of non-adherence to pulmonary rehabilitation questionnaire: a clinical tool for patients with chronic obstructive pulmonary diseases. *Clin Ter* 2014;165(3):123-128. [doi: [10.7417/CT.2014.1708](https://doi.org/10.7417/CT.2014.1708)] [Medline: [24999563](https://pubmed.ncbi.nlm.nih.gov/24999563/)]
147. Pedrosa RBDS, Rodrigues RCM. Adaptation and evaluation of the measurement properties of the Brazilian version of the Self-efficacy for Appropriate Medication Adherence Scale. *Rev Lat Am Enfermagem* 2016;24:e2692 [FREE Full text] [doi: [10.1590/1518-8345.0167.2692](https://doi.org/10.1590/1518-8345.0167.2692)] [Medline: [27192417](https://pubmed.ncbi.nlm.nih.gov/27192417/)]
148. Rathbun RC, Farmer KC, Lockhart SM, Stephens JR. Validity of a stage of change instrument in assessing medication adherence in indigent patients with HIV infection. *Ann Pharmacother* 2007 Feb;41(2):208-214. [doi: [10.1345/aph.1H383](https://doi.org/10.1345/aph.1H383)] [Medline: [17213294](https://pubmed.ncbi.nlm.nih.gov/17213294/)]

149. Fernandes R, Wales S, Crisp J, Kyngas H. Modification and testing of the chronic disease compliance instrument to measure treatment compliance in adolescents with diabetes. *J Clin Nurs* 2011 May;20(9-10):1273-1281. [doi: [10.1111/j.1365-2702.2010.03566.x](https://doi.org/10.1111/j.1365-2702.2010.03566.x)] [Medline: [21492273](https://pubmed.ncbi.nlm.nih.gov/21492273/)]
150. Jacobsen R, Møldrup C, Christrup L, Sjøgren P, Hansen OB. The Danish version of the Medication Adherence Report Scale: preliminary validation in cancer pain patients. *Pain Pract* 2009;9(1):1-7. [doi: [10.1111/j.1533-2500.2008.00245.x](https://doi.org/10.1111/j.1533-2500.2008.00245.x)] [Medline: [19019056](https://pubmed.ncbi.nlm.nih.gov/19019056/)]
151. Sampaio R, Azevedo LF, Dias CC, Horne R, Castro Lopes JM. Portuguese version of the Medication Adherence Report Scale (MARS-9): Validation in a population of chronic pain patients. *J Eval Clin Pract* 2019 Apr;25(2):346-352. [doi: [10.1111/jep.13098](https://doi.org/10.1111/jep.13098)] [Medline: [30648328](https://pubmed.ncbi.nlm.nih.gov/30648328/)]
152. Pedrosa RBDS, Rodrigues RCM, Oliveira HC, Alexandre NMC. Construct Validity of the Brazilian Version of the Self-Efficacy for Appropriate Medication Adherence Scale. *J Nurs Meas* 2016 Apr 01;24(1):18-31. [doi: [10.1891/1061-3749.24.1.E18](https://doi.org/10.1891/1061-3749.24.1.E18)] [Medline: [29025523](https://pubmed.ncbi.nlm.nih.gov/29025523/)]
153. Kristina SA, Putri LR, Riani DA, Ikawati Z, Endarti D. VALIDITY OF SELF-REPORTED MEASURE OF MEDICATION ADHERENCE AMONG DIABETIC PATIENTS IN INDONESIA. *Int Res J Pharm* 2019 Aug 2;10(7):144-148. [doi: [10.7897/2230-8407.1007234](https://doi.org/10.7897/2230-8407.1007234)]
154. Shilbayeh SAR, Almutairi WA, Alyahya SA, Alshammari NH, Shaheen E, Adam A. Validation of knowledge and adherence assessment tools among patients on warfarin therapy in a Saudi hospital anticoagulant clinic. *Int J Clin Pharm* 2018 Feb;40(1):56-66. [doi: [10.1007/s11096-017-0569-5](https://doi.org/10.1007/s11096-017-0569-5)] [Medline: [29189976](https://pubmed.ncbi.nlm.nih.gov/29189976/)]
155. Kalichman SC, Amaral CM, Swetzes C, Jones M, Macy R, Kalichman MO, et al. A simple single-item rating scale to measure medication adherence: further evidence for convergent validity. *J Int Assoc Physicians AIDS Care (Chic)* 2009;8(6):367-374 [FREE Full text] [doi: [10.1177/1545109709352884](https://doi.org/10.1177/1545109709352884)] [Medline: [19952289](https://pubmed.ncbi.nlm.nih.gov/19952289/)]
156. Kalichman SC, Cain D, Fuhrel A, Eaton L, Di Fonzo K, Ertl T. Assessing medication adherence self-efficacy among low-literacy patients: development of a pictographic visual analogue scale. *Health Educ Res* 2005 Feb;20(1):24-35. [doi: [10.1093/her/cyg106](https://doi.org/10.1093/her/cyg106)] [Medline: [15253999](https://pubmed.ncbi.nlm.nih.gov/15253999/)]
157. Wilks SE, Spivey CA, Chisholm-Burns MA. Psychometric re-evaluation of the immunosuppressant therapy adherence scale among solid-organ transplant recipients. *J Eval Clin Pract* 2010 Feb;16(1):64-68. [doi: [10.1111/j.1365-2753.2008.01115.x](https://doi.org/10.1111/j.1365-2753.2008.01115.x)] [Medline: [20367816](https://pubmed.ncbi.nlm.nih.gov/20367816/)]
158. Fernandez S, Chaplin W, Schoenthaler AM, Ogedegbe G. Revision and validation of the medication adherence self-efficacy scale (MASES) in hypertensive African Americans. *J Behav Med* 2008 Dec;31(6):453-462 [FREE Full text] [doi: [10.1007/s10865-008-9170-7](https://doi.org/10.1007/s10865-008-9170-7)] [Medline: [18784996](https://pubmed.ncbi.nlm.nih.gov/18784996/)]
159. Bennett SJ, Milgrom LB, Champion V, Huster GA. Beliefs about medication and dietary compliance in people with heart failure: an instrument development study. *Heart Lung* 1997;26(4):273-279. [doi: [10.1016/s0147-9563\(97\)90084-4](https://doi.org/10.1016/s0147-9563(97)90084-4)] [Medline: [9257137](https://pubmed.ncbi.nlm.nih.gov/9257137/)]
160. Shalansky SJ, Levy AR, Ignaszewski AP. Self-reported Morisky score for identifying nonadherence with cardiovascular medications. *Ann Pharmacother* 2004 Sep;38(9):1363-1368. [doi: [10.1345/aph.1E071](https://doi.org/10.1345/aph.1E071)] [Medline: [15238622](https://pubmed.ncbi.nlm.nih.gov/15238622/)]
161. Koneru S, Shishov M, Ware A, Farhey Y, Mongey A, Graham TB, et al. Effectively measuring adherence to medications for systemic lupus erythematosus in a clinical setting. *Arthritis Rheum* 2007 Aug 15;57(6):1000-1006 [FREE Full text] [doi: [10.1002/art.22898](https://doi.org/10.1002/art.22898)] [Medline: [17665465](https://pubmed.ncbi.nlm.nih.gov/17665465/)]
162. Matza LS, Yu-Isenberg KS, Coyne KS, Park J, Wakefield J, Skinner EP, et al. Further testing of the reliability and validity of the ASK-20 adherence barrier questionnaire in a medical center outpatient population. *Current Medical Research and Opinion* 2008 Oct 14;24(11):3197-3206. [doi: [10.1185/03007990802463642](https://doi.org/10.1185/03007990802463642)]
163. Mueller S, Wilke T, Gorasso V, Erhart M, Kittner JM. Adaption and validation of the adherence barriers questionnaire for HIV patients on antiretroviral therapy (ABQ-HIV). *BMC Infect Dis* 2018 Nov 28;18(1):599 [FREE Full text] [doi: [10.1186/s12879-018-3530-x](https://doi.org/10.1186/s12879-018-3530-x)] [Medline: [30486795](https://pubmed.ncbi.nlm.nih.gov/30486795/)]
164. Müller S, Kohlmann T, Wilke T. Validation of the Adherence Barriers Questionnaire - an instrument for identifying potential risk factors associated with medication-related non-adherence. *BMC Health Serv Res* 2015 Apr 10;15:153 [FREE Full text] [doi: [10.1186/s12913-015-0809-0](https://doi.org/10.1186/s12913-015-0809-0)] [Medline: [25884193](https://pubmed.ncbi.nlm.nih.gov/25884193/)]
165. Okello S, Nasasira B, Muir ANW, Musingo A. Validity and Reliability of a Self-Reported Measure of Antihypertensive Medication Adherence in Uganda. *PLoS One* 2016;11(7):e0158499 [FREE Full text] [doi: [10.1371/journal.pone.0158499](https://doi.org/10.1371/journal.pone.0158499)] [Medline: [27367542](https://pubmed.ncbi.nlm.nih.gov/27367542/)]
166. Hahn SR, Park J, Skinner EP, Yu-Isenberg KS, Weaver MB, Crawford B, et al. Development of the ASK-20 adherence barrier survey. *Curr Med Res Opin* 2008 Jul;24(7):2127-2138. [doi: [10.1185/03007990802174769](https://doi.org/10.1185/03007990802174769)] [Medline: [18554431](https://pubmed.ncbi.nlm.nih.gov/18554431/)]
167. Smith SR, Wahed AS, Kelley SS, Conjeevaram HS, Robuck PR, Fried MW, Virahep-C Study Group. Assessing the validity of self-reported medication adherence in hepatitis C treatment. *Ann Pharmacother* 2007 Jul;41(7):1116-1123. [doi: [10.1345/aph.1K024](https://doi.org/10.1345/aph.1K024)] [Medline: [17519299](https://pubmed.ncbi.nlm.nih.gov/17519299/)]
168. Sidorkiewicz S, Tran V, Cousyn C, Perrodeau E, Ravaud P. Development and validation of an instrument to assess treatment adherence for each individual drug taken by a patient. *BMJ Open* 2016 May 10;6(5):e010510 [FREE Full text] [doi: [10.1136/bmjopen-2015-010510](https://doi.org/10.1136/bmjopen-2015-010510)] [Medline: [27165645](https://pubmed.ncbi.nlm.nih.gov/27165645/)]

169. Sriwarakorn S, Krittiyanunt S, Sakulbumrungsil R. Sensitivity and specificity of Thai-version brief medication questionnaire. *Journal of Health Research* 2010;24(3):129-134 [[FREE Full text](#)]
170. Stjernswärd S, Persson K, Nielsen R, Tuninger E, Levander S. A modified Drug Attitude Inventory used in long-term patients in sheltered housing. *Eur Neuropsychopharmacol* 2013 Oct;23(10):1296-1299. [doi: [10.1016/j.euroneuro.2012.11.011](#)] [Medline: [23265955](#)]
171. Ashur ST, Shamsuddin K, Shah SA, Bosseri S, Morisky DE. Reliability and known-group validity of the Arabic version of the 8-item Morisky Medication Adherence Scale among type 2 diabetes mellitus patients. *East Mediterr Health J* 2015 Dec 13;21(10):722-728 [[FREE Full text](#)] [doi: [10.26719/2015.21.10.722](#)] [Medline: [26750162](#)]
172. Wales S, Crisp J, Fernandes R, Kyngas H. Modification and testing of the chronic disease compliance instrument to measure treatment compliance in adolescents with asthma. *Contemp Nurse* 2011 Oct;39(2):147-156. [doi: [10.5172/conu.2011.147](#)] [Medline: [22551427](#)]
173. Kerr T, Hogg RS, Yip B, Tyndall MW, Montaner J, Wood E. Validity of self-reported adherence among injection drug users. *J Int Assoc Physicians AIDS Care (Chic)* 2008;7(4):157-159. [doi: [10.1177/1545109708320686](#)] [Medline: [18626123](#)]
174. Mumtaz T, Haider SA, Malik JA, La Greca AM. Translation, validation and effectiveness of self-care inventory in assessing adherence to diabetes treatment. *J Pak Med Assoc* 2016 Jul;66(7):853-858 [[FREE Full text](#)] [Medline: [27427135](#)]
175. Giordano TP, Guzman D, Clark R, Charlebois ED, Bangsberg DR. Measuring adherence to antiretroviral therapy in a diverse population using a visual analogue scale. *HIV Clin Trials* 2004;5(2):74-79. [doi: [10.1310/JFXH-G3X2-EYM6-D6UG](#)] [Medline: [15116282](#)]
176. Hogan TP, Awad AG, Eastwood R. A self-report scale predictive of drug compliance in schizophrenics: reliability and discriminative validity. *Psychol Med* 1983 Feb;13(1):177-183. [Medline: [6133297](#)]
177. Phillips T, Brittain K, Mellins CA, Zerbe A, Remien RH, Abrams EJ, et al. A Self-Reported Adherence Measure to Screen for Elevated HIV Viral Load in Pregnant and Postpartum Women on Antiretroviral Therapy. *AIDS Behav* 2017 Feb;21(2):450-461 [[FREE Full text](#)] [doi: [10.1007/s10461-016-1448-0](#)] [Medline: [27278548](#)]
178. Andy UU, Harvie HS, Smith AL, Probert KJ, Bogner HR, Arya LA. Validation of a self-administered instrument to measure adherence to anticholinergic drugs in women with overactive bladder. *Neurourol Urodyn* 2015 Jun;34(5):424-428 [[FREE Full text](#)] [doi: [10.1002/nau.22605](#)] [Medline: [24719232](#)]
179. Breuil V, Cortet B, Cotté F, Arnould B, Dias-Barbosa C, Gaudin A, et al. Validation of the adherence evaluation of osteoporosis treatment (ADEOS) questionnaire for osteoporotic post-menopausal women. *Osteoporos Int* 2012 Feb;23(2):445-455 [[FREE Full text](#)] [doi: [10.1007/s00198-011-1555-8](#)] [Medline: [21625889](#)]
180. Korb-Savoldelli V, Gillaizeau F, Pouchot J, Lenain E, Postel-Vinay N, Plouin P, et al. Validation of a French version of the 8-item Morisky medication adherence scale in hypertensive adults. *J Clin Hypertens (Greenwich)* 2012 Jul;14(7):429-434 [[FREE Full text](#)] [doi: [10.1111/j.1751-7176.2012.00634.x](#)] [Medline: [22747615](#)]
181. Vreeman RC, Nyandiko WM, Ayaya SO, Walumbe EG, Inui TS. Cognitive interviewing for cross-cultural adaptation of pediatric antiretroviral therapy adherence measurement items. *Int J Behav Med* 2014 Feb;21(1):186-196. [doi: [10.1007/s12529-012-9283-9](#)] [Medline: [23188670](#)]
182. Da W, Li X, Qiao S, Zhou Y, Shen Z. Evaluation of self-report adherence measures and their associations with detectable viral load among people living with HIV (PLHIV) in China. *PLoS One* 2018;13(8):e0203032 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0203032](#)] [Medline: [30161177](#)]
183. Doucette WR, Farris KB, Youland KM, Newland BA, Egerton SJ, Barnes JM. Development of the Drug Adherence Work-up (DRAW) tool. *J Am Pharm Assoc (2003)* 2012;52(6):e199-e204. [doi: [10.1331/JAPhA.2012.12001](#)] [Medline: [23945734](#)]
184. Lee W, Ahn J, Kim J, Hong Y, Hong SK, Kim YT, et al. Reliability and validity of a self-reported measure of medication adherence in patients with type 2 diabetes mellitus in Korea. *J Int Med Res* 2013 Aug;41(4):1098-1110. [doi: [10.1177/0300060513484433](#)] [Medline: [23860015](#)]
185. He W, Bonner A, Anderson D. Patient reported adherence to hypertension treatment: A revalidation study. *European Journal of Cardiovascular Nursing* 2015 Aug 25;15(2):150-156. [doi: [10.1177/1474515115603902](#)]
186. Kao Y, Liu Y. Compliance and schizophrenia: the predictive potential of insight into illness, symptoms, and side effects. *Compr Psychiatry* 2010;51(6):557-565. [doi: [10.1016/j.comppsy.2010.03.007](#)] [Medline: [20965300](#)]
187. Kim Y, Evangelista LS, Phillips LR, Pavlish C, Kopple JD. The End-Stage Renal Disease Adherence Questionnaire (ESRD-AQ): testing the psychometric properties in patients receiving in-center hemodialysis. *Nephrol Nurs J* 2010;37(4):377-393. [Medline: [20830945](#)]
188. Wang Y, Lee J, Toh MPHS, Tang WE, Ko Y. Validity and reliability of a self-reported measure of medication adherence in patients with Type 2 diabetes mellitus in Singapore. *Diabet Med* 2012 Sep;29(9):e338-e344. [doi: [10.1111/j.1464-5491.2012.03733.x](#)] [Medline: [22672497](#)]
189. Wang Y, Kong MC, Ko Y. Psychometric properties of the 8-item Morisky Medication Adherence Scale in patients taking warfarin. *Thromb Haemost* 2012 Oct;108(4):789-795. [doi: [10.1160/TH12-05-0368](#)] [Medline: [22836920](#)]
190. Garfield S, Eliasson L, Clifford S, Willson A, Barber N. Developing the Diagnostic Adherence to Medication Scale (the DAMS) for use in clinical practice. *BMC Health Serv Res* 2012 Oct 08;12:350 [[FREE Full text](#)] [doi: [10.1186/1472-6963-12-350](#)] [Medline: [23039138](#)]



191. Gökdoğan F, Kes D. Validity and reliability of the Turkish Adherence to Refills and Medications Scale. *Int J Nurs Pract* 2017 Oct;23(5):532-538. [doi: [10.1111/ijn.12566](https://doi.org/10.1111/ijn.12566)] [Medline: [28726281](https://pubmed.ncbi.nlm.nih.gov/28726281/)]
192. Kripalani S, Risser J, Gatti ME, Jacobson TA. Development and evaluation of the Adherence to Refills and Medications Scale (ARMS) among low-literacy patients with chronic disease. *Value Health* 2009;12(1):118-123 [FREE Full text] [doi: [10.1111/j.1524-4733.2008.00400.x](https://doi.org/10.1111/j.1524-4733.2008.00400.x)] [Medline: [19911444](https://pubmed.ncbi.nlm.nih.gov/19911444/)]
193. Chui E, Wong KL, Chan KY, Wong M. Validation study of the Brief Medication Adherence Scale (BMAS) in patients with schizophrenia and related disorders in Hong Kong. *Asian J Psychiatr* 2018 Oct;37:154-160. [doi: [10.1016/j.ajp.2018.09.005](https://doi.org/10.1016/j.ajp.2018.09.005)] [Medline: [30265964](https://pubmed.ncbi.nlm.nih.gov/30265964/)]
194. Cornelius T, Voils CI, Umland RC, Kronish IM. Validity Of The Self-Reported Domains Of Subjective Extent Of Nonadherence (DOSE-Nonadherence) Scale In Comparison With Electronically Monitored Adherence To Cardiovascular Medications. *Patient Prefer Adherence* 2019;13:1677-1684 [FREE Full text] [doi: [10.2147/PPA.S225460](https://doi.org/10.2147/PPA.S225460)] [Medline: [31631982](https://pubmed.ncbi.nlm.nih.gov/31631982/)]
195. Blalock DV, Zullig LL, Bosworth HB, Taylor SS, Voils CI. Self-reported medication nonadherence predicts cholesterol levels over time. *J Psychosom Res* 2019 Mar;118:49-55. [doi: [10.1016/j.jpsychores.2019.01.010](https://doi.org/10.1016/j.jpsychores.2019.01.010)] [Medline: [30782354](https://pubmed.ncbi.nlm.nih.gov/30782354/)]
196. Liao YW, Cheow C, Leung KTY, Tan H, Low SF, Cheen HHM, et al. A cultural adaptation and validation study of a self-report measure of the extent of and reasons for medication nonadherence among patients with diabetes in Singapore. *Patient Prefer Adherence* 2019;13:1241-1252 [FREE Full text] [doi: [10.2147/PPA.S208736](https://doi.org/10.2147/PPA.S208736)] [Medline: [31551654](https://pubmed.ncbi.nlm.nih.gov/31551654/)]
197. Fairley CK, Permana A, Read TRH. Long-term utility of measuring adherence by self-report compared with pharmacy record in a routine clinic setting. *HIV Med* 2005 Sep;6(5):366-369 [FREE Full text] [doi: [10.1111/j.1468-1293.2005.00322.x](https://doi.org/10.1111/j.1468-1293.2005.00322.x)] [Medline: [16156886](https://pubmed.ncbi.nlm.nih.gov/16156886/)]
198. Naqvi AA, AlShayban DM, Ghorri SA, Mahmoud MA, Haseeb A, Faidah HS, et al. Validation of the General Medication Adherence Scale in Saudi Patients With Chronic Diseases. *Front Pharmacol* 2019;10:633 [FREE Full text] [doi: [10.3389/fphar.2019.00633](https://doi.org/10.3389/fphar.2019.00633)] [Medline: [31231222](https://pubmed.ncbi.nlm.nih.gov/31231222/)]
199. Naqvi AA, Hassali MA, Jahangir A, Nadir MN, Kachela B. Translation and validation of the English version of the general medication adherence scale (GMAS) in patients with chronic illnesses. *J Drug Assess* 2019;8(1):36-42 [FREE Full text] [doi: [10.1080/21556660.2019.1579729](https://doi.org/10.1080/21556660.2019.1579729)] [Medline: [30863660](https://pubmed.ncbi.nlm.nih.gov/30863660/)]
200. Naqvi AA, Hassali MA, Rizvi M, Zehra A, Iffat W, Haseeb A, et al. Development and Validation of a Novel General Medication Adherence Scale (GMAS) for Chronic Illness Patients in Pakistan. *Front Pharmacol* 2018;9:1124 [FREE Full text] [doi: [10.3389/fphar.2018.01124](https://doi.org/10.3389/fphar.2018.01124)] [Medline: [30356775](https://pubmed.ncbi.nlm.nih.gov/30356775/)]
201. Athavale AS, Bentley JP, Banahan BF, McCaffrey DJ, Pace PF, Vorhies DW. Development of the medication adherence estimation and differentiation scale (MEDS). *Curr Med Res Opin* 2019 Apr;35(4):577-585. [doi: [10.1080/03007995.2018.1512478](https://doi.org/10.1080/03007995.2018.1512478)] [Medline: [30106315](https://pubmed.ncbi.nlm.nih.gov/30106315/)]
202. Lee S, Bae YH, Worley M, Law A. Validating the Modified Drug Adherence Work-Up (M-DRAW) Tool to Identify and Address Barriers to Medication Adherence. *Pharmacy (Basel)* 2017 Sep 08;5(3):50-52 [FREE Full text] [doi: [10.3390/pharmacy5030052](https://doi.org/10.3390/pharmacy5030052)] [Medline: [28970464](https://pubmed.ncbi.nlm.nih.gov/28970464/)]
203. Lee S, Bae-Shaaw YH, Gogineni H, Worley MM, Law AV. Triple strength utility of the Modified Drug Adherence Work-Up (M-DRAW) tool in a veterans affairs outpatient diabetes clinic. *Res Social Adm Pharm* 2019 Oct 01:914-920. [doi: [10.1016/j.sapharm.2019.09.063](https://doi.org/10.1016/j.sapharm.2019.09.063)] [Medline: [31629655](https://pubmed.ncbi.nlm.nih.gov/31629655/)]
204. Muñoz-Moreno JA, Fumaz CR, Ferrer MJ, Tuldrà A, Rovira T, Viladrich C, SERAD Validation Team. Assessing self-reported adherence to HIV therapy by questionnaire: the SERAD (Self-Reported Adherence) Study. *AIDS Res Hum Retroviruses* 2007 Oct;23(10):1166-1175. [doi: [10.1089/aid.2006.0120](https://doi.org/10.1089/aid.2006.0120)] [Medline: [17961100](https://pubmed.ncbi.nlm.nih.gov/17961100/)]
205. Knobel H, Alonso J, Casado JL, Collazos J, González J, Ruiz I, et al. Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study. *AIDS* 2002 Mar 8;16(4):605-613. [Medline: [11873004](https://pubmed.ncbi.nlm.nih.gov/11873004/)]
206. Wilson IB, Fowler FJ, Cosenza CA, Michaud J, Bentkover J, Rana A, et al. Cognitive and field testing of a new set of medication adherence self-report items for HIV care. *AIDS Behav* 2014 Dec;18(12):2349-2358 [FREE Full text] [doi: [10.1007/s10461-013-0610-1](https://doi.org/10.1007/s10461-013-0610-1)] [Medline: [24077970](https://pubmed.ncbi.nlm.nih.gov/24077970/)]
207. Wilson IB, Lee Y, Michaud J, Fowler FJ, Rogers WH. Validation of a New Three-Item Self-Report Measure for Medication Adherence. *AIDS Behav* 2016 Nov;20(11):2700-2708 [FREE Full text] [doi: [10.1007/s10461-016-1406-x](https://doi.org/10.1007/s10461-016-1406-x)] [Medline: [27098408](https://pubmed.ncbi.nlm.nih.gov/27098408/)]
208. Ueno H, Yamazaki Y, Yonekura Y, Park MJ, Ishikawa H, Kiuchi T. Reliability and validity of a 12-item medication adherence scale for patients with chronic disease in Japan. *BMC Health Serv Res* 2018 Jul 31;18(1):592 [FREE Full text] [doi: [10.1186/s12913-018-3380-7](https://doi.org/10.1186/s12913-018-3380-7)] [Medline: [30064422](https://pubmed.ncbi.nlm.nih.gov/30064422/)]
209. Hughes LD, Done J, Young A. A 5 item version of the Compliance Questionnaire for Rheumatology (CQR5) successfully identifies low adherence to DMARDs. *BMC Musculoskelet Disord* 2013 Oct 08;14:286 [FREE Full text] [doi: [10.1186/1471-2474-14-286](https://doi.org/10.1186/1471-2474-14-286)] [Medline: [24103582](https://pubmed.ncbi.nlm.nih.gov/24103582/)]
210. Schatz M, Zeiger RS, Yang S, Weinstein AG, Chen W, Saris-Baglama RN, et al. Development and preliminary validation of the Adult Asthma Adherence Questionnaire™. *J Allergy Clin Immunol Pract* 2013;1(3):280-288. [doi: [10.1016/j.jaip.2013.03.001](https://doi.org/10.1016/j.jaip.2013.03.001)] [Medline: [24565486](https://pubmed.ncbi.nlm.nih.gov/24565486/)]

211. Gabriel A, Violato C. Knowledge of and attitudes towards depression and adherence to treatment: the Antidepressant Adherence Scale (AAS). *J Affect Disord* 2010 Nov;126(3):388-394. [doi: [10.1016/j.jad.2010.07.013](https://doi.org/10.1016/j.jad.2010.07.013)] [Medline: [20708273](https://pubmed.ncbi.nlm.nih.gov/20708273/)]
212. Martins MJRV, Pinto AM, Castilho P, Macedo AF, Pereira AT, Bajouco M, et al. Assessing beliefs and attitudes towards antipsychotic medication from a recovery-based perspective: Psychometric properties of a new scale. *Psychiatry Res* 2019 Mar;273:325-330. [doi: [10.1016/j.psychres.2019.01.043](https://doi.org/10.1016/j.psychres.2019.01.043)] [Medline: [30677722](https://pubmed.ncbi.nlm.nih.gov/30677722/)]
213. Demirtaş A, Akbayrak N. Development of an assessment scale for treatment compliance in type 2 Diabetes Mellitus in Turkish population: Psychometric evaluation. *Int J Nurs Sci* 2017 Jul 10;4(3):244-251 [FREE Full text] [doi: [10.1016/j.ijnss.2017.06.002](https://doi.org/10.1016/j.ijnss.2017.06.002)] [Medline: [31406748](https://pubmed.ncbi.nlm.nih.gov/31406748/)]
214. Dolder CR, Lacro JP, Warren KA, Golshan S, Perkins DO, Jeste DV. Brief evaluation of medication influences and beliefs: development and testing of a brief scale for medication adherence. *J Clin Psychopharmacol* 2004 Aug;24(4):404-409. [doi: [10.1097/01.jcp.0000130554.63254.3a](https://doi.org/10.1097/01.jcp.0000130554.63254.3a)] [Medline: [15232332](https://pubmed.ncbi.nlm.nih.gov/15232332/)]
215. Tan J, Luo L, Zhang M, Chen H, Zhang D, Dong C, et al. A Chinese and Western medication adherence scale in patients with Chronic Kidney Disease. *Patient Prefer Adherence* 2019;13:1487-1495 [FREE Full text] [doi: [10.2147/PPA.S207693](https://doi.org/10.2147/PPA.S207693)] [Medline: [31507316](https://pubmed.ncbi.nlm.nih.gov/31507316/)]
216. Zheng J, Wang Y, Ye X, Xiao L, Ye J, Li X, et al. Validation of diabetes medication self-efficacy scale in Chinese with type 2 diabetes. *Patient Prefer Adherence* 2018;12:2517-2525 [FREE Full text] [doi: [10.2147/PPA.S170144](https://doi.org/10.2147/PPA.S170144)] [Medline: [30568430](https://pubmed.ncbi.nlm.nih.gov/30568430/)]
217. Kerr SJ, Avihingsanon A, Puthachoen O, Chetchotisakd P, Layton M, Ubolyam S, et al. Assessing adherence in Thai patients taking combination antiretroviral therapy. *Int J STD AIDS* 2012 Mar;23(3):160-165. [doi: [10.1258/ijsa.2009.009152](https://doi.org/10.1258/ijsa.2009.009152)] [Medline: [22581867](https://pubmed.ncbi.nlm.nih.gov/22581867/)]
218. Mehta SN, Nansel TR, Volkening LK, Butler DA, Haynie DL, Laffel LMB. Validation of a contemporary adherence measure for children with Type 1 diabetes: the Diabetes Management Questionnaire. *Diabet Med* 2015 Sep;32(9):1232-1238 [FREE Full text] [doi: [10.1111/dme.12682](https://doi.org/10.1111/dme.12682)] [Medline: [26280463](https://pubmed.ncbi.nlm.nih.gov/26280463/)]
219. Sleath B, Carpenter DM, Blalock SJ, Davis SA, Hickson RP, Lee C, et al. Development of a new diabetes medication self-efficacy scale and its association with both reported problems in using diabetes medications and self-reported adherence. *Patient Prefer Adherence* 2016;10:1003-1010 [FREE Full text] [doi: [10.2147/PPA.S101349](https://doi.org/10.2147/PPA.S101349)] [Medline: [27354769](https://pubmed.ncbi.nlm.nih.gov/27354769/)]
220. Vreeman RC, Scanlon ML, Tu W, Slaven JE, McAteer CI, Kerr SJ, et al. Validation of a self-report adherence measurement tool among a multinational cohort of children living with HIV in Kenya, South Africa and Thailand. *J Int AIDS Soc* 2019 May;22(5):e25304 [FREE Full text] [doi: [10.1002/jia2.25304](https://doi.org/10.1002/jia2.25304)] [Medline: [31148372](https://pubmed.ncbi.nlm.nih.gov/31148372/)]
221. Jesus-Nunes AP, Morais-de-Jesus M, Dantas-Duarte A, Moreira TM, Argolo FC, Castro ADO, et al. The Portuguese Version of the Immunosuppressant Therapy Adherence Scale (ITAS) among Liver Transplant Recipient Patients: Translation and Psychometric Properties. *Ann Hepatol* 2018;17(1):104-109 [FREE Full text] [doi: [10.5604/01.3001.0010.7541](https://doi.org/10.5604/01.3001.0010.7541)] [Medline: [29311393](https://pubmed.ncbi.nlm.nih.gov/29311393/)]
222. Mikhael EM, Hussain SA, Shawky N, Hassali MA. Validity and reliability of anti-diabetic medication adherence scale among patients with diabetes in Baghdad, Iraq: a pilot study. *BMJ Open Diabetes Res Care* 2019;7(1):e000658 [FREE Full text] [doi: [10.1136/bmjdr-2019-000658](https://doi.org/10.1136/bmjdr-2019-000658)] [Medline: [31354953](https://pubmed.ncbi.nlm.nih.gov/31354953/)]
223. Wohl DA, Panter AT, Kirby C, Magnus BE, Hudgens MG, Allmon AG, et al. Estimating HIV Medication Adherence and Persistence: Two Instruments for Clinical and Research Use. *AIDS Behav* 2018 Mar;22(3):948-960. [doi: [10.1007/s10461-017-1772-z](https://doi.org/10.1007/s10461-017-1772-z)] [Medline: [28447269](https://pubmed.ncbi.nlm.nih.gov/28447269/)]
224. Nelsen A, Trautner BW, Petersen NJ, Gupta S, Rodriguez-Barradas M, Giordano TP, et al. Development and validation of a measure for intention to adhere to HIV treatment. *AIDS Patient Care STDS* 2012 Jun;26(6):329-334 [FREE Full text] [doi: [10.1089/apc.2011.0318](https://doi.org/10.1089/apc.2011.0318)] [Medline: [22680281](https://pubmed.ncbi.nlm.nih.gov/22680281/)]
225. Gonzalez JS, Schneider HE, Wexler DJ, Psaros C, Delahanty LM, Cagliero E, et al. Validity of medication adherence self-reports in adults with type 2 diabetes. *Diabetes Care* 2013 Apr;36(4):831-837 [FREE Full text] [doi: [10.2337/dc12-0410](https://doi.org/10.2337/dc12-0410)] [Medline: [23204245](https://pubmed.ncbi.nlm.nih.gov/23204245/)]
226. Battistella M, Fleites R, Wong R, Jassal SV. Development, validation, and implementation of a medication adherence survey to seek a better understanding of the hemodialysis patient. *Clin Nephrol* 2016 Jan;85(1):12-22. [doi: [10.5414/CN108654](https://doi.org/10.5414/CN108654)] [Medline: [26636327](https://pubmed.ncbi.nlm.nih.gov/26636327/)]
227. Wicks P, Massagli M, Kulkarni A, Dastani H. Use of an online community to develop patient-reported outcome instruments: the Multiple Sclerosis Treatment Adherence Questionnaire (MS-TAQ). *J Med Internet Res* 2011;13(1):e12 [FREE Full text] [doi: [10.2196/jmir.1687](https://doi.org/10.2196/jmir.1687)] [Medline: [21266318](https://pubmed.ncbi.nlm.nih.gov/21266318/)]
228. Qi B, Resnick B. Reliability and validity of the Chinese versions of self-efficacy and outcome expectations for osteoporosis medication adherence scales in Chinese immigrants. *J Nurs Meas* 2014;22(3):472-488. [doi: [10.1891/1061-3749.22.3.472](https://doi.org/10.1891/1061-3749.22.3.472)] [Medline: [25608433](https://pubmed.ncbi.nlm.nih.gov/25608433/)]
229. Fredericksen RJ, Yang FM, Gibbons LE, Edwards TC, Brown S, Fitzsimmons E, et al. Development and content validation of measures assessing adherence barriers and behaviors for use in clinical care. *Res Social Adm Pharm* 2019 Sep;15(9):1168-1176. [doi: [10.1016/j.sapharm.2018.10.001](https://doi.org/10.1016/j.sapharm.2018.10.001)] [Medline: [30327183](https://pubmed.ncbi.nlm.nih.gov/30327183/)]



230. Almeida-Brasil CC, Nascimento ED, Silveira MR, Bonolo PDF, Ceccato MDGB. New patient-reported outcome measure to assess perceived barriers to antiretroviral therapy adherence: the PEDIA scale. *Cad Saude Publica* 2019;35(5):e00184218 [FREE Full text] [doi: [10.1590/0102-311x00184218](https://doi.org/10.1590/0102-311x00184218)] [Medline: [31166419](https://pubmed.ncbi.nlm.nih.gov/31166419/)]
231. Poveda V, Amado L, Filgueiras M, Teixeira L, Miranda V, Santos-Silva A, et al. End-stage renal disease adherence questionnaire: translation and validation to the portuguese language. *Ren Fail* 2016 Nov;38(10):1633-1638. [doi: [10.1080/0886022X.2016.1209063](https://doi.org/10.1080/0886022X.2016.1209063)] [Medline: [27764984](https://pubmed.ncbi.nlm.nih.gov/27764984/)]
232. Teixeira A, Oliveira C, Teixeira M, Rita Gaio A, Lobo JMS, de Almeida IFM, et al. Development and Validation of a Novel Questionnaire for Adherence with Topical Treatments in Psoriasis (QATOP). *Am J Clin Dermatol* 2017 Aug;18(4):571-581. [doi: [10.1007/s40257-017-0272-2](https://doi.org/10.1007/s40257-017-0272-2)] [Medline: [28321796](https://pubmed.ncbi.nlm.nih.gov/28321796/)]
233. Abe RY, Wen LDC, Barker GT, Mansberger SL. Psychometric Properties of the Glaucoma Treatment Compliance Assessment Tool (GTCAT) in a Brazilian Population. *J Glaucoma* 2018 Mar;27(3):257-265. [doi: [10.1097/IJG.0000000000000876](https://doi.org/10.1097/IJG.0000000000000876)] [Medline: [29369851](https://pubmed.ncbi.nlm.nih.gov/29369851/)]
234. Barker GT, Mansberger SL. Psychometric Properties of the Reduced Version of the Glaucoma Treatment Compliance Assessment Tool (GTCAT). *Ophthalmic Epidemiol* 2019 Feb;26(1):55-62. [doi: [10.1080/09286586.2018.1516785](https://doi.org/10.1080/09286586.2018.1516785)] [Medline: [30204034](https://pubmed.ncbi.nlm.nih.gov/30204034/)]
235. Kandrotaitė K, Smigelskas K, Janusauskiene D, Jievaltas M, Maciulaitis R, Briedis V. Development of a short questionnaire to identify the risk of nonadherence to antibiotic treatment. *Curr Med Res Opin* 2013 Nov;29(11):1555-1563. [doi: [10.1185/03007995.2013.835255](https://doi.org/10.1185/03007995.2013.835255)] [Medline: [23952367](https://pubmed.ncbi.nlm.nih.gov/23952367/)]
236. Daouphars M, Ouvry M, Lenain P, Rouvet J, Jardin F, Bubenheim M, et al. Preliminary validation of self-assessment tool to measure imatinib adherence in patients with chronic myeloid leukemia. *Pharmacotherapy* 2013 Feb;33(2):152-156. [doi: [10.1002/phar.1174](https://doi.org/10.1002/phar.1174)] [Medline: [23359430](https://pubmed.ncbi.nlm.nih.gov/23359430/)]
237. Buscher A, Hartman C, Kallen MA, Giordano TP. Validity of self-report measures in assessing antiretroviral adherence of newly diagnosed, HAART-naïve, HIV patients. *HIV Clin Trials* 2011;12(5):244-254 [FREE Full text] [doi: [10.1310/hct1205-244](https://doi.org/10.1310/hct1205-244)] [Medline: [22180522](https://pubmed.ncbi.nlm.nih.gov/22180522/)]
238. Erickson SR, Coombs JH, Kirking DM, Azimi AR. Compliance from self-reported versus pharmacy claims data with metered-dose inhalers. *Ann Pharmacother* 2001 Sep;35(9):997-1003. [doi: [10.1345/aph.10379](https://doi.org/10.1345/aph.10379)] [Medline: [11573875](https://pubmed.ncbi.nlm.nih.gov/11573875/)]
239. Godin G, Gagné C, Naccache H. Validation of a self-reported questionnaire assessing adherence to antiretroviral medication. *AIDS Patient Care STDS* 2003 Jul;17(7):325-332. [doi: [10.1089/108729103322231268](https://doi.org/10.1089/108729103322231268)] [Medline: [12952734](https://pubmed.ncbi.nlm.nih.gov/12952734/)]
240. Plaza V, Fernández-Rodríguez C, Melero C, Cosío BG, Entrenas LM, de LLP, et al. Validation of the 'Test of the Adherence to Inhalers' (TAI) for Asthma and COPD Patients. *J Aerosol Med Pulm Drug Deliv* 2016 Apr;29(2):142-152 [FREE Full text] [doi: [10.1089/jamp.2015.1212](https://doi.org/10.1089/jamp.2015.1212)] [Medline: [26230150](https://pubmed.ncbi.nlm.nih.gov/26230150/)]
241. Mancebo MC, Pinto A, Rasmussen SA, Eisen JL. Development of the Treatment Adherence Survey-patient version (TAS-P) for OCD. *J Anxiety Disord* 2008;22(1):32-43 [FREE Full text] [doi: [10.1016/j.janxdis.2007.01.009](https://doi.org/10.1016/j.janxdis.2007.01.009)] [Medline: [17324553](https://pubmed.ncbi.nlm.nih.gov/17324553/)]
242. Cella D, Gershon R, Lai J, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* 2007;16 Suppl 1:133-141. [doi: [10.1007/s11136-007-9204-6](https://doi.org/10.1007/s11136-007-9204-6)] [Medline: [17401637](https://pubmed.ncbi.nlm.nih.gov/17401637/)]
243. Kripalani S, Gatti ME, Jacobson TA. Association of age, health literacy, and medication management strategies with cardiovascular medication adherence. *Patient Educ Couns* 2010 Nov;81(2):177-181. [doi: [10.1016/j.pec.2010.04.030](https://doi.org/10.1016/j.pec.2010.04.030)] [Medline: [20684870](https://pubmed.ncbi.nlm.nih.gov/20684870/)]
244. Eindhoven DC, Hilt AD, Zwaan TC, Schalijs MJ, Borleffs CJW. Age and gender differences in medical adherence after myocardial infarction: Women do not receive optimal treatment - The Netherlands claims database. *Eur J Prev Cardiol* 2018 Jan;25(2):181-189. [doi: [10.1177/2047487317744363](https://doi.org/10.1177/2047487317744363)] [Medline: [29164916](https://pubmed.ncbi.nlm.nih.gov/29164916/)]
245. Lai J, Dineen K, Reeve BB, Von Roenn J, Shervin D, McGuire M, et al. An item response theory-based pain item bank can enhance measurement precision. *J Pain Symptom Manage* 2005 Sep;30(3):278-288 [FREE Full text] [doi: [10.1016/j.jpainsymman.2005.03.009](https://doi.org/10.1016/j.jpainsymman.2005.03.009)] [Medline: [16183012](https://pubmed.ncbi.nlm.nih.gov/16183012/)]
246. Haley SM, Ni P, Hambleton RK, Slavin MD, Jette AM. Computer adaptive testing improved accuracy and precision of scores over random item selection in a physical functioning item bank. *J Clin Epidemiol* 2006 Nov;59(11):1174-1182. [doi: [10.1016/j.jclinepi.2006.02.010](https://doi.org/10.1016/j.jclinepi.2006.02.010)] [Medline: [17027428](https://pubmed.ncbi.nlm.nih.gov/17027428/)]
247. Inui TS, Carter WB, Pecoraro RE, Pearlman RA, Dohan JJ. Variations in patient compliance with common long-term drugs. *Med Care* 1980 Oct;18(10):986-993. [doi: [10.1097/00005650-198010000-00002](https://doi.org/10.1097/00005650-198010000-00002)] [Medline: [7453267](https://pubmed.ncbi.nlm.nih.gov/7453267/)]
248. Piette JD, Heisler M, Ganoczy D, McCarthy JF, Valenstein M. Differential medication adherence among patients with schizophrenia and comorbid diabetes and hypertension. *Psychiatr Serv* 2007 Feb;58(2):207-212. [doi: [10.1176/ps.2007.58.2.207](https://doi.org/10.1176/ps.2007.58.2.207)] [Medline: [17287377](https://pubmed.ncbi.nlm.nih.gov/17287377/)]
249. Lai J, Cella D, Chang C, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Qual Life Res* 2003 Aug;12(5):485-501. [Medline: [13677494](https://pubmed.ncbi.nlm.nih.gov/13677494/)]
250. Dirven L, Groenvold M, Taphoorn MJB, Conroy T, Tomaszewski KA, Young T, EORTC Quality of Life Group. Psychometric evaluation of an item bank for computerized adaptive testing of the EORTC QLQ-C30 cognitive functioning dimension in cancer patients. *Qual Life Res* 2017 Nov;26(11):2919-2929 [FREE Full text] [doi: [10.1007/s11136-017-1648-8](https://doi.org/10.1007/s11136-017-1648-8)] [Medline: [28707048](https://pubmed.ncbi.nlm.nih.gov/28707048/)]

251. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009 Oct;62(10):e1-34 [FREE Full text] [doi: [10.1016/j.jclinepi.2009.06.006](https://doi.org/10.1016/j.jclinepi.2009.06.006)] [Medline: [19631507](https://pubmed.ncbi.nlm.nih.gov/19631507/)]
252. Evans JP, Smith A, Gibbons C, Alonso J, Valderas JM. The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS): a view from the UK. *Patient Relat Outcome Meas* 2018;9:345-352 [FREE Full text] [doi: [10.2147/PROM.S141378](https://doi.org/10.2147/PROM.S141378)] [Medline: [30498382](https://pubmed.ncbi.nlm.nih.gov/30498382/)]
253. Terwee CB, Prinsen CAC, Ricci GMG, Suman A, de VHCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res* 2016 Apr;25(4):767-779 [FREE Full text] [doi: [10.1007/s11136-015-1122-4](https://doi.org/10.1007/s11136-015-1122-4)] [Medline: [26346986](https://pubmed.ncbi.nlm.nih.gov/26346986/)]
254. Kubica A, Kosobucka A, Michalski P, Fabiszak T, Felsmann M. Self-reported questionnaires for assessment adherence to treatment in patients with cardiovascular diseases. *Medical Research Journal* 2018 Mar 30;2(4):115-122. [doi: [10.5603/MRJ.2017.0015](https://doi.org/10.5603/MRJ.2017.0015)]
255. Nguyen T, La CA, Cottrell N. What are validated self-report adherence scales really measuring?: a systematic review. *Br J Clin Pharmacol* 2014 Mar;77(3):427-445 [FREE Full text] [doi: [10.1111/bcp.12194](https://doi.org/10.1111/bcp.12194)] [Medline: [23803249](https://pubmed.ncbi.nlm.nih.gov/23803249/)]
256. Jin L, Acharya L. Cultural Beliefs Underlying Medication Adherence in People of Chinese Descent in the United States. *Health Commun* 2016;31(5):513-521. [doi: [10.1080/10410236.2014.974121](https://doi.org/10.1080/10410236.2014.974121)] [Medline: [26422467](https://pubmed.ncbi.nlm.nih.gov/26422467/)]
257. McQuaid EL, Landier W. Cultural Issues in Medication Adherence: Disparities and Directions. *J Gen Intern Med* 2018 Feb;33(2):200-206 [FREE Full text] [doi: [10.1007/s11606-017-4199-3](https://doi.org/10.1007/s11606-017-4199-3)] [Medline: [29204971](https://pubmed.ncbi.nlm.nih.gov/29204971/)]

## Abbreviations

**CAT:** computer-adaptive tests

**DOSE-Nonadherence:** Domains of Subjective Extent of Nonadherence

**IRT:** item response theory

**PROMs:** patient-reported outcome measures

**PRISMA:** Preferred reporting items for systematic review and meta-analysis

**PROMIS:** Patient-Reported Outcomes Measurement Information System

*Edited by G Eysenbach; submitted 07.04.20; peer-reviewed by P Nieuwkerk, S Wee, T Patel; comments to author 08.06.20; revised version received 12.06.20; accepted 14.06.20; published 08.10.20.*

*Please cite as:*

Kwan YH, Oo LJY, Loh DHF, Phang JK, Weng SD, Blalock DV, Chew EH, Yap KZ, Tan CYK, Yoon S, Fong W, Østbye T, Low LL, Bosworth HB, Thumboo J

*Development of an Item Bank to Measure Medication Adherence: Systematic Review*

*J Med Internet Res* 2020;22(10):e19089

URL: <https://www.jmir.org/2020/10/e19089>

doi: [10.2196/19089](https://doi.org/10.2196/19089)

PMID: [33030441](https://pubmed.ncbi.nlm.nih.gov/33030441/)

©Yu Heng Kwan, Livia Jia Yi Oo, Dionne Hui Fang Loh, Jie Kie Phang, Si Dun Weng, Dan V Blalock, Eng Hui Chew, Kai Zhen Yap, Corrinne Yong Koon Tan, Sungwon Yoon, Warren Fong, Truls Østbye, Lian Leng Low, Hayden Barry Bosworth, Julian Thumboo. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Review

# Measurement Properties of Existing Patient-Reported Outcome Measures on Medication Adherence: Systematic Review

Yu Heng Kwan<sup>1,2\*</sup>, PhD; Si Dun Weng<sup>2\*</sup>, BSc; Dionne Hui Fang Loh<sup>3\*</sup>, BEng; Jie Kie Phang<sup>4</sup>, BSc; Livia Jia Yi Oo<sup>2</sup>, BSc; Dan V Blalock<sup>5,6</sup>, PhD; Eng Hui Chew<sup>2</sup>, PhD; Kai Zhen Yap<sup>2</sup>, PhD; Corrinne Yong Koon Tan<sup>7</sup>, MBA; Sungwon Yoon<sup>1,3</sup>, PhD; Warren Fong<sup>4,8,9</sup>, FAMS; Truls Østbye<sup>1\*</sup>, PhD; Lian Leng Low<sup>3,10,11\*</sup>, FCFP(S); Hayden Barry Bosworth<sup>5,6,12,13\*</sup>, PhD; Julian Thumboo<sup>1,4,9\*</sup>, FRCP

<sup>1</sup>Program in Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore

<sup>2</sup>Department of Pharmacy, National University of Singapore, Singapore, Singapore

<sup>3</sup>PULSES Centre Grant, SingHealth Regional Health System, Singapore, Singapore

<sup>4</sup>Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore

<sup>5</sup>Center of Innovation to Accelerate Discovery and Practice Transformation (ADAPT), Durham Veterans Affairs Health Care System, Durham, NC, United States

<sup>6</sup>Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, United States

<sup>7</sup>Pharmacy Transformation Office, National Healthcare Group Pharmacy, Singapore, Singapore

<sup>8</sup>Duke-NUS Medical School, Singapore, Singapore

<sup>9</sup>NUS Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

<sup>10</sup>Department of Family Medicine and Continuing Care, Singapore General Hospital, Singapore, Singapore

<sup>11</sup>Post Acute and Continuing Care, Outram Community Hospital, Singapore, Singapore

<sup>12</sup>Department of Population Health Sciences, Duke University Medical Center, Durham, NC, United States

<sup>13</sup>School of Nursing, Duke University Medical Center, Durham, NC, United States

\* these authors contributed equally

## Corresponding Author:

Yu Heng Kwan, PhD

Program in Health Services and Systems Research,

Duke-NUS Medical School

8 College Road

Singapore, 169857

Singapore

Phone: 65 63778465

Email: [yuheng\\_02071988@hotmail.com](mailto:yuheng_02071988@hotmail.com)

## Abstract

**Background:** Medication adherence is essential for improving the health outcomes of patients. Various patient-reported outcome measures (PROMs) have been developed to measure medication adherence in patients. However, no study has summarized the psychometric properties of these PROMs to guide selection for use in clinical practice or research.

**Objective:** This study aims to evaluate the quality of the PROMs used to measure medication adherence.

**Methods:** This study was guided by the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) guidelines. Relevant articles were retrieved from the EMBASE, PubMed, Cochrane Library, Web of Science, and CINAHL (Cumulative Index to Nursing and Allied Health Literature) databases. The PROMs were then evaluated based on the CoConsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines.

**Results:** A total of 121 unique medication adherence PROMs from 214 studies were identified. *Hypotheses testing for construct validity* and *internal consistency* were the most frequently assessed measurement properties. PROMs with at least a *moderate* level of evidence for  $\geq 5$  measurement properties include the Adherence Starts with Knowledge 20, Compliance Questionnaire-Rheumatology, General Medication Adherence Scale, Hill-Bone Scale, Immunosuppressant Therapy Barrier Scale, Medication Adherence Reasons Scale (MAR-Scale) revised, 5-item Medication Adherence Rating Scale (MARS-5), 9-item MARS (MARS-9), 4-item Morisky Medication Adherence Scale (MMAS-4), 8-item MMAS (MMAS-8), Self-efficacy for

Appropriate Medication Adherence Scale, Satisfaction with Iron Chelation Therapy, Test of Adherence to Inhalers, and questionnaire by Voils. The MAR-Scale revised, MMAS-4, and MMAS-8 have been administered electronically.

**Conclusions:** This study identified 121 PROMs for medication adherence and provided synthesized evidence for the measurement properties of these PROMs. The findings from this study may assist clinicians and researchers in selecting suitable PROMs to assess medication adherence.

(*J Med Internet Res* 2020;22(10):e19179) doi:[10.2196/19179](https://doi.org/10.2196/19179)

## KEYWORDS

systematic review; reliability and validity; medication adherence; patient reported outcome measures

## Introduction

Medication adherence is known as “the degree to which the person’s behaviour corresponds with the agreed recommendations from a healthcare provider” [1]. Medication adherence is essential when it comes to improving the health outcomes of patients, especially for chronic diseases [2,3]. However, only approximately half of the patients worldwide adhere to their prescribed medication regimen [4]. Reasons for medication nonadherence include complexity of the treatment regimen, poor communication with health care providers, and concerns about side effects from taking medications [5]. Poor medication adherence may lead to worse health outcomes and higher rates of mortality and morbidity [1,6]. Nonadherence also incurs a high cost burden to the health care system by increasing hospital visits as well as causing unnecessary escalation to more expensive treatments [1]. Therefore, improving medication adherence is key to improving treatment outcomes [7,8].

To successfully improve medication adherence, there is a need for the accurate assessment of medication adherence. Current practices for measuring medication include direct measures such as drug assays of blood or urine as well as indirect measures of adherence such as pill count, electronic monitoring devices, and the use of big data such as review of prescription records and claims [9,10]. Some of these measures are not time efficient and are likely to be unsustainable in clinical practice. Various patient-reported outcome measures (PROMs) such as the Morisky Medication Adherence Scale (MMAS) [11], Hill-Bone Compliance Scale [12], and Medication Adherence Rating Scale (MARS) [13] have been developed to measure self-reported adherence to medications. These PROMs may be useful in clinical practice because they are easy to administer. On the basis of the patients’ PROM ratings, health care professionals may be able to provide timely feedback. Thus, underlying issues that contribute to medication noncompliance can be addressed at the point of care [14].

A number of previous studies have been conducted to validate PROMs on medication adherence [12,15-17], and a previous systematic review found 14 PROMs that assessed adherence to inhaled asthma maintenance medication alone [18]. However, to date, there is no comprehensive review that summarizes the psychometric properties of PROMs for medication adherence, which is essential to guide the selection of suitable PROMs to evaluate medication adherence in patients. Hence, we carried out a systematic review to identify studies that investigated

PROMs for medication adherence and to evaluate the quality of these PROMs.

## Methods

This study was conducted with reference to the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) statement [19]. The measurement properties of each PROM were evaluated based on the COnsensus-based Standards for the selection of health Measurement Instruments(COSMIN) guidelines [20,21]. The COSMIN guidelines evaluate PROM development and the following 9 measurement properties: “content validity,” “structural validity,” “internal consistency,” “cross-cultural validity/measurement invariance,” test-retest “reliability,” “measurement error,” “criterion validity,” “hypotheses testing for construct validity,” and “responsiveness” [21,22]. Of note, the assessment of “convergent validity,” “discriminant validity,” and “known-group validity” falls under “hypotheses testing for construct validity” [23,24]. We also assumed that “concurrent validity” and “predictive validity” can be evaluated by the same measurement property, “hypothesis testing for construct validity,” and sensitivity to change can be evaluated under “responsiveness” as well [22,25,26].

## Search Strategy

The EMBASE, PubMed, Cochrane Library, Web of Science, and CINAHL databases were searched for relevant studies published before November 1, 2019. A search strategy (Multimedia Appendix 1) consisting of adherence, PROMs, and measurement properties was used. Search filters created by Terwee et al [27], which consists of a combination of search terms, were also used to enhance the sensitivity of searches, where available. For the adherence construct, synonyms such as compliance were used in the search. Duplicates were removed from the final search library.

## Study Selection

Articles included in this study were full-text publications in English, validating medication adherence PROMs, with the assessment of at least one measurement property listed in the COSMIN [24,28]. We excluded animal studies, case studies, narrative reviews, expert opinions, conference abstracts, and PROMs completed by proxy [29]. In total, 2 independent reviewers (SW and LO) screened the titles and abstracts of the studies according to the inclusion and exclusion criteria. Opinions from a third reviewer (YK) were sought in the event of any disagreements. The remaining full-text articles were then



evaluated by the same 2 independent reviewers for inclusion and exclusion. Additional articles for inclusion into the final pool of articles were also identified from the reference list of articles during the full-text screening phase [30].

### Extraction of Data

The following data were then extracted from the included articles by 2 independent reviewers (SW and LO), where available:

- General characteristics of the study population: age, sample size, gender, and country.
- Characteristics of disease or condition: disease studied, duration of illness, or treatment.
- PROM characteristics: methods of administration, availability of electronic administration, language, response scale, domains, and number of items.

### Assessment of Methodological Quality

The methodological quality of the studies was assessed by 2 independent reviewers (SW and LO). Any disagreement was resolved in consultation with a third reviewer (JP). Each measurement property was assessed based on a 4-point scale: “inadequate,” “doubtful,” “adequate,” or “very good” [20,22]. The item with the worst rating under each measurement property would determine the overall rating for the specific measurement property [31].

On the basis of the COSMIN guidelines, it is recommended for the review team to determine before assessing the methodological quality of studies which outcome measurement instrument can be considered a reasonable gold standard [32]. The study team decided that there is currently no gold standard in the field of patient-reported outcomes that measure medication adherence [33,34]; thus, the assessment of criterion validity of the PROMs was not performed except when an abridged PROM was compared with the original long version, which will be regarded as the gold standard. This is in line with the consensus from the COSMIN panel that no gold standard exists for PROMs [35].

### Assessment of Psychometric Quality

The psychometric quality for each medication adherence instrument was assessed using the quality criteria by Terwee et al [36]. For each of the measurement properties evaluated in the included studies, a “positive (+),” “indeterminate (?),” or

“negative (–)” rating was assigned based on the psychometric results. For example, for the internal consistency measurement property, the rating will be “+” if “at least low evidence for sufficient structural validity AND Cronbach’s alpha(s)  $\geq 0.70$  for each unidimensional scale or subscale.” The rating will be “–” if there is “at least low evidence for sufficient structural validity AND Cronbach’s alpha(s)  $< 0.70$  for each unidimensional scale or subscale.” The rating will be “?” if the “criteria for at least low evidence for sufficient structural validity is not met” [36].

### Evidence Synthesis

For each PROM, an evidence synthesis across studies was conducted. First, we determined whether each measurement property for a PROM had overall “sufficient (+),” “insufficient (–),” “inconsistent ( $\pm$ ),” or “indeterminate (?)” evidence for each measurement property of the PROM. For PROMs that were assessed in more than one study, the overall rating of the level of evidence for the PROM would be sufficient (+), indeterminate (?), or insufficient (–) if the individual studies were all consistently rated as positive (+), indeterminate (?), or negative (–), respectively. If the results of individual studies were inconsistent, the overall rating of the level of evidence for the PROM would be inconsistent.

We also graded the quality of evidence for each measurement property of PROM as “high,” “moderate,” “low,” or “very low” level of evidence based on the guidelines from the modified Grading of Recommendations Assessment, Development and Evaluation approach for systematic reviews of clinical trials [22,37].

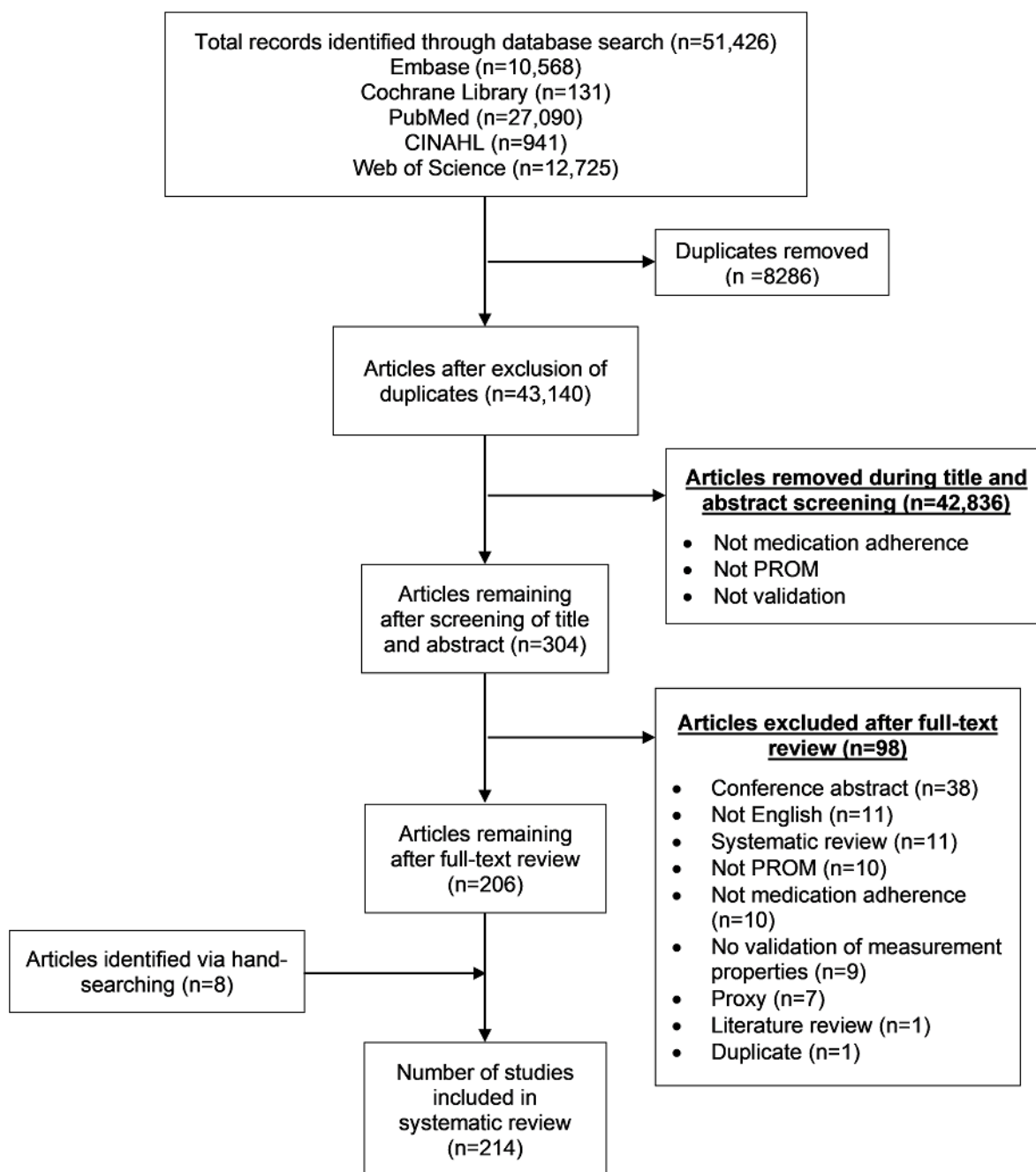
## Results

### Search Results and Study Characteristics

In total, 51,426 articles were retrieved from the 5 databases. After removing 8286 duplicates and 42,836 articles during title and abstract screening, 304 articles remained for full-text review. A total of 98 articles were further eliminated during full-text article screening. An additional 8 relevant articles were identified through hand-searching of the reference lists from the included articles to obtain a final list of 214 articles (Figure 1). A total of 240 PROMs were evaluated across 214 studies, and 121 unique medication adherence PROMs in 32 languages from 48 countries were identified (Table 1).



**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram for the systematic review. CINAHL: Cumulative Index to Nursing and Allied Health Literature; PROM: patient-reported outcome measure.



**Table 1.** Study characteristics of included articles (N=214).

Study characteristics	Values, n (%)
<b>Sample size<sup>a</sup></b>	
<50	17 (7.9)
50-99	31 (14.4)
100-199	62 (29.0)
200-299	32 (15.0)
300-399	23 (10.7)
400-499	15 (7.0)
>500	35 (16.4)
<b>Mean age (years)<sup>a,b</sup></b>	
<20	9 (4.2)
20-39	28 (13.1)
40-59	97 (45.3)
60-79	55 (25.7)
>80	2 (0.9)
<b>Proportion of males (%)<sup>a</sup></b>	
0-20	15 (7.0)
21-40	47 (22.0)
41-60	78 (36.4)
61-80	38 (17.8)
81-100	23 (10.7)
<b>Mean treatment/disease duration<sup>a</sup> (years)</b>	
<1	3 (1.4)
1-5	11 (5.1)
6-10	27 (12.6)
11-15	16 (7.5)
>15	2 (0.9)

<sup>a</sup>Some values were not reported.<sup>b</sup>Some values were reported as median, range, or not reported.

## PROM Characteristics

The characteristics of the medication adherence PROM are presented in [Multimedia Appendix 2](#) [38-51]. The most extensively studied PROMs were the 8-item MMAS (MMAS-8) and 4-item MMAS (MMAS-4), which were studied in 27 and 18 studies, respectively. Among the 121 PROMs, 15 PROMs have been administered electronically—Adult AIDS Clinical Trials Group (AACTG) adherence items [52], Fredericksen et al [38], Item Response Theory-30 [53], Least Absolute Shrinkage and Selection Operator-10 [53], Medication Adherence Reasons Scale (MAR-Scale) [54], MAR-Scale (revised) [55], MARS [56], Medication Intake Survey-Asthma [57], MMAS-4 [54], MMAS-8 [58], Multiple Sclerosis Treatment Adherence Questionnaire [59], Probabilistic Medication Adherence Scale [56], Self-Rating Scale Item [52], visual analog scale [52], and WedAd-Q Questionnaire [60].

## Assessment of Methodological and Psychometric Quality

Outcomes on the assessment of methodological quality and study quality of PROMs are summarized in [Multimedia Appendix 3](#) [13,17,38-54,56,57,59,61-248]. In terms of validity, “hypotheses testing for construct validity,” “structural validity,” and “content validity” were assessed in 181, 92, and 92 studies, respectively. In terms of reliability, “internal consistency” and “reliability” were assessed in 138 and 27 studies, respectively.

No studies assessed “measurement error” or “cross-cultural validity/measurement invariance.” Of note, 46 studies performed translation of PROMs, but none of them assessed measurement invariance or differential item functioning. Furthermore, although 75 studies assessed *PROM development*, only one of them obtained *very good* methodological quality.

## Evidence Synthesis

The results for the evidence synthesis for the PROMs are summarized in [Table 2](#). PROMs with at least a *moderate* level of evidence for  $\geq 5$  measurement properties include the Adherence Starts with Knowledge 20 (ASK-20), Compliance Questionnaire-Rheumatology (CQR), General Medication

Adherence Scale (GMAS), Hill-Bone Scale, Immunosuppressant Therapy Barrier Scale (ITBS), MAR-Scale revised, MARS-5, MARS-9, MMAS-4, MMAS-8, Self-efficacy for Appropriate Medication Adherence Scale (SEAMS), Satisfaction with Iron Chelation Therapy (SICT), Test of Adherence to Inhalers (TAI), and the questionnaire by Voils.

**Table 2.** Evidence synthesis of measurement properties for each patient-reported outcome measure.

PROM <sup>a</sup>	Number of studies	PROM development	Content validity	Structural validity	Internal consistency	Cross-cultural validity/measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
Adult Asthma Adherence Questionnaire	1	0 <sup>b</sup>	? <sup>c</sup> /B <sup>d</sup>	?/B	– <sup>e</sup> /C <sup>f</sup>	0	0	0	0	–/B	+ <sup>g</sup> /B
Adult AIDS Clinical Trials Group	8	0	?/D <sup>h</sup>	?/B	+/ <sup>i</sup> A	0	0	0	0	± <sup>j</sup> /C	+/ <sup>k</sup> C
Antidepressant Adherence Scale	1	D	?/D	0	–/D	0	0	0	0	–/C	0
Adherence Barrier Questionnaire	2	D	0	?/B	+/ <sup>l</sup> A	0	0	0	0	–/B	0
Adherence tool for chronic myelomonocytic leukemia	1	D	0	0	–/C	0	0	0	0	–/D	–/D
Adherence Evaluation of Osteoporosis Treatment Questionnaire-12	1	0	0	0	0	0	0	0	0	+/ <sup>m</sup> A	+/ <sup>n</sup> A
Antipsychotic Medication Beliefs and Attitudes Scale	1	D	+/ <sup>o</sup> C	–/B	+/ <sup>p</sup> A	0	0	0	0	+/ <sup>q</sup> B	0
Adherence to Pulmonary Rehabilitation Questionnaire	1	C	+/ <sup>r</sup> C	?/B	+/ <sup>s</sup> D	0	0	0	0	0	0
ARMS <sup>k</sup>	1	C	+/ <sup>t</sup> C	?/B	+/ <sup>u</sup> A	0	?/C	0	0	+/ <sup>v</sup> B	0
ARMS-7	1	0	?/C	–/A	+/ <sup>w</sup> A	0	+/ <sup>x</sup> B	0	0	0	0
ASK <sup>l</sup> -12	1	0	0	?/B	+/ <sup>y</sup> A	0	+/ <sup>z</sup> B	0	0	+/ <sup>aa</sup> A	0
ASK-20	3	B	+/ <sup>ab</sup> B	?/B	+/ <sup>ac</sup> D	0	+/ <sup>ad</sup> B	0	0	±/ <sup>ae</sup> A	–/B
Attitudes to mesalamine questionnaire	1	D	0	0	0	0	0	0	0	+/ <sup>af</sup> A	+/ <sup>ag</sup> A
Adherence self-report questionnaire	1	0	+/ <sup>ah</sup> C	0	0	0	–/D	0	0	0	–/B
Axelsson et al <sup>m</sup> [39]	1	D	0	?/B	+/ <sup>ai</sup> A	0	0	0	0	+/ <sup>aj</sup> C	0
Basel Assessment of Adherence to Immunosuppressive Medications Scale	1	0	+/ <sup>ak</sup> B	?/B	+/ <sup>al</sup> D	— <sup>n</sup>	+/ <sup>am</sup> A	0	0	+/ <sup>an</sup> B	0
Brief Evaluation of Medication Influences and Beliefs	1	D	?/D	?/C	–/D	0	?/D	0	0	–/D	+/ <sup>ao</sup> C
Beliefs Related to Medication Adherence	1	D	?/D	?/D	+/ <sup>ap</sup> D	0	0	0	0	+/ <sup>aq</sup> D	0
Brief Medication Adherence Scale	1	D	0	?/B	?/D	0	+/ <sup>ar</sup> B	0	0	+/ <sup>as</sup> C	0
Beliefs about Medication Compliance Scale	1	D	+/ <sup>at</sup> C	?/B	+/ <sup>au</sup> A	0	0	0	0	0	0
Brief Medication Questionnaire	4	D	?/C	0	0	0	+/ <sup>av</sup> B	0	0	+/ <sup>aw</sup> B	+/ <sup>ax</sup> C
Center for Adherence Support Evaluation Adherence Index	2	0	0	0	0	0 <sup>n</sup>	?/D	0	0	+/ <sup>ay</sup> B	–/B

PROM <sup>a</sup>	Number of studies	PROM development	Content validity	Structural validity	Internal consistency	Cross-cultural validity/measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
Chronic Disease Compliance Instrument	3	A	+A	?/B	+/B	— <sup>n</sup>	0	0	0	—/B	0
CEAT-VIH	1	0	+/B	0	+/D	0	0	0	0	+/C	0
Chaiyachattiet al <sup>m</sup> [40]	1	0	0	0	0	0	0	0	0	0	—/C
Compliance assessment	1	0	0	0	?/D	0	0	0	0	+/B	0
Cohort Study of Medication Adherence Among Older Adults self-report tool	1	D	0	?/D	0	0	0	0	0	+/B	+/B
CQR <sup>o</sup>	4	B	+/B	?/B	+A	0 <sup>n</sup>	±/B	0	0	±/B	+A
CQR-5	1	0	+A	+D	+A	0	0	0	0	+B	—/B
Da et al <sup>m</sup> [41]	1	0	0	0	0	0	0	0	0	—/C	—/B
DAI <sup>p</sup>	1	D	0	0	+D	0	?/D	0	0	+B	0
DAI-10	2	0	0	0	+C	0	0	0	0	±/A	—/B
DAI-9	1	0	0	?/C	0	0	—/D	0	0	—/D	0
Diagnostic Adherence to Medication Scale	1	B	+B	0	0	0	0	0	0	+A	0
Demirtas et al <sup>m</sup> [42]	1	B	+B	?/A	+A	0	?/C	0	0	+C	0
Danish version of Medication Adherence Report Scale-4	1	0	?/D	0	+C	0	0	0	0	—/D	0
Diabetes Management Questionnaire	1	D	?/C	0	+A	0	—/C	0	0	+B	0
Diabetes Medication Self-efficacy Scale	2	D	+B	?/B	+D	0 <sup>n</sup>	+B	0	0	+C	—/B
Environmental Barriers to Adherence Scale	1	0	0	0	+A	0	?/D	0	0	+D	0
Eye-Drop Satisfaction Questionnaire	2	D	+C	?/B	+A	0	0	0	0	?/C	0
End-Stage Renal Disease Adherence Questionnaire	2	D	+B	0	0	0 <sup>n</sup>	+C	0	0	—/B	0
Every Visit Adherence Questionnaire	1	0	0	0	0	0	0	0	0	+D	0
Five-dimension adherence model	1	D	0	0	0	0	0	0	0	0	+A
Fredericksen et al <sup>m</sup> [249]	1	C	+A	0	0	0	+A	0	0	0	0
General adherence tendency measure	1	D	0	0	0	0	0	0	0	—/B	—/B
General Medicine Adherence Scale	3	C	+B	+A	+A	0 <sup>n</sup>	±/B	0	0	+B	+C
Godin et al <sup>m</sup> [43]	2	C	?/C	0	0	0	0	0	0	—/C	+C
GTCAT <sup>d</sup>	1	D	?/B	?/D	—/B	0	?/D	0	0	+B	0
GTCAT (reduced)	2	0	?/B	—/C	—/A	0	+B	0	0	—/B	0



PROM <sup>a</sup>	Number of studies	PROM development	Content validity	Structural validity	Internal consistency	Cross-cultural validity/measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
Hill-Bone Scale	5	B	+A	?/B	+B	0 <sup>n</sup>	0	0	0	+A	0
Hill-Bone Scale (modified)	1	0	?/C	0	−/D	0 <sup>n</sup>	0	0	0	0	0
HIV Intention Measure	1	C	+B	?/B	+A	0	0	0	0	+C	0
HIV Symptom Quality of Life Adherence Questionnaire	1	D	0	?/B	+C	0	0	0	0	+B	0
Iraqi Anti-Diabetic Medication Adherence Scale	1	C	?/C	0	+B	0 <sup>n</sup>	?/C	0	0	+B	+C
Item Response Theory-30	1	B	+B	?/B	0	0	0	0	0	?/C	+B
Immunosuppressant Therapy Adherence Scale	2	0	0	?/B	+D	0 <sup>n</sup>	0	0	0	±B	0
Immunosuppressant Therapy Barrier Scale	1	C	+B	?/B	+A	0	0	0	0	+A	+A
Kennedy et al <sup>m</sup> [44]	1	D	+B	0	0	0	0	0	0	0	0
Kerr et al <sup>m</sup> [45]	1	D	0	0	0	0	0	0	0	−/D	−/D
Least absolute shrinkage and selection operator-10	1	B	+B	?/B	0	0	0	0	0	?/C	+B
Long-Term Medication Behaviour Self-Efficacy Scale	1	D	?/D	0	0	0	0	0	0	0	0
Modified Drug Adherence Work-up Tool	2	0	?/D	?/D	+D	0	0	0	0	−/B	+B
Medication Adherence Questionnaire	1	D	?/D	0	+D	0 <sup>n</sup>	+D	0	0	0	0
MAR-Scale <sup>r</sup>	1	C	?/D	?/B	−/A	0	−/D	0	0	+D	0
MAR-Scale (revised)	2	B	+B	?/B	+A	0	0	0	0	+B	0
MARS <sup>s</sup>	8	D	+D	?/B	±A	0 <sup>n</sup>	?/D	0	0	±B	±B
MARS-10	1	0	?/D	?/B	+A	0 <sup>n</sup>	0	0	0	+B	+B
MARS-5	8	0	+B	?/B	+B	0 <sup>n</sup>	±B	0	0	±A	±B
MARS-9	2	0	+B	?/B	+A	0	−/B	0	0	±B	0
MASES <sup>t</sup>	2	D	+A	?/B	+A	0	−/D	0	0	±A	0
MASES-R	1	0	0	?/B	+D	0	−/C	0	0	−/B	0
Medication Adherence Self-Report Inventory	6	D	0	0	+B	0	+D	0	0	+A	+A
Medication adherence scale	1	D	+C	−/A	+A	0	0	0	0	−/B	0
Medication adherence survey	1	B	+B	0	0	0	0	0	0	+D	0
Medication Adherence Estimation and Differentiation Scale	1	C	?/B	+A	+A	0	0	0	0	+D	−/A

PROM <sup>a</sup>	Number of studies	PROM development	Content validity	Structural validity	Internal consistency	Cross-cultural validity/measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
Medication Intake Survey-Asthma	1	B	+B	0	0	0	−/B	0	0	+B	0
MMAS <sup>u</sup> -4	18	0	+C	?/B	−/B	0	±/B	0	0	±/A	−/B
MMAS-7	1	0	0	0	0	0 <sup>n</sup>	0	0	0	+B	0
MMAS-8	27	0	+B	±/A	±/A	0 <sup>n</sup>	+B	0	0	±/A	−/B
MMAS-9	1	0	0	+A	−/A	0	?/C	0	0	−/B	0
Medication Nonpersistence Scale	1	C	?/C	+A	+D	0	0	0	0	+C	−/D
Medical Outcomes Study General Adherence Scale	1	0	0	0	+D	0	0	0	0	−/C	0
Multiple Sclerosis Treatment Adherence Questionnaire	1	C	?/B	0	−/A	0	0	0	0	+A	0
Outcome Expectations for Osteoporosis Medication Adherence Scale	2	D	+B	±/A	+A	0	0	0	0	−/C	0
Perceived Barriers to Antiretroviral Therapy Adherence Scale	1	B	+C	?/B	?/C	0 <sup>n</sup>	−/C	0	0	+C	0
Pictographic Self-Efficacy Scale	1	D	+B	0	−/A	0	−/D	0	0	−/B	0
Patient Rating of Compliance Scale	1	0	0	0	0	0	0	0	0	+B	0
Patient Preference Questionnaire	1	D	+D	0	+D	0	0	0	0	0	0
Probabilistic Medication Adherence Scale	1	C	±/C	−/A	+C	0	0	0	0	0	0
Number of pills taken or prescribed	2	0	0	0	?/D	0	−/D	0	0	±/C	0
Questionnaire for Adherence with Topical Treatments in Psoriasis	1	D	+B	0	0	0	0	0	0	±/D	0
Question of Interest	1	0	?/D	0	0	0	0	0	0	−/D	−/D
SCI <sup>v</sup>	2	0	?/C	+A	0	0	−/B	0	0	±/A	−/B
SCI-R	1	0	+B	?/B	+A	0 <sup>n</sup>	?/D	0	0	+D	+D
Strathclyde Compliance Risk Assessment Tool	2	0	0	0	+B	0	0	0	0	+B	+B
Summary of Diabetes Self-care Activities	2	0	0	0	0	0	0	0	0	±/C	0
Self-Efficacy for Appropriate Medication Adherence Scale	3	B	+B	?/B	+A	0	+A	0	0	±/A	0
Self-efficacy scale	1	D	?/B	?/B	+A	0	0	0	0	?/D	0
Self-report measures of adherence	1	0	0	0	0	0	?/B	0	0	+B	0
Self-report on adherence	1	0	0	0	0	0	0	0	0	−/C	+C

PROM <sup>a</sup>	Number of studies	PROM development	Content validity	Structural validity	Internal consistency	Cross-cultural validity/measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
Self-Efficacy for Osteoporosis Medication Adherence Scale	2	D	+B	−/A	+A	0	0	0	0	+C	0
Self-Reported Adherence Questionnaire	1	0	D	0	0	0	0	0	0	?/D	0
Satisfaction with Iron Chelation Therapy	1	B	+B	?/B	+A	0	0	0	0	+A	0
Sidorkiewicz et al <sup>m</sup> [46]	1	B	+B	0	0	0	0	0	0	+C	0
Simplified Medication Adherence Questionnaire	2	D	?/C	0	+C	0	+A	0	0	+B	−/A
Stages of Change Model Questionnaire	2	D	0	0	0	0	0	0	0	+C	0
Special Projects of National Significance Adherence Survey	1	0	0	0	+D	0	0	0	0	+C	0
Self-Rating Scale Item	4	0	0	0	0	0	0	0	0	+C	+C
Test of Adherence to Inhalers	1	B	?/B	?/B	+D	0	+A	0	0	−/C	−/A
Tan et al [47]	1	C	?/C	?/B	+A	0	+A	0	0	+B	0
Treatment Adherence Survey-Patient Version	1	D	0	0	0	0	+B	0	0	−/C	0
Therapeutic Adherence Scale for Hypertensive Patients	1	0	0	+A	+A	0	0	0	0	+B	−/B
Topical Therapy Adherence Questionnaire	1	D	+D	0	+D	0	0	0	0	0	0
Turcu-tioică et al <sup>m</sup> [48]	1	D	+D	0	−/D	0	0	0	0	−/D	0
Visual analog scale	4	0	0	0	0	0	−/D	0	0	+B	0
Validated Hemophilia Regimen Treatment Adherence Scale—On-Demand	1	D	+D	0	+B	0	+B	0	0	±/B	0
Validated Hemophilia Regimen Treatment Regimen Treatment Adherence Scale-Prophylaxis	1	D	+D	0	+C	0	+C	0	0	−/C	0
Voils et al <sup>m</sup> [49]	4	C	+B	+A	+A	0 <sup>n</sup>	−/A	0	0	±/A	−/B
Vreeman et al <sup>m</sup> [50]	2	C	?/C	0	0	0 <sup>n</sup>	0	0	0	−/C	−/C
Web-Ad-Q Questionnaire	1	C	+C	0	0	0	+D	0	0	+C	0

PROM <sup>a</sup>	Number of studies	PROM development	Content validity	Structural validity	Internal consistency	Cross-cultural validity/measurement invariance	Reliability	Measurement error	Criterion validity	Hypotheses testing for construct validity	Responsiveness
Wilson et al <sup>m</sup> [51]	3	B	+B	0	+A	0	0	0	0	+D	–/A

<sup>a</sup>PROM: patient-reported outcome measurement.

<sup>b</sup>0: Measurement properties were not assessed by the study.

<sup>c</sup>?: intermediate.

<sup>d</sup>B: moderate.

<sup>e</sup>–: insufficient.

<sup>f</sup>C: low.

<sup>g</sup>+: sufficient.

<sup>h</sup>D: very low.

<sup>i</sup>A: high.

<sup>j</sup>±: inconsistent.

<sup>k</sup>ARMS: Adherence to Refills and Medications Scale.

<sup>l</sup>ASK: Adherence Starts with Knowledge questionnaire.

<sup>m</sup>PROMs without proper names are labeled based on the last name of the first author who developed the instrument.

<sup>n</sup>Only translation was done. Cross-cultural validation was not the aim of the study.

<sup>o</sup>CQR: Compliance Questionnaire on Rheumatology.

<sup>p</sup>DAI: Drug Attitude Inventory.

<sup>q</sup>GTCAT: Glaucoma Treatment Compliance Assessment Tool.

<sup>r</sup>MAR-Scale: Medication Adherence Reasons Scale.

<sup>s</sup>MARS: Medication Adherence Rating Scale.

<sup>t</sup>MASES: Medication Adherence Self-efficacy Scale.

<sup>u</sup>MMAS: Morisky Medication Adherence Scale.

<sup>v</sup>SCI: Self-Care Inventory.

## Discussion

To the best of our knowledge, this is the first systematic review that comprehensively summarized PROMs for medication adherence based on the COSMIN guidelines [21,22]. Among the 214 included articles, we identified 121 unique PROMs for medication adherence. Our study revealed the most commonly evaluated medication adherence PROMs to be the MMAS-8, MMAS-4, AACTG, MARS, and MARS-5. However, being more commonly evaluated does not mean that these PROMs have the best psychometric properties. Moreover, based on the number of studies for each PROM in Table 2, most of the PROMs have too few studies to provide a strong evidence base for their use.

Among the 15 PROMs that have been administered electronically, 3 PROMs (MAR-Scale revised, MMAS-4, and MMAS-8) have at least a *moderate* level of evidence for  $\geq 5$  measurement properties. Electronic administration of PROMs to measure medication adherence may be appealing in health care settings, as it may reduce the administrative burden for data collection and data entry. In addition, as web-based interventions to improve medication adherence become increasingly commonplace [250], electronic PROMs may be incorporated into web-based platforms to assess the effectiveness of these web-based interventions.

Despite a few studies claiming the use of certain PROMs and objective measures as the *gold standard* for measuring medication adherence, we deliberately omitted evaluating *criterion validity* for these studies. As mentioned in the introduction, although objective measures such as pill count, electronic monitoring devices, and big data may measure adherence indirectly, these measures are laborious, costly, and sometimes invasive, making them unsuitable for routine clinical use. Furthermore, all these surrogate measures, including PROMs, do not predict any real biological outcomes such as a reduction in viral load, blood pressure, or glucose concentration in determining medication adherence in patients. Hence, none of these measures can be deemed as a *gold standard* [10].

Measurement error was not evaluated because none of the studies reported the standard error of measurement, smallest detectable change, or limits of agreements required by the COSMIN. In addition, although translations of PROMs were performed in 46 studies, none of these studies assessed measurement invariance or differential item functioning; therefore, cross-cultural validity was not evaluated for any of the PROMs in this study. Moreover, only one study examined the interpretability of PROMs in the form of minimal detectable change for the MMAS-8 [251]. Further study on measurement error, cross-cultural validity, and interpretability of medication adherence PROMs is warranted.

The strengths of this study include using COSMIN guidelines, which are well regarded as a consensus-based standard for

evaluating the measurement properties of PROMs [23]. The COSMIN Risk of Bias checklist employed in this study is an improvement from the original COSMIN checklist with several improvements in the standards for evaluation [21,22]. We also used sensitive search filters to retrieve and include as many potentially relevant articles as possible.

One limitation related to this study was that the selection of articles and evaluation of psychometric properties were subjective in nature and may have been prone to judgment bias. Nevertheless, the requirement by COSMIN to have 2 independent reviewers and the need for a third reviewer to reach a consensus in the case of any discrepancy occurring has helped reduce the risk of judgment bias [22,252].

## Conclusions

In summary, 121 unique medication adherence PROMs were identified in 214 studies. On the basis of the COSMIN guidelines, PROMs with at least a *moderate* level of evidence for  $\geq 5$  measurement properties include the ASK-20, CQR, GMAS, Hill-Bone Scale, ITBS, MAR-Scale revised, MARS-5, MARS-9, MMAS-4, MMAS-8, SEAMS, SICT, TAI, and questionnaire by Voils. Of these, only the GMAS has *sufficient* (+) ratings for at least four measurement properties. We believe this study would assist clinicians and researchers in selecting suitable PROMs to measure medication adherence among patients. Future research may consider validating measurement errors and cross-cultural validity to further improve the insights on the measurement properties of these PROMs.

## Acknowledgments

The authors would like to thank Librarian Ms Wong Suei Nee at the National University of Singapore Medical Library for her advice on search strategies. This research was supported by the SingHealth Regional Health System Population-based, Unified, Learning System for Enhanced and Sustainable Health (PULSES) Centre Grant—*Measuring Medication Adherence in Singapore* (CGDec19S03), and PULSES Fellowship (CGDec19F02).

Kwan Yu Heng, Weng Si Dun, and Dionne Hui Fang Loh contributed equally as co-first authors. Truls Østbye, Low Lian Leng, Hayden Barry Bosworth, and Julian Thumboo contributed equally as senior coauthors.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Search strategy.

[DOCX File, 33 KB - [jmir\\_v22i10e19179\\_app1.docx](#)]

### Multimedia Appendix 2

Characteristics of patient reported outcome measures.

[DOCX File, 45 KB - [jmir\\_v22i10e19179\\_app2.docx](#)]

### Multimedia Appendix 3

Assessment of psychometric properties of patient-reported outcome measures.

[DOCX File, 64 KB - [jmir\\_v22i10e19179\\_app3.docx](#)]

## References

1. Dal-Fabbro AL. Adherence to long term therapies: evidence for action. *Cad. Saúde Pública* 2005 Aug;21(4):1297-1298. [doi: [10.1590/s0102-311x2005000400037](#)]
2. Asche C, LaFleur J, Conner C. A review of diabetes treatment adherence and the association with clinical and economic outcomes. *Clin Ther* 2011 Jan;33(1):74-109. [doi: [10.1016/j.clinthera.2011.01.019](#)] [Medline: [21397776](#)]
3. Han E, Suh D, Lee S, Jang S. The impact of medication adherence on health outcomes for chronic metabolic diseases: a retrospective cohort study. *Res Social Adm Pharm* 2014;10(6):e87-e98. [doi: [10.1016/j.sapharm.2014.02.001](#)] [Medline: [25088545](#)]
4. Sabaté E. editor. , ed. Adherence to Long-Term Therapies evidence for Action. Geneva, Switzerland: World Health Organization; 2003:E.
5. Lee S, Raamkumar AS, Li J, Cao Y, Witedwittayanusat K, Chen L, et al. Reasons for primary medication nonadherence: a systematic review and metric analysis. *J Manag Care Spec Pharm* 2018 Aug;24(8):778-794 [FREE Full text] [doi: [10.18553/jmcp.2018.24.8.778](#)] [Medline: [30058985](#)]
6. Ho PM, Rumsfeld JS, Masoudi FA, McClure DL, Plomondon ME, Steiner JF, et al. Effect of medication nonadherence on hospitalization and mortality among patients with diabetes mellitus. *Arch Intern Med* 2006 Sep 25;166(17):1836-1841. [doi: [10.1001/archinte.166.17.1836](#)] [Medline: [17000939](#)]



7. Haynes RB, Ackloo E, Sahota N, McDonald HP, Yao X. Interventions for enhancing medication adherence. *Cochrane Database Syst Rev* 2008 Apr 16(2):CD000011. [doi: [10.1002/14651858.CD000011.pub3](https://doi.org/10.1002/14651858.CD000011.pub3)] [Medline: [18425859](#)]
8. Brown M, Bussell J. Medication adherence: WHO cares? *Mayo Clin Proc* 2011 Apr;86(4):304-314 [FREE Full text] [doi: [10.4065/mcp.2010.0575](https://doi.org/10.4065/mcp.2010.0575)] [Medline: [21389250](#)]
9. Zhang JX, Meltzer DO. Identifying patients with cost-related medication non-adherence: a big-data approach. *J Med Econ* 2016 Aug;19(8):806-811 [FREE Full text] [doi: [10.1080/13696998.2016.1176031](https://doi.org/10.1080/13696998.2016.1176031)] [Medline: [27052465](#)]
10. Lam WY, Fresco P. Medication adherence measures: an overview. *Biomed Res Int* 2015;2015:217047 [FREE Full text] [doi: [10.1155/2015/217047](https://doi.org/10.1155/2015/217047)] [Medline: [26539470](#)]
11. Moon SJ, Lee W, Hwang JS, Hong YP, Morisky DE. Accuracy of a screening tool for medication adherence: a systematic review and meta-analysis of the Morisky medication adherence scale-8. *PLoS One* 2017;12(11):e0187139 [FREE Full text] [doi: [10.1371/journal.pone.0187139](https://doi.org/10.1371/journal.pone.0187139)] [Medline: [29095870](#)]
12. Kim M, Hill M, Bone L, Levine D. Development and testing of the hill-bone compliance to high blood pressure therapy scale. *Prog Cardiovasc Nurs* 2000;15(3):90-96. [doi: [10.1111/j.1751-7117.2000.tb00211.x](https://doi.org/10.1111/j.1751-7117.2000.tb00211.x)] [Medline: [10951950](#)]
13. Thompson K, Kulkarni J, Sergejew A. Reliability and validity of a new medication adherence rating scale (MARS) for the psychoses. *Schizophr Res* 2000 May 5;42(3):241-247. [doi: [10.1016/s0920-9964\(99\)00130-9](https://doi.org/10.1016/s0920-9964(99)00130-9)] [Medline: [10785582](#)]
14. Voils CI, Hoyle RH, Thorpe CT, Maciejewski ML, Yancy WS. Improving the measurement of self-reported medication nonadherence. *J Clin Epidemiol* 2011 Mar;64(3):250-254. [doi: [10.1016/j.jclinepi.2010.07.014](https://doi.org/10.1016/j.jclinepi.2010.07.014)] [Medline: [21194887](#)]
15. Voils CI, Maciejewski ML, Hoyle RH, Reeve BB, Gallagher P, Bryson CL, et al. Initial validation of a self-report measure of the extent of and reasons for medication nonadherence. *Med Care* 2012 Dec;50(12):1013-1019 [FREE Full text] [doi: [10.1097/MLR.0b013e318269e121](https://doi.org/10.1097/MLR.0b013e318269e121)] [Medline: [22922431](#)]
16. Korb-Savoldelli V, Gillaizeau F, Pouchot J, Lenain E, Postel-Vinay N, Plouin PF, et al. Validation of a French version of the 8-item Morisky medication adherence scale in hypertensive adults. *J Clin Hypertens (Greenwich)* 2012 Jul;14(7):429-434 [FREE Full text] [doi: [10.1111/j.1751-7176.2012.00634.x](https://doi.org/10.1111/j.1751-7176.2012.00634.x)] [Medline: [22747615](#)]
17. Fialko L, Garety PA, Kuipers E, Dunn G, Bebbington PE, Fowler D, et al. A large-scale validation study of the medication adherence rating scale (MARS). *Schizophr Res* 2008 Mar;100(1-3):53-59. [doi: [10.1016/j.schres.2007.10.029](https://doi.org/10.1016/j.schres.2007.10.029)] [Medline: [18083007](#)]
18. Gagné M, Boulet L, Pérez N, Moisan J. Patient-reported outcome instruments that evaluate adherence behaviours in adults with asthma: a systematic review of measurement properties. *Br J Clin Pharmacol* 2018 Sep;84(9):1928-1940 [FREE Full text] [doi: [10.1111/bcp.13623](https://doi.org/10.1111/bcp.13623)] [Medline: [29710423](#)]
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](#)]
20. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010 May;19(4):539-549 [FREE Full text] [doi: [10.1007/s11136-010-9606-8](https://doi.org/10.1007/s11136-010-9606-8)] [Medline: [20169472](#)]
21. Mokkink LB, de Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018 May;27(5):1171-1179 [FREE Full text] [doi: [10.1007/s11136-017-1765-4](https://doi.org/10.1007/s11136-017-1765-4)] [Medline: [29260445](#)]
22. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018 May;27(5):1147-1157 [FREE Full text] [doi: [10.1007/s11136-018-1798-3](https://doi.org/10.1007/s11136-018-1798-3)] [Medline: [29435801](#)]
23. Mokkink LB, Prinsen CA, Bouter LM, Vet HC, Terwee CB. The consensus-based standards for the selection of health measurement instruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther* 2016 Jan 19;20(2):105-113 [FREE Full text] [doi: [10.1590/bjpt-rbf.2014.0143](https://doi.org/10.1590/bjpt-rbf.2014.0143)] [Medline: [26786084](#)]
24. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010 Jul;63(7):737-745. [doi: [10.1016/j.jclinepi.2010.02.006](https://doi.org/10.1016/j.jclinepi.2010.02.006)] [Medline: [20494804](#)]
25. Davidson M, Keating J. Patient-reported outcome measures (PROMs): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. *Br J Sports Med* 2014 May;48(9):792-796. [doi: [10.1136/bjsports-2012-091704](https://doi.org/10.1136/bjsports-2012-091704)] [Medline: [23258849](#)]
26. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000 May;53(5):459-468. [doi: [10.1016/s0895-4356\(99\)00206-1](https://doi.org/10.1016/s0895-4356(99)00206-1)] [Medline: [10812317](#)]
27. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009 Oct;18(8):1115-1123 [FREE Full text] [doi: [10.1007/s11136-009-9528-5](https://doi.org/10.1007/s11136-009-9528-5)] [Medline: [19711195](#)]
28. Terwee CB, Prinsen CA, Ricci Garotti MG, Suman A, de Vet HC, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res* 2016 Apr;25(4):767-779 [FREE Full text] [doi: [10.1007/s11136-015-1122-4](https://doi.org/10.1007/s11136-015-1122-4)] [Medline: [26346986](#)]

29. Scherer RW, Saldanha IJ. How should systematic reviewers handle conference abstracts? A view from the trenches. *Syst Rev* 2019 Nov 7;8(1):264 [FREE Full text] [doi: [10.1186/s13643-019-1188-0](https://doi.org/10.1186/s13643-019-1188-0)] [Medline: [31699124](#)]
30. Richards D. Handsearching still a valuable element of the systematic review. *Evid Based Dent* 2008;9(3):85. [doi: [10.1038/sj.ebd.6400602](https://doi.org/10.1038/sj.ebd.6400602)] [Medline: [18927572](#)]
31. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012 May;21(4):651-657 [FREE Full text] [doi: [10.1007/s11136-011-9960-1](https://doi.org/10.1007/s11136-011-9960-1)] [Medline: [21732199](#)]
32. Mokkink L, Prinsen C, Patrick D. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User manual 2018;78:1.
33. Vitolins MZ, Rand CS, Rapp SR, Ribisl PM, Sevic MA. Measuring adherence to behavioral and medical interventions. *Control Clin Trials* 2000 Oct;21(5 Suppl):188S-194S. [doi: [10.1016/s0197-2456\(00\)00077-5](https://doi.org/10.1016/s0197-2456(00)00077-5)] [Medline: [11018574](#)]
34. Smith H, Hankins M, Hodson A, George C. Measuring the adherence to medication of elderly patients with heart failure: is there a gold standard? *Int J Cardiol* 2010 Nov 5;145(1):122-123. [doi: [10.1016/j.ijcard.2009.06.031](https://doi.org/10.1016/j.ijcard.2009.06.031)] [Medline: [19615767](#)]
35. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010 Mar 18;10:22 [FREE Full text] [doi: [10.1186/1471-2288-10-22](https://doi.org/10.1186/1471-2288-10-22)] [Medline: [20298572](#)]
36. Terwee CB, Bot SD, de BM, van DW, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007 Jan;60(1):34-42. [doi: [10.1016/j.jclinepi.2006.03.012](https://doi.org/10.1016/j.jclinepi.2006.03.012)] [Medline: [17161752](#)]
37. Dijkers M. -. Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development 2020:a.
38. Fredericksen R, Yang F, Gibbons L, Edwards T, Brown S, Fitzsimmons E, et al. Development and content validation of measures assessing adherence barriers and behaviors for use in clinical care. *Res Social Adm Pharm* 2019 Sep;15(9):1168-1176. [doi: [10.1016/j.sapharm.2018.10.001](https://doi.org/10.1016/j.sapharm.2018.10.001)] [Medline: [30327183](#)]
39. Axelsson M, Ekerljung L, Lundbäck B, Lötvall J. Personality and unachieved treatment goals related to poor adherence to asthma medication in a newly developed adherence questionnaire - a population-based study. *Multidiscip Respir Med* 2016;11:42 [FREE Full text] [doi: [10.1186/s40248-016-0078-8](https://doi.org/10.1186/s40248-016-0078-8)] [Medline: [27980735](#)]
40. Chaiyachati K, Hirschhorn LR, Tanser F, Newell M, Barnighausen T. Validating five questions of antiretroviral nonadherence in a public-sector treatment program in rural South Africa. *AIDS Patient Care STDS* 2011 Mar;25(3):163-170 [FREE Full text] [doi: [10.1089/apc.2010.0257](https://doi.org/10.1089/apc.2010.0257)] [Medline: [21269131](#)]
41. Da W, Li X, Qiao S, Zhou Y, Shen Z. Evaluation of self-report adherence measures and their associations with detectable viral load among people living with HIV (PLHIV) in China. *PLoS One* 2018;13(8):e0203032 [FREE Full text] [doi: [10.1371/journal.pone.0203032](https://doi.org/10.1371/journal.pone.0203032)] [Medline: [30161177](#)]
42. Demirtaş A, Akbayrak N. Development of an assessment scale for treatment compliance in type 2 Diabetes Mellitus in Turkish population: Psychometric evaluation. *International Journal of Nursing Sciences* 2017 Jul;4(3):244-251. [doi: [10.1016/j.ijnss.2017.06.002](https://doi.org/10.1016/j.ijnss.2017.06.002)]
43. Godin G, Gagné C, Naccache H. Validation of a self-reported questionnaire assessing adherence to antiretroviral medication. *AIDS Patient Care STDS* 2003 Jul;17(7):325-332. [doi: [10.1089/108729103322231268](https://doi.org/10.1089/108729103322231268)] [Medline: [12952734](#)]
44. Kennedy SB. Developing a self-administered tool to predict adherence to antiretroviral therapy: design, method, and objectives. *AIDS Patient Care STDS* 2000 Jun;14(6):309-316. [doi: [10.1089/10872910050046331](https://doi.org/10.1089/10872910050046331)] [Medline: [10897503](#)]
45. Kerr T, Hogg RS, Yip B, Tyndall MW, Montaner J, Wood E. Validity of self-reported adherence among injection drug users. *J Int Assoc Physicians AIDS Care (Chic)* 2008;7(4):157-159. [doi: [10.1177/1545109708320686](https://doi.org/10.1177/1545109708320686)] [Medline: [18626123](#)]
46. Sidorkiewicz S, Tran V, Cousyn C, Perrodeau E, Ravaut P. Development and validation of an instrument to assess treatment adherence for each individual drug taken by a patient. *BMJ Open* 2016 May 10;6(5):e010510 [FREE Full text] [doi: [10.1136/bmjopen-2015-010510](https://doi.org/10.1136/bmjopen-2015-010510)] [Medline: [27165645](#)]
47. Tan J, Luo L, Zhang M, Chen H, Zhang D, Dong C, et al. A Chinese and Western medication adherence scale in patients with chronic kidney disease. *Patient Prefer Adherence* 2019;13:1487-1495 [FREE Full text] [doi: [10.2147/PPA.S207693](https://doi.org/10.2147/PPA.S207693)] [Medline: [31507316](#)]
48. Turcu-Stolică A. Developing of a new tool for evaluation of therapeutic adherence in COPD. *Farmacia* 2018 Oct 13;66(5):920-924. [doi: [10.31925/farmacia.2018.5.25](https://doi.org/10.31925/farmacia.2018.5.25)]
49. Cornelius T, Voils CI, Umland RC, Kronish IM. Validity of the self-reported domains of subjective extent of nonadherence (DOSE-nonadherence) scale in comparison with electronically monitored adherence to cardiovascular medications. *Patient Prefer Adherence* 2019;13:1677-1684 [FREE Full text] [doi: [10.2147/PPA.S225460](https://doi.org/10.2147/PPA.S225460)] [Medline: [31631982](#)]
50. Vreeman RC, Nyandiko WM, Ayaya SO, Walumbe EG, Inui TS. Cognitive interviewing for cross-cultural adaptation of pediatric antiretroviral therapy adherence measurement items. *Int J Behav Med* 2014 Feb;21(1):186-196. [doi: [10.1007/s12529-012-9283-9](https://doi.org/10.1007/s12529-012-9283-9)] [Medline: [23188670](#)]
51. Wilson IB, Fowler FJ, Cosenza CA, Michaud J, Bentkover J, Rana A, et al. Cognitive and field testing of a new set of medication adherence self-report items for HIV care. *AIDS Behav* 2014 Dec;18(12):2349-2358 [FREE Full text] [doi: [10.1007/s10461-013-0610-1](https://doi.org/10.1007/s10461-013-0610-1)] [Medline: [24077970](#)]

52. Feldman BJ, Fredericksen RJ, Crane PK, Safren SA, Mugavero MJ, Willig JH, et al. Evaluation of the single-item self-rating adherence scale for use in routine clinical care of people living with HIV. *AIDS Behav* 2013 Jan;17(1):307-318 [FREE Full text] [doi: [10.1007/s10461-012-0326-7](https://doi.org/10.1007/s10461-012-0326-7)] [Medline: [23108721](#)]
53. Wohl DA, Panter AT, Kirby C, Magnus BE, Hudgens MG, Allmon AG, et al. Estimating HIV medication adherence and persistence: two instruments for clinical and research use. *AIDS Behav* 2018 Mar;22(3):948-960. [doi: [10.1007/s10461-017-1772-z](https://doi.org/10.1007/s10461-017-1772-z)] [Medline: [28447269](#)]
54. Unni EJ, Farris KB. Development of a new scale to measure self-reported medication nonadherence. *Res Social Adm Pharm* 2015;11(3):e133-e143. [doi: [10.1016/j.sapharm.2009.06.005](https://doi.org/10.1016/j.sapharm.2009.06.005)] [Medline: [21272524](#)]
55. Unni E, Sternbach N, Goren A. Using the medication adherence reasons scale (MAR-Scale) to identify the reasons for non-adherence across multiple disease conditions. *Patient Prefer Adherence* 2019;13:993-1004 [FREE Full text] [doi: [10.2147/PPA.S205359](https://doi.org/10.2147/PPA.S205359)] [Medline: [31308635](#)]
56. Kleppe M, Lacroix J, Ham J, Midden C. The development of the ProMAS: a probabilistic medication adherence scale. *Patient Prefer Adherence* 2015;9:355-367 [FREE Full text] [doi: [10.2147/PPA.S76749](https://doi.org/10.2147/PPA.S76749)] [Medline: [25784791](#)]
57. Dima AL, van Ganse E, Laforest L, Texier N, de Bruin M, The Astro-Lab Group. Measuring medication adherence in asthma: development of a novel self-report tool. *Psychol Health* 2017 Oct;32(10):1288-1307. [doi: [10.1080/08870446.2017.1290248](https://doi.org/10.1080/08870446.2017.1290248)] [Medline: [28276742](#)]
58. Tran V, Harrington M, Montori VM, Barnes C, Wicks P, Ravaud P. Adaptation and validation of the treatment burden questionnaire (TBQ) in English using an internet platform. *BMC Med* 2014 Jul 2;12:109 [FREE Full text] [doi: [10.1186/1741-7015-12-109](https://doi.org/10.1186/1741-7015-12-109)] [Medline: [24989988](#)]
59. Wicks P, Massagli M, Kulkarni A, Dastani H. Use of an online community to develop patient-reported outcome instruments: the multiple sclerosis treatment adherence questionnaire (MS-TAQ). *J Med Internet Res* 2011 Jan 24;13(1):e12 [FREE Full text] [doi: [10.2196/jmir.1687](https://doi.org/10.2196/jmir.1687)] [Medline: [21266318](#)]
60. Vale FC, Santa-Helena ET, Santos MA, Carvalho WM, Menezes PR, Basso CR, et al. Development and validation of the WebAd-questionnaire to monitor adherence to HIV therapy. *Rev Saude Publica* 2018;52:62 [FREE Full text] [doi: [10.11606/s1518-8787.2018052000337](https://doi.org/10.11606/s1518-8787.2018052000337)] [Medline: [29846437](#)]
61. Schatz M, Zeiger RS, Yang S, Weinstein AG, Chen W, Saris-Baglama RN, et al. Development and preliminary validation of the adult asthma adherence questionnaire. *J Allergy Clin Immunol Pract* 2013;1(3):280-288. [doi: [10.1016/j.jaip.2013.03.001](https://doi.org/10.1016/j.jaip.2013.03.001)] [Medline: [24565486](#)]
62. Buscher A, Hartman C, Kallen MA, Giordano TP. Validity of self-report measures in assessing antiretroviral adherence of newly diagnosed, HAART-naïve, HIV patients. *HIV Clin Trials* 2011;12(5):244-254 [FREE Full text] [doi: [10.1310/hct1205-244](https://doi.org/10.1310/hct1205-244)] [Medline: [22180522](#)]
63. Gonzalez JS, Schneider HE, Wexler DJ, Psaros C, Delahanty LM, Cagliero E, et al. Validity of medication adherence self-reports in adults with type 2 diabetes. *Diabetes Care* 2013 Apr;36(4):831-837 [FREE Full text] [doi: [10.2337/dc12-0410](https://doi.org/10.2337/dc12-0410)] [Medline: [23204245](#)]
64. Kalichman SC, Amaral CM, Swetzes C, Jones M, Macy R, Kalichman MO, et al. A simple single-item rating scale to measure medication adherence: further evidence for convergent validity. *J Int Assoc Physicians AIDS Care (Chic)* 2009;8(6):367-374 [FREE Full text] [doi: [10.1177/1545109709352884](https://doi.org/10.1177/1545109709352884)] [Medline: [19952289](#)]
65. Reynolds NR, Sun J, Nagaraja HN, Gifford AL, Wu AW, Chesney MA. Optimizing measurement of self-reported adherence with the ACTG adherence questionnaire: a cross-protocol analysis. *J Acquir Immune Defic Syndr* 2007 Dec 1;46(4):402-409. [doi: [10.1097/qai.0b013e318158a44f](https://doi.org/10.1097/qai.0b013e318158a44f)] [Medline: [18077832](#)]
66. Simoni JM, Huh D, Wang Y, Wilson IB, Reynolds NR, Remien RH, et al. The validity of self-reported medication adherence as an outcome in clinical trials of adherence-promotion interventions: findings from the MACH14 study. *AIDS Behav* 2014 Dec;18(12):2285-2290 [FREE Full text] [doi: [10.1007/s10461-014-0905-x](https://doi.org/10.1007/s10461-014-0905-x)] [Medline: [25280447](#)]
67. Smith SR, Wahed AS, Kelley SS, Conjeevaram HS, Robuck PR, Fried MW, Virahep-C Study Group. Assessing the validity of self-reported medication adherence in hepatitis C treatment. *Ann Pharmacother* 2007 Jul;41(7):1116-1123. [doi: [10.1345/aph.1K024](https://doi.org/10.1345/aph.1K024)] [Medline: [17519299](#)]
68. van den Boogaard J, Lyimo R, Boeree M, Kibiki G, Aarnoutse R. Electronic monitoring of treatment adherence and validation of alternative adherence measures in tuberculosis patients: a pilot study. *Bull World Health Organ* 2011 Sep 1;89(9):632-639 [FREE Full text] [doi: [10.2471/BLT.11.086462](https://doi.org/10.2471/BLT.11.086462)] [Medline: [21897483](#)]
69. Chesney MA, Ickovics JR, Chambers DB, Gifford AL, Neidig J, Zwickl B, et al. Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: the AACTG adherence instruments. Patient care committee & adherence working group of the outcomes committee of the adult AIDS clinical trials group (AACTG). *AIDS Care* 2000 Jun;12(3):255-266. [doi: [10.1080/09540120050042891](https://doi.org/10.1080/09540120050042891)] [Medline: [10928201](#)]
70. Gabriel A, Violato C. Knowledge of and attitudes towards depression and adherence to treatment: the antidepressant adherence scale (AAS). *J Affect Disord* 2010 Nov;126(3):388-394. [doi: [10.1016/j.jad.2010.07.013](https://doi.org/10.1016/j.jad.2010.07.013)] [Medline: [20708273](#)]
71. Müller S, Kohlmann T, Wilke T. Validation of the adherence barriers questionnaire - an instrument for identifying potential risk factors associated with medication-related non-adherence. *BMC Health Serv Res* 2015 Apr 10;15:153 [FREE Full text] [doi: [10.1186/s12913-015-0809-0](https://doi.org/10.1186/s12913-015-0809-0)] [Medline: [25884193](#)]



72. Mueller S, Wilke T, Gorasso V, Erhart M, Kittner JM. Adaption and validation of the adherence barriers questionnaire for HIV patients on antiretroviral therapy (ABQ-HIV). *BMC Infect Dis* 2018 Nov 28;18(1):599 [FREE Full text] [doi: [10.1186/s12879-018-3530-x](https://doi.org/10.1186/s12879-018-3530-x)] [Medline: [30486795](#)]
73. Daouphars M, Ouvry M, Lenain P, Rouvet J, Jardin F, Bubenheim M, et al. Preliminary validation of self-assessment tool to measure imatinib adherence in patients with chronic myeloid leukemia. *Pharmacotherapy* 2013 Feb;33(2):152-156. [doi: [10.1002/phar.1174](https://doi.org/10.1002/phar.1174)] [Medline: [23359430](#)]
74. Breuil V, Cortet B, Cotté FE, Arnould B, Dias-Barbosa C, Gaudin A, et al. Validation of the adherence evaluation of osteoporosis treatment (ADEOS) questionnaire for osteoporotic post-menopausal women. *Osteoporos Int* 2012 Feb;23(2):445-455. [doi: [10.1007/s00198-011-1555-8](https://doi.org/10.1007/s00198-011-1555-8)] [Medline: [21625889](#)]
75. Martins MJ, Pinto AM, Castilho P, Macedo AF, Pereira AT, Bajouco M, et al. Assessing beliefs and attitudes towards antipsychotic medication from a recovery-based perspective: psychometric properties of a new scale. *Psychiatry Res* 2019 Mar;273:325-330. [doi: [10.1016/j.psychres.2019.01.043](https://doi.org/10.1016/j.psychres.2019.01.043)] [Medline: [30677722](#)]
76. Ayiesah R, Leonard J, Chong C. Development and validation of non-adherence to pulmonary rehabilitation questionnaire: a clinical tool for patients with chronic obstructive pulmonary diseases. *Clin Ter* 2014;165(3):123-128. [doi: [10.7417/CT.2014.1708](https://doi.org/10.7417/CT.2014.1708)] [Medline: [24999563](#)]
77. Kripalani S, Risser J, Gatti ME, Jacobson TA. Development and evaluation of the adherence to refills and medications scale (ARMS) among low-literacy patients with chronic disease. *Value Health* 2009;12(1):118-123 [FREE Full text] [doi: [10.1111/j.1524-4733.2008.00400.x](https://doi.org/10.1111/j.1524-4733.2008.00400.x)] [Medline: [19911444](#)]
78. Gökdoğan F, Kes D. Validity and reliability of the Turkish adherence to refills and medications scale. *Int J Nurs Pract* 2017 Oct;23(5). [doi: [10.1111/ijn.12566](https://doi.org/10.1111/ijn.12566)] [Medline: [28726281](#)]
79. Matza LS, Park J, Coyne KS, Skinner EP, Malley KG, Wolever RQ. Derivation and validation of the ASK-12 adherence barrier survey. *Ann Pharmacother* 2009 Oct;43(10):1621-1630. [doi: [10.1345/aph.1M174](https://doi.org/10.1345/aph.1M174)] [Medline: [19776298](#)]
80. Atsuta R, To Y, Sakamoto S, Mukai I, Kobayashi A, Kinoshita A, et al. Assessing usability of the 'adherence starts with knowledge 20' (ASK-20) questionnaire for Japanese adults with bronchial asthma receiving inhaled corticosteroids long term. *Allergol Int* 2017 Jul;66(3):411-417 [FREE Full text] [doi: [10.1016/j.alit.2016.09.001](https://doi.org/10.1016/j.alit.2016.09.001)] [Medline: [27712949](#)]
81. Hahn SR, Park J, Skinner EP, Yu-Isenberg KS, Weaver MB, Crawford B, et al. Development of the ASK-20 adherence barrier survey. *Curr Med Res Opin* 2008 Jul;24(7):2127-2138. [doi: [10.1185/03007990802174769](https://doi.org/10.1185/03007990802174769)] [Medline: [18554431](#)]
82. Matza LS, Yu-Isenberg KS, Coyne KS, Park J, Wakefield J, Skinner EP, et al. Further testing of the reliability and validity of the ASK-20 adherence barrier questionnaire in a medical center outpatient population. *Curr Med Res Opin* 2008 Nov;24(11):3197-3206. [doi: [10.1185/03007990802463642](https://doi.org/10.1185/03007990802463642)] [Medline: [18922213](#)]
83. Zeller A, Schroeder K, Peters TJ. An adherence self-report questionnaire facilitated the differentiation between nonadherence and nonresponse to antihypertensive treatment. *J Clin Epidemiol* 2008 Mar;61(3):282-288. [doi: [10.1016/j.jclinepi.2007.04.007](https://doi.org/10.1016/j.jclinepi.2007.04.007)] [Medline: [18226752](#)]
84. Moss AC, Lillis Y, Edwards George JB, Choudhry NK, Berg AH, Cheifetz AS, et al. Attitudes to mesalamine questionnaire: a novel tool to predict mesalamine nonadherence in patients with IBD. *Am J Gastroenterol* 2014 Dec;109(12):1850-1855. [doi: [10.1038/ajg.2014.158](https://doi.org/10.1038/ajg.2014.158)] [Medline: [24913040](#)]
85. Marsicano ED, Fernandes ND, Colugnati F, Grincenkov FR, Fernandes NM, De Geest S, et al. Transcultural adaptation and initial validation of Brazilian-Portuguese version of the Basel assessment of adherence to immunosuppressive medications scale (BAASIS) in kidney transplants. *BMC Nephrol* 2013 May 21;14:108 [FREE Full text] [doi: [10.1186/1471-2369-14-108](https://doi.org/10.1186/1471-2369-14-108)] [Medline: [23692889](#)]
86. Dolder CR, Lacro JP, Warren KA, Golshan S, Perkins DO, Jeste DV. Brief evaluation of medication influences and beliefs: development and testing of a brief scale for medication adherence. *J Clin Psychopharmacol* 2004 Aug;24(4):404-409. [doi: [10.1097/01.jcp.0000130554.63254.3a](https://doi.org/10.1097/01.jcp.0000130554.63254.3a)] [Medline: [15232332](#)]
87. McDonald-Miszczak L, Maris P, Fitzgibbon T, Ritchie G. A pilot study examining older adults' beliefs related to medication adherence: the BERMA survey. *J Aging Health* 2004 Nov;16(5):591-614. [doi: [10.1177/0898264304265772](https://doi.org/10.1177/0898264304265772)] [Medline: [15448274](#)]
88. Chui E, Wong K, Chan K, Wong M. Validation study of the brief medication adherence scale (BMAS) in patients with schizophrenia and related disorders in Hong Kong. *Asian J Psychiatr* 2018 Oct;37:154-160. [doi: [10.1016/j.ajp.2018.09.005](https://doi.org/10.1016/j.ajp.2018.09.005)] [Medline: [30265964](#)]
89. Bennett SJ, Milgrom LB, Champion V, Huster GA. Beliefs about medication and dietary compliance in people with heart failure: an instrument development study. *Heart Lung* 1997;26(4):273-279. [doi: [10.1016/s0147-9563\(97\)90084-4](https://doi.org/10.1016/s0147-9563(97)90084-4)] [Medline: [9257137](#)]
90. Sriwarakorn S, Krittiyanunt S, Sakulbumrungsil R. SENSITIVITY AND SPECIFICITY OF THAI-VERSION BRIEF MEDICATION QUESTIONNAIRE. *Journal of Health Research*. Jul-Sep 2010;24(3):129-134.
91. Svarstad BL, Chewning BA, Sleath BL, Claesson C. The brief medication questionnaire: a tool for screening patient adherence and barriers to adherence. *Patient Educ Couns* 1999 Jun;37(2):113-124. [doi: [10.1016/s0738-3991\(98\)00107-4](https://doi.org/10.1016/s0738-3991(98)00107-4)] [Medline: [14528539](#)]

92. Choo PW, Rand CS, Inui TS, Lee MT, Cain E, Cordeiro-Breault M, et al. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. *Med Care* 1999 Sep;37(9):846-857. [doi: [10.1097/00005650-199909000-00002](https://doi.org/10.1097/00005650-199909000-00002)] [Medline: [10493464](https://pubmed.ncbi.nlm.nih.gov/10493464/)]
93. Kerr SJ, Avihingsanon A, Puthachoen O, Chetchotisakd P, Layton M, Ubolyam S, et al. Assessing adherence in Thai patients taking combination antiretroviral therapy. *Int J STD AIDS* 2012 Mar;23(3):160-165. [doi: [10.1258/ijsa.2009.009152](https://doi.org/10.1258/ijsa.2009.009152)] [Medline: [22581867](https://pubmed.ncbi.nlm.nih.gov/22581867/)]
94. Mannheimer SB, Mukherjee R, Hirschhorn LR, Dougherty J, Celano SA, Ciccarone D, et al. The CASE adherence index: a novel method for measuring adherence to antiretroviral therapy. *AIDS Care* 2006 Oct;18(7):853-861 [FREE Full text] [doi: [10.1080/095401205000465160](https://doi.org/10.1080/095401205000465160)] [Medline: [16971298](https://pubmed.ncbi.nlm.nih.gov/16971298/)]
95. Kyngäs HA, Skaar-Chandler CA, Duffy ME. The development of an instrument to measure the compliance of adolescents with a chronic disease. *J Adv Nurs* 2000 Dec;32(6):1499-1506. [doi: [10.1046/j.1365-2648.2000.01611.x](https://doi.org/10.1046/j.1365-2648.2000.01611.x)] [Medline: [11136419](https://pubmed.ncbi.nlm.nih.gov/11136419/)]
96. Wales S, Crisp J, Fernandes R, Kyngas H. Modification and testing of the chronic disease compliance instrument to measure treatment compliance in adolescents with asthma. *Contemp Nurse* 2011 Oct;39(2):147-156. [doi: [10.5172/conu.2011.147](https://doi.org/10.5172/conu.2011.147)] [Medline: [22551427](https://pubmed.ncbi.nlm.nih.gov/22551427/)]
97. Fernandes R, Wales S, Crisp J, Kyngas H. Modification and testing of the chronic disease compliance instrument to measure treatment compliance in adolescents with diabetes. *J Clin Nurs* 2011 May;20(9-10):1273-1281. [doi: [10.1111/j.1365-2702.2010.03566.x](https://doi.org/10.1111/j.1365-2702.2010.03566.x)] [Medline: [21492273](https://pubmed.ncbi.nlm.nih.gov/21492273/)]
98. Dima AL, Schweitzer A, Diaconu R, Remor E, Wanless RS. Adherence to ARV medication in Romanian young adults: self-reported behaviour and psychological barriers. *Psychol Health Med* 2013;18(3):343-354. [doi: [10.1080/13548506.2012.722648](https://doi.org/10.1080/13548506.2012.722648)] [Medline: [22985131](https://pubmed.ncbi.nlm.nih.gov/22985131/)]
99. Kampman O, Lehtinen K, Lassila V. The reliability of compliance assessments performed by doctors and patients during neuroleptic treatment: a comparison of compliance ratings. *Acta Psychiatr Scand* Oct 2001;104(4):- . [doi: [10.1111/j.1600-0447.2001.00159.x](https://doi.org/10.1111/j.1600-0447.2001.00159.x)]
100. Krousel-Wood M, Joyce C, Holt EW, Levitan EB, Dornelles A, Webber LS, et al. Development and evaluation of a self-report tool to predict low pharmacy refill adherence in elderly patients with uncontrolled hypertension. *Pharmacotherapy* 2013 Aug;33(8):798-811 [FREE Full text] [doi: [10.1002/phar.1275](https://doi.org/10.1002/phar.1275)] [Medline: [23649849](https://pubmed.ncbi.nlm.nih.gov/23649849/)]
101. de Klerk E, van der Heijde D, van der Tempel H, van der Linden S. Development of a questionnaire to investigate patient compliance with antirheumatic drug therapy. *J Rheumatol* 1999 Dec;26(12):2635-2641. [Medline: [10606375](https://pubmed.ncbi.nlm.nih.gov/10606375/)]
102. Salt E, Hall L, Peden AR, Home R. Psychometric properties of three medication adherence scales in patients with rheumatoid arthritis. *J Nurs Meas* 2012;20(1):59-72. [doi: [10.1891/1061-3749.20.1.59](https://doi.org/10.1891/1061-3749.20.1.59)] [Medline: [22679710](https://pubmed.ncbi.nlm.nih.gov/22679710/)]
103. Lee JY, Lee SY, Hahn HJ, Son IJ, Hahn SG, Lee EB. Cultural adaptation of a compliance questionnaire for patients with rheumatoid arthritis to a Korean version. *Korean J Intern Med* 2011 Mar;26(1):28-33 [FREE Full text] [doi: [10.3904/kjim.2011.26.1.28](https://doi.org/10.3904/kjim.2011.26.1.28)] [Medline: [21437159](https://pubmed.ncbi.nlm.nih.gov/21437159/)]
104. Cinar F, Cinar M, Yilmaz S, Acikel C, Erdem H, Pay S, et al. AB0658 Cross-cultural adaptation, reliability, and validity of the turkish version of the compliance questionnaire on rheumatology (CQR-T) in patients with Behçet's disease. *Ann Rheum Dis* 2015 Jun 9;74(Suppl 2):1118.3-1118.9. [doi: [10.1136/annrheumdis-2015-eular.3171](https://doi.org/10.1136/annrheumdis-2015-eular.3171)]
105. Hughes LD, Done J, Young A. A 5 item version of the compliance questionnaire for rheumatology (CQR5) successfully identifies low adherence to DMARDs. *BMC Musculoskelet Disord* 2013 Oct 8;14:286 [FREE Full text] [doi: [10.1186/1471-2474-14-286](https://doi.org/10.1186/1471-2474-14-286)] [Medline: [24103582](https://pubmed.ncbi.nlm.nih.gov/24103582/)]
106. Hogan TP, Awad AG, Eastwood R. A self-report scale predictive of drug compliance in schizophrenics: reliability and discriminative validity. *Psychol Med* 1983 Feb;13(1):177-183. [doi: [10.1017/s0033291700050182](https://doi.org/10.1017/s0033291700050182)] [Medline: [6133297](https://pubmed.ncbi.nlm.nih.gov/6133297/)]
107. Friemann K, Wciórka J. [Four measures of treatment compliance among patients recovering from psychotic episodes--a comparative study]. *Psychiatr Pol* 2013;47(5):759-773. [doi: [10.12740/pp/18341](https://doi.org/10.12740/pp/18341)] [Medline: [25011225](https://pubmed.ncbi.nlm.nih.gov/25011225/)]
108. Kikkert MJ, Koeter MW, Dekker JJ, Burti L, Robson D, Puschner B, et al. The predictive validity of subjective adherence measures in patients with schizophrenia. *Int J Methods Psychiatr Res* 2011 Jun;20(2):73-81 [FREE Full text] [doi: [10.1002/mpr.335](https://doi.org/10.1002/mpr.335)] [Medline: [21557378](https://pubmed.ncbi.nlm.nih.gov/21557378/)]
109. Stjernswärd S, Persson K, Nielsen R, Tuninger E, Levander S. A modified drug attitude inventory used in long-term patients in sheltered housing. *Eur Neuropsychopharmacol* 2013 Oct;23(10):1296-1299. [doi: [10.1016/j.euroneuro.2012.11.011](https://doi.org/10.1016/j.euroneuro.2012.11.011)] [Medline: [23265955](https://pubmed.ncbi.nlm.nih.gov/23265955/)]
110. Garfield S, Eliasson L, Clifford S, Willson A, Barber N. Developing the diagnostic adherence to medication scale (the DAMS) for use in clinical practice. *BMC Health Serv Res* 2012 Oct 8;12:350 [FREE Full text] [doi: [10.1186/1472-6963-12-350](https://doi.org/10.1186/1472-6963-12-350)] [Medline: [23039138](https://pubmed.ncbi.nlm.nih.gov/23039138/)]
111. Jacobsen R, Møldrup C, Christrup L, Sjøgren P, Hansen O. The Danish version of the medication adherence report scale: preliminary validation in cancer pain patients. *Pain Pract* 2009;9(1):1-7. [doi: [10.1111/j.1533-2500.2008.00245.x](https://doi.org/10.1111/j.1533-2500.2008.00245.x)] [Medline: [19019056](https://pubmed.ncbi.nlm.nih.gov/19019056/)]
112. Mehta SN, Nansel TR, Volkening LK, Butler DA, Haynie DL, Laffel LM. Validation of a contemporary adherence measure for children with type 1 diabetes: the diabetes management questionnaire. *Diabet Med* 2015 Sep;32(9):1232-1238 [FREE Full text] [doi: [10.1111/dme.12682](https://doi.org/10.1111/dme.12682)] [Medline: [26280463](https://pubmed.ncbi.nlm.nih.gov/26280463/)]



113. Sleath B, Carpenter D, Blalock S, Davis S, Hickson R, Lee C, et al. Development of a new diabetes medication self-efficacy scale and its association with both reported problems in using diabetes medications and self-reported adherence. *Patient Prefer Adherence* 2016;10:1003-1010 [[FREE Full text](#)] [doi: [10.2147/PPA.S101349](#)] [Medline: [27354769](#)]
114. Zheng J, Wang Y, Ye X, Xiao L, Ye J, Li X, et al. Validation of diabetes medication self-efficacy scale in Chinese with type 2 diabetes. *Patient Prefer Adherence* 2018;12:2517-2525 [[FREE Full text](#)] [doi: [10.2147/PPA.S170144](#)] [Medline: [30568430](#)]
115. Irvine AA, Saunders JT, Blank MB, Carter WR. Validation of scale measuring environmental barriers to diabetes-regimen adherence. *Diabetes Care* 1990 Jul;13(7):705-711. [doi: [10.2337/diacare.13.7.705](#)] [Medline: [2387191](#)]
116. Regnault A, Viala-Danten M, Gilet H, Berdeaux G. Scoring and psychometric properties of the eye-drop satisfaction questionnaire (EDSQ), an instrument to assess satisfaction and compliance with glaucoma treatment. *BMC Ophthalmol* 2010 Feb 1;10:1 [[FREE Full text](#)] [doi: [10.1186/1471-2415-10-1](#)] [Medline: [20122146](#)]
117. Nordmann J, Denis P, Vigneux M, Trudeau E, Guillemin I, Berdeaux G. Development of the conceptual framework for the eye-drop satisfaction questionnaire (EDSQ) in glaucoma using a qualitative study. *BMC Health Serv Res* 2007 Aug 6;7:124 [[FREE Full text](#)] [doi: [10.1186/1472-6963-7-124](#)] [Medline: [17683594](#)]
118. Kim Y, Evangelista LS, Phillips LR, Pavlish C, Kopple JD. The end-stage renal disease adherence questionnaire (ESRD-AQ): testing the psychometric properties in patients receiving in-center hemodialysis. *Nephrol Nurs J* 2010;37(4):377-393 [[FREE Full text](#)] [Medline: [20830945](#)]
119. Poveda V, Amado L, Filgueiras M, Teixeira L, Miranda V, Santos-Silva A, et al. End-stage renal disease adherence questionnaire: translation and validation to the Portuguese language. *Ren Fail* 2016 Nov;38(10):1633-1638. [doi: [10.1080/0886022X.2016.1209063](#)] [Medline: [27764984](#)]
120. Fairley C, Permana A, Read T. Long-term utility of measuring adherence by self-report compared with pharmacy record in a routine clinic setting. *HIV Med* 2005 Sep;6(5):366-369 [[FREE Full text](#)] [doi: [10.1111/j.1468-1293.2005.00322.x](#)] [Medline: [16156886](#)]
121. Kandrotaitė K, Smigelskas K, Janusauskienė D, Jievaltas M, Maciulaitis R, Briedis V. Development of a short questionnaire to identify the risk of nonadherence to antibiotic treatment. *Curr Med Res Opin* 2013 Nov;29(11):1555-1563. [doi: [10.1185/03007995.2013.835255](#)] [Medline: [23952367](#)]
122. Zongo A, Guénette L, Moisan J, Grégoire JP. Predictive validity of self-reported measures of adherence to noninsulin antidiabetes medication against control of glycated hemoglobin levels. *Can J Diabetes* 2016 Feb;40(1):58-65. [doi: [10.1016/j.cjcd.2015.06.008](#)] [Medline: [26507401](#)]
123. Naqvi AA, AlShayban DM, Ghorri SA, Mahmoud MA, Haseeb A, Faidah HS, et al. Validation of the general medication adherence scale in Saudi patients with chronic diseases. *Front Pharmacol* 2019;10:633 [[FREE Full text](#)] [doi: [10.3389/fphar.2019.00633](#)] [Medline: [31231222](#)]
124. Naqvi AA, Hassali MA, Jahangir A, Nadir MN, Kachela B. Translation and validation of the English version of the general medication adherence scale (GMAS) in patients with chronic illnesses. *J Drug Assess* 2019;8(1):36-42 [[FREE Full text](#)] [doi: [10.1080/21556660.2019.1579729](#)] [Medline: [30863660](#)]
125. Naqvi AA, Hassali MA, Rizvi M, Zehra A, Iffat W, Haseeb A, et al. Development and validation of a novel general medication adherence scale (GMAS) for chronic illness patients in Pakistan. *Front Pharmacol* 2018 Oct 9;9:1124 [[FREE Full text](#)] [doi: [10.3389/fphar.2018.01124](#)] [Medline: [30356775](#)]
126. Mansberger SL, Sheppler CR, McClure TM, Vanalstine CL, Swanson IL, Stoumbos Z, et al. Psychometrics of a new questionnaire to assess glaucoma adherence: the glaucoma treatment compliance assessment tool (an American ophthalmological society thesis). *Trans Am Ophthalmol Soc* 2013 Sep;111:1-16 [[FREE Full text](#)] [Medline: [24072942](#)]
127. Abe RY, Wen LD, Barker GT, Mansberger SL. Psychometric Properties of the glaucoma treatment compliance assessment tool (GTCAT) in a Brazilian population. *J Glaucoma* 2018 Mar;27(3):257-265. [doi: [10.1097/IJG.0000000000000876](#)] [Medline: [29369851](#)]
128. Barker GT, Mansberger SL. Psychometric properties of the reduced version of the glaucoma treatment compliance assessment tool (GTCAT). *Ophthalmic Epidemiol* 2019 Feb;26(1):55-62. [doi: [10.1080/09286586.2018.1516785](#)] [Medline: [30204034](#)]
129. Song Y, Han H, Song H, Nam S, Nguyen T, Kim MT. Psychometric evaluation of hill-bone medication adherence subscale. *Asian Nurs Res (Korean Soc Nurs Sci)* 2011 Sep;5(3):183-188 [[FREE Full text](#)] [doi: [10.1016/j.anr.2011.09.007](#)] [Medline: [25030368](#)]
130. Koschack J, Marx G, Schnakenberg J, Kochen MM, Himmel W. Comparison of two self-rating instruments for medication adherence assessment in hypertension revealed insufficient psychometric properties. *J Clin Epidemiol* 2010 Mar;63(3):299-306. [doi: [10.1016/j.jclinepi.2009.06.011](#)] [Medline: [19762213](#)]
131. Krousel-Wood M, Muntner P, Jannu A, Desalvo K, Re RN. Reliability of a medication adherence measure in an outpatient setting. *Am J Med Sci* 2005 Sep;330(3):128-133. [doi: [10.1097/00000441-200509000-00006](#)] [Medline: [16174996](#)]
132. Nogueira-Silva L, Sá-Sousa A, Lima MJ, Monteiro A, Dennison-Himmelfarb C, Fonseca JA. Translation and cultural adaptation of the hill-bone compliance to high blood pressure therapy scale to Portuguese. *Rev Port Cardiol* 2016 Feb;35(2):93-97 [[FREE Full text](#)] [doi: [10.1016/j.repc.2015.07.013](#)] [Medline: [26852304](#)]
133. Alsolami FJ, Hou X, Correa-Velez I. An arabic instrument to measure medication adherence in Saudi hypertensive patients. *ME-JFM* 2013 Sep;11(7):17-23. [doi: [10.5742/mefm.2014.92409](#)]

134. Nelsen A, Trautner BW, Petersen NJ, Gupta S, Rodriguez-Barradas M, Giordano TP, et al. Development and validation of a measure for intention to adhere to HIV treatment. *AIDS Patient Care STDS* 2012 Jun;26(6):329-334 [[FREE Full text](#)] [doi: [10.1089/apc.2011.0318](#)] [Medline: [22680281](#)]
135. Spire B, Arnould B, Barbier F, Durant J, Gilquin J, Landman R, et al. Simplification and first validation of a short battery of patient questionnaires for clinical management of HIV-infected patients: the HIV-SQUAD (symptom quality of life adherence) questionnaire. *HIV Clin Trials* 2009;10(4):215-232. [doi: [10.1310/hct1004-215](#)] [Medline: [19723610](#)]
136. Mikhael EM, Hussain SA, Shawky N, Hassali MA. Validity and reliability of anti-diabetic medication adherence scale among patients with diabetes in Baghdad, Iraq: a pilot study. *BMJ Open Diabetes Res Care* 2019;7(1):e000658 [[FREE Full text](#)] [doi: [10.1136/bmjdr-2019-000658](#)] [Medline: [31354953](#)]
137. Jesus-Nunes AP, Morais-de-Jesus M, Dantas-Duarte A, Moreira TM, Argolo FC, Castro AD, et al. The Portuguese version of the immunosuppressant therapy adherence scale (ITAS) among liver transplant recipient patients: translation and psychometric properties. *Ann Hepatol* 2018;17(1):104-109 [[FREE Full text](#)] [doi: [10.5604/01.3001.0010.7541](#)] [Medline: [29311393](#)]
138. Wilks S, Spivey C, Chisholm-Burns M. Psychometric re-evaluation of the immunosuppressant therapy adherence scale among solid-organ transplant recipients. *J Eval Clin Pract* 2010 Feb;16(1):64-68. [doi: [10.1111/j.1365-2753.2008.01115.x](#)] [Medline: [20367816](#)]
139. Chisholm MA, Lance CE, Williamson GM, Mulloy LL. Development and validation of an immunosuppressant therapy adherence barrier instrument. *Nephrol Dial Transplant* 2005 Jan;20(1):181-188. [doi: [10.1093/ndt/gfh576](#)] [Medline: [15572384](#)]
140. de Geest S, Abraham I, Gemoets H, Evers G. Development of the long-term medication behaviour self-efficacy scale: qualitative study for item development. *J Adv Nurs* 1994 Feb;19(2):233-238. [doi: [10.1111/j.1365-2648.1994.tb01076.x](#)] [Medline: [8188953](#)]
141. Lee S, Bae Y, Worley M, Law A. Validating the modified drug adherence work-up (M-DRAW) tool to identify and address barriers to medication adherence. *Pharmacy (Basel)* 2017 Sep 8;5(3):- [[FREE Full text](#)] [doi: [10.3390/pharmacy5030052](#)] [Medline: [28970464](#)]
142. Lee S, Bae-Shaaw YH, Gogineni H, Worley MM, Law AV. Triple strength utility of the modified drug adherence work-up (M-DRAW) tool in a veterans affairs outpatient diabetes clinic. *Res Social Adm Pharm* 2020 Jul;16(7):914-920. [doi: [10.1016/j.sapharm.2019.09.063](#)] [Medline: [31629655](#)]
143. Sadakathulla I, Mateti U, Kellarai A, Bhat K. Adhering to antihypertensive treatment is vitally important. *Article Salud(i)Ciencia* 2019;23(4):314-324. [doi: [10.21840/sic/157368](#)]
144. Unni EJ, Olson JL, Farris KB. Revision and validation of medication adherence reasons scale (MAR-scale). *Curr Med Res Opin* 2014 Feb;30(2):211-221. [doi: [10.1185/03007995.2013.851075](#)] [Medline: [24102296](#)]
145. Unni E, Sternbach N, Goren A. Using the medication adherence reasons scale (MAR-Scale) to identify the reasons for non-adherence across multiple disease conditions. *Patient Prefer Adherence* 2019;13:993-1004 [[FREE Full text](#)] [doi: [10.2147/PPA.S205359](#)] [Medline: [31308635](#)]
146. Fond G, Boyer L, Boucekine M, Aden LA, Schürhoff F, Tessier A, FACE-SZ (Fonda and Mental Academic Centers of Expertise for Schizophrenia) group. Validation study of the medication adherence rating scale. Results from the FACE-SZ national dataset. *Schizophr Res* 2017 Apr;182:84-89. [doi: [10.1016/j.schres.2016.10.023](#)] [Medline: [27789187](#)]
147. Kao Y, Liu Y. Compliance and schizophrenia: the predictive potential of insight into illness, symptoms, and side effects. *Compr Psychiatry* 2010;51(6):557-565. [doi: [10.1016/j.comppsy.2010.03.007](#)] [Medline: [20965300](#)]
148. Owie GO, Olotu SO, James BO. Reliability and validity of the medication adherence rating scale in a cohort of patients with schizophrenia from Nigeria. *Trends Psychiatry Psychother* 2018;40(2):85-92 [[FREE Full text](#)] [doi: [10.1590/2237-6089-2017-0077](#)] [Medline: [29768528](#)]
149. Sowunmi O, Onifade P. Psychometric evaluation of medication adherence rating scale (MARS) among Nigerian patients with schizophrenia. *Niger J Clin Pract* 2019 Sep;22(9):1281-1285 [[FREE Full text](#)] [doi: [10.4103/njcp.njcp\\_325\\_18](#)] [Medline: [31489867](#)]
150. Zemmour K, Tinland A, Boucekine M, Girard V, Loubière S, Resseguier N, French Housing First Study Group. Validation of the medication adherence rating scale in homeless patients with schizophrenia: results from the French housing first experience. *Sci Rep* 2016 Aug 18;6:31598 [[FREE Full text](#)] [doi: [10.1038/srep31598](#)] [Medline: [27534796](#)]
151. Cohen JL, Mann DM, Wisnivesky JP, Home R, Leventhal H, Musumeci-Szabó TJ, et al. Assessing the validity of self-reported medication adherence among inner-city asthmatic adults: the medication adherence report scale for asthma. *Ann Allergy Asthma Immunol* 2009 Oct;103(4):325-331. [doi: [10.1016/s1081-1206\(10\)60532-7](#)] [Medline: [19852197](#)]
152. Alsous M, Alhalaiqa F, Abu Farha R, Abdel Jalil M, McElnay J, Horne R. Reliability and validity of arabic translation of medication adherence report scale (MARS) and beliefs about medication questionnaire (BMQ)-specific for use in children and their parents. *PLoS One* 2017;12(2):e0171863 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0171863](#)] [Medline: [28192467](#)]
153. Garcia-Marcos PW, Brand PL, Kaptein AA, Klok T. Is the MARS questionnaire a reliable measure of medication adherence in childhood asthma? *J Asthma* 2016 Dec;53(10):1085-1089. [doi: [10.1080/02770903.2016.1180699](#)] [Medline: [27177241](#)]

154. Jónsdóttir H, Opjordsmoen S, Birkenaes AB, Engh JA, Ringen PA, Vaskinn A, et al. Medication adherence in outpatients with severe mental disorders: relation between self-reports and serum level. *J Clin Psychopharmacol* 2010 Apr;30(2):169-175. [doi: [10.1097/JCP.0b013e3181d2191e](https://doi.org/10.1097/JCP.0b013e3181d2191e)] [Medline: [20520290](https://pubmed.ncbi.nlm.nih.gov/20520290/)]
155. Lee L, El-Den S, Horne R, Carter SR. Patient satisfaction with information, concerns, beliefs and adherence to topical corticosteroids. *Patient Educ Couns* 2019 Jun;102(6):1203-1209. [doi: [10.1016/j.pec.2019.01.019](https://doi.org/10.1016/j.pec.2019.01.019)] [Medline: [30928342](https://pubmed.ncbi.nlm.nih.gov/30928342/)]
156. Lin C, Ou H, Nikoobakht M, Broström A, Årestedt K, Pakpour A. Validation of the 5-item medication adherence report scale in older stroke patients in Iran. *J Cardiovasc Nurs* 2018;33(6):536-543. [doi: [10.1097/JCN.0000000000000488](https://doi.org/10.1097/JCN.0000000000000488)] [Medline: [29649015](https://pubmed.ncbi.nlm.nih.gov/29649015/)]
157. Tommelein E, Mehuys E, van Tongelen I, Brusselle G, Boussery K. Accuracy of the medication adherence report scale (MARS-5) as a quantitative measure of adherence to inhalation medication in patients with COPD. *Ann Pharmacother* 2014 May;48(5):589-595. [doi: [10.1177/1060028014522982](https://doi.org/10.1177/1060028014522982)] [Medline: [24523393](https://pubmed.ncbi.nlm.nih.gov/24523393/)]
158. van de Steeg N, Sielk M, Pentzek M, Bakx C, Altiner A. Drug-adherence questionnaires not valid for patients taking blood-pressure-lowering drugs in a primary health care setting. *J Eval Clin Pract* 2009 Jun;15(3):468-472. [doi: [10.1111/j.1365-2753.2008.01038.x](https://doi.org/10.1111/j.1365-2753.2008.01038.x)] [Medline: [19366389](https://pubmed.ncbi.nlm.nih.gov/19366389/)]
159. Mahler C, Hermann K, Horne R, Ludt S, Haefeli WE, Szecsenyi J, et al. Assessing reported adherence to pharmacological treatment recommendations. Translation and evaluation of the medication adherence report scale (MARS) in Germany. *J Eval Clin Pract* 2010 Jun;16(3):574-579. [doi: [10.1111/j.1365-2753.2009.01169.x](https://doi.org/10.1111/j.1365-2753.2009.01169.x)] [Medline: [20210821](https://pubmed.ncbi.nlm.nih.gov/20210821/)]
160. Sampaio R, Azevedo LF, Dias CC, Horne R, Castro Lopes JM. Portuguese version of the medication adherence report scale (MARS-9): validation in a population of chronic pain patients. *J Eval Clin Pract* 2019 Apr;25(2):346-352. [doi: [10.1111/jep.13098](https://doi.org/10.1111/jep.13098)] [Medline: [30648328](https://pubmed.ncbi.nlm.nih.gov/30648328/)]
161. Ogedegbe G, Mancuso CA, Allegrante JP, Charlson ME. Development and evaluation of a medication adherence self-efficacy scale in hypertensive African-American patients. *J Clin Epidemiol* 2003 Jun;56(6):520-529. [doi: [10.1016/s0895-4356\(03\)00053-2](https://doi.org/10.1016/s0895-4356(03)00053-2)] [Medline: [12873646](https://pubmed.ncbi.nlm.nih.gov/12873646/)]
162. Saffari M, Zeidi IM, Fridlund B, Chen H, Pakpour AH. A persian adaptation of medication adherence self-efficacy scale (MASES) in hypertensive patients: psychometric properties and factor structure. *High Blood Press Cardiovasc Prev* 2015 Sep;22(3):247-255. [doi: [10.1007/s40292-015-0101-8](https://doi.org/10.1007/s40292-015-0101-8)] [Medline: [25986076](https://pubmed.ncbi.nlm.nih.gov/25986076/)]
163. Andy UU, Harvie HS, Smith AL, Propert KJ, Bogner HR, Arya LA. Validation of a self-administered instrument to measure adherence to anticholinergic drugs in women with overactive bladder. *Neurourol Urodyn* 2015 Jun;34(5):424-428 [FREE Full text] [doi: [10.1002/nau.22605](https://doi.org/10.1002/nau.22605)] [Medline: [24719232](https://pubmed.ncbi.nlm.nih.gov/24719232/)]
164. Koneru S, Shishov M, Ware A, Farhey Y, Mongey A, Graham TB, et al. Effectively measuring adherence to medications for systemic lupus erythematosus in a clinical setting. *Arthritis Rheum* 2007 Aug 15;57(6):1000-1006 [FREE Full text] [doi: [10.1002/art.22898](https://doi.org/10.1002/art.22898)] [Medline: [17665465](https://pubmed.ncbi.nlm.nih.gov/17665465/)]
165. Kosilov K, Loparev S, Kuzina I, Kosilova L, Ivanovskaya M, Prokofyeva A. Effectiveness of a new tool for self-evaluation of adherence to antimuscarinic drug treatment in older patients of both sexes with urge incontinence. *Geriatr Gerontol Int* 2018 Jan;18(1):115-122. [doi: [10.1111/ggi.13150](https://doi.org/10.1111/ggi.13150)] [Medline: [28921819](https://pubmed.ncbi.nlm.nih.gov/28921819/)]
166. Kosilov K, Loparev S, Kuzina I, Shakirova O, Zhuravskaya N, Lobodenko A. Self-assessment of treatment compliance with antimuscarinic drugs and lower urinary tract condition among women with urinary incontinence. *Int Urogynecol J* 2017 Nov;28(11):1663-1669. [doi: [10.1007/s00192-017-3333-4](https://doi.org/10.1007/s00192-017-3333-4)] [Medline: [28429053](https://pubmed.ncbi.nlm.nih.gov/28429053/)]
167. Kosilov K, Loparev S, Kuzina I, Shakirova O, Zhuravskaya N, Lobodenko A. The effective tool for self-assessment of adherence to treatment in patients with benign prostatic obstruction and overactive bladder symptoms. *Aging Male* 2017 Mar;20(1):39-44. [doi: [10.1080/13685538.2016.1247435](https://doi.org/10.1080/13685538.2016.1247435)] [Medline: [28074677](https://pubmed.ncbi.nlm.nih.gov/28074677/)]
168. Kosilov KV, Loparev SA, Kuzina IG, Shakirova OV, Gainullina YI, Kosilova LV, et al. A new tool for self-evaluation of adherence to antimuscarinic drugs treatment in patients with urinary incontinence. *Arab J Urol* 2017 Dec;15(4):372-379 [FREE Full text] [doi: [10.1016/j.aju.2017.10.007](https://doi.org/10.1016/j.aju.2017.10.007)] [Medline: [29234543](https://pubmed.ncbi.nlm.nih.gov/29234543/)]
169. Ueno H, Yamazaki Y, Yonekura Y, Park M, Ishikawa H, Kiuchi T. Reliability and validity of a 12-item medication adherence scale for patients with chronic disease in Japan. *BMC Health Serv Res* 2018 Jul 31;18(1):592 [FREE Full text] [doi: [10.1186/s12913-018-3380-7](https://doi.org/10.1186/s12913-018-3380-7)] [Medline: [30064422](https://pubmed.ncbi.nlm.nih.gov/30064422/)]
170. Battistella M, Fleites R, Wong R, Jassal SV. Development, validation, and implementation of a medication adherence survey to seek a better understanding of the hemodialysis patient. *Clin Nephrol* 2016 Jan;85(1):12-22. [doi: [10.5414/CN108654](https://doi.org/10.5414/CN108654)] [Medline: [26636327](https://pubmed.ncbi.nlm.nih.gov/26636327/)]
171. Athavale AS, Bentley JP, Banahan BF, McCaffrey DJ, Pace PF, Vorhies DW. Development of the medication adherence estimation and differentiation scale (MEDS). *Curr Med Res Opin* 2019 Apr;35(4):577-585. [doi: [10.1080/03007995.2018.1512478](https://doi.org/10.1080/03007995.2018.1512478)] [Medline: [30106315](https://pubmed.ncbi.nlm.nih.gov/30106315/)]
172. Cate H, Bhattacharya D, Clark A, Holland R, Broadway DC. A comparison of measures used to describe adherence to glaucoma medication in a randomised controlled trial. *Clin Trials* 2015 Dec;12(6):608-617. [doi: [10.1177/1740774515592636](https://doi.org/10.1177/1740774515592636)] [Medline: [26183437](https://pubmed.ncbi.nlm.nih.gov/26183437/)]
173. Cohen HW, Shmukler C, Ullman R, Rivera CM, Walker EA. Measurements of medication adherence in diabetic patients with poorly controlled HbA(1c). *Diabet Med* 2010 Feb;27(2):210-216 [FREE Full text] [doi: [10.1111/j.1464-5491.2009.02898.x](https://doi.org/10.1111/j.1464-5491.2009.02898.x)] [Medline: [20546266](https://pubmed.ncbi.nlm.nih.gov/20546266/)]



174. Guénette L, Moisan J, Préville M, Boyer R. Measures of adherence based on self-report exhibited poor agreement with those based on pharmacy records. *J Clin Epidemiol* 2005 Sep;58(9):924-933. [doi: [10.1016/j.jclinepi.2005.02.002](https://doi.org/10.1016/j.jclinepi.2005.02.002)] [Medline: [16085196](https://pubmed.ncbi.nlm.nih.gov/16085196/)]
175. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care* 1986 Jan;24(1):67-74. [doi: [10.1097/00005650-198601000-00007](https://doi.org/10.1097/00005650-198601000-00007)] [Medline: [3945130](https://pubmed.ncbi.nlm.nih.gov/3945130/)]
176. Shalansky SJ, Levy AR, Ignaszewski AP. Self-reported Morisky score for identifying nonadherence with cardiovascular medications. *Ann Pharmacother* 2004 Sep;38(9):1363-1368. [doi: [10.1345/aph.1E071](https://doi.org/10.1345/aph.1E071)] [Medline: [15238622](https://pubmed.ncbi.nlm.nih.gov/15238622/)]
177. Brooks CM, Richards JM, Kohler CL, Soong S, Martin B, Windsor RA, et al. Assessing adherence to asthma medication and inhaler regimens: a psychometric analysis of adult self-report scales. *Med Care* 1994 Mar;32(3):298-307. [doi: [10.1097/00005650-199403000-00008](https://doi.org/10.1097/00005650-199403000-00008)] [Medline: [8145604](https://pubmed.ncbi.nlm.nih.gov/8145604/)]
178. Erickson SR, Coombs JH, Kirking DM, Azimi AR. Compliance from self-reported versus pharmacy claims data with metered-dose inhalers. *Ann Pharmacother* 2001 Sep;35(9):997-1003. [doi: [10.1345/aph.10379](https://doi.org/10.1345/aph.10379)] [Medline: [11573875](https://pubmed.ncbi.nlm.nih.gov/11573875/)]
179. Kristina SA, Putri LR, Riani DA, Ikawati Z, Endarti D. Validity of self-reported measure of medication adherence among diabetic patients in Indonesia. *Int Res J Pharm* 2019 Aug 2;10(7):144-148. [doi: [10.7897/2230-8407.1007234](https://doi.org/10.7897/2230-8407.1007234)]
180. Wang Y, Lee J, Toh M, Tang W, Ko Y. Validity and reliability of a self-reported measure of medication adherence in patients with type 2 diabetes mellitus in Singapore. *Diabet Med* 2012 Sep;29(9):e338-e344. [doi: [10.1111/j.1464-5491.2012.03733.x](https://doi.org/10.1111/j.1464-5491.2012.03733.x)] [Medline: [22672497](https://pubmed.ncbi.nlm.nih.gov/22672497/)]
181. Prado JC, Kupek E, Mion D. Validity of four indirect methods to measure adherence in primary care hypertensives. *J Hum Hypertens* 2007 Jul;21(7):579-584. [doi: [10.1038/sj.jhh.1002196](https://doi.org/10.1038/sj.jhh.1002196)] [Medline: [17443212](https://pubmed.ncbi.nlm.nih.gov/17443212/)]
182. Jerant A, DiMatteo R, Arnsten J, Moore-Hill M, Franks P. Self-report adherence measures in chronic illness: retest reliability and predictive validity. *Med Care* 2008 Nov;46(11):1134-1139. [doi: [10.1097/MLR.0b013e31817924e4](https://doi.org/10.1097/MLR.0b013e31817924e4)] [Medline: [18953223](https://pubmed.ncbi.nlm.nih.gov/18953223/)]
183. Fahey M, Abdulmajeed A, Sabra K. Measurement of adherence to anti-hypertensive medication as perceived by doctors and patients. *Qatar Med J* 2006 Jun;2006(1):16. [doi: [10.5339/qmj.2006.1.16](https://doi.org/10.5339/qmj.2006.1.16)]
184. Ali MA, Abou-Taleb DA, Mohamed RR. Treatment adherence and beliefs about medicines among Egyptian vitiligo patients. *Dermatol Ther* 2016 Nov;29(6):413-418. [doi: [10.1111/dth.12397](https://doi.org/10.1111/dth.12397)] [Medline: [27594551](https://pubmed.ncbi.nlm.nih.gov/27594551/)]
185. Pandey A, Raza F, Velasco A, Brinker S, Ayers C, Das SR, et al. Comparison of Morisky medication adherence scale with therapeutic drug monitoring in apparent treatment-resistant hypertension. *J Am Soc Hypertens* 2015 Jun;9(6):420-6.e2. [doi: [10.1016/j.jash.2015.04.004](https://doi.org/10.1016/j.jash.2015.04.004)] [Medline: [26051923](https://pubmed.ncbi.nlm.nih.gov/26051923/)]
186. Ashur S, Shamsuddin K, Shah S, Bosseri S, Morisky DE. Reliability and known-group validity of the Arabic version of the 8-item Morisky medication adherence scale among type 2 diabetes mellitus patients. *East Mediterr Health J* 2015 Dec 13;21(10):722-728. [doi: [10.26719/2015.21.10.722](https://doi.org/10.26719/2015.21.10.722)] [Medline: [26750162](https://pubmed.ncbi.nlm.nih.gov/26750162/)]
187. de las Cuevas C, Peñate W. Psychometric properties of the eight-item Morisky medication adherence scale (MMAS-8) in a psychiatric outpatient setting. *Int J Clin Health Psychol* 2015;15(2):121-129 [FREE Full text] [doi: [10.1016/j.ijchp.2014.11.003](https://doi.org/10.1016/j.ijchp.2014.11.003)] [Medline: [30487829](https://pubmed.ncbi.nlm.nih.gov/30487829/)]
188. de Oliveira-Filho AD, Morisky DE, Neves SJF, Costa FA, de Lyra DP. The 8-item Morisky medication adherence scale: validation of a Brazilian-Portuguese version in hypertensive adults. *Res Social Adm Pharm* 2014;10(3):554-561. [doi: [10.1016/j.sapharm.2013.10.006](https://doi.org/10.1016/j.sapharm.2013.10.006)] [Medline: [24268603](https://pubmed.ncbi.nlm.nih.gov/24268603/)]
189. DiBonaventura M, Wintfeld N, Huang J, Goren A. The association between nonadherence and glycated hemoglobin among type 2 diabetes patients using basal insulin analogs. *Patient Prefer Adherence* 2014;8:873-882 [FREE Full text] [doi: [10.2147/PPA.S55550](https://doi.org/10.2147/PPA.S55550)] [Medline: [24971002](https://pubmed.ncbi.nlm.nih.gov/24971002/)]
190. Fabbrini G, Abbruzzese G, Barone P, Antonini A, Tinazzi M, Castegnaro G, REASON study group. Adherence to anti-Parkinson drug therapy in the 'REASON' sample of Italian patients with Parkinson's disease: the linguistic validation of the Italian version of the 'Morisky medical adherence scale-8 items'. *Neurol Sci* 2013 Nov;34(11):2015-2022. [doi: [10.1007/s10072-013-1438-1](https://doi.org/10.1007/s10072-013-1438-1)] [Medline: [23728715](https://pubmed.ncbi.nlm.nih.gov/23728715/)]
191. George J, Vuong T, Bailey MJ, Kong DC, Marriott JL, Stewart K. Medication regimen complexity and adherence in patients at risk of medication misadventure. *J Pharm Pract Res* 2015 Apr 13;36(2):99-102. [doi: [10.1002/j.2055-2335.2006.tb00580.x](https://doi.org/10.1002/j.2055-2335.2006.tb00580.x)]
192. Jankowska-Polanska B, Uchmanowicz I, Chudiak A, Dudek K, Morisky D, Szymanska-Chabowska A. Psychometric properties of the Polish version of the eight-item Morisky medication adherence scale in hypertensive adults. *Patient Prefer Adherence* 2016;10:1759-1766 [FREE Full text] [doi: [10.2147/PPA.S101904](https://doi.org/10.2147/PPA.S101904)] [Medline: [27672314](https://pubmed.ncbi.nlm.nih.gov/27672314/)]
193. Kim J, Lee W, Hong Y, Ryu W, Lee KJ, Lee W, et al. Psychometric properties of a short self-reported measure of medication adherence among patients with hypertension treated in a busy clinical setting in Korea. *J Epidemiol* 2014;24(2):132-140 [FREE Full text] [doi: [10.2188/jea.je20130064](https://doi.org/10.2188/jea.je20130064)] [Medline: [24463958](https://pubmed.ncbi.nlm.nih.gov/24463958/)]
194. Korb-Savoldelli V, Gillaizeau F, Pouchot J, Lenain E, Postel-Vinay N, Plouin PF, et al. Validation of a French version of the 8-item Morisky medication adherence scale in hypertensive adults. *J Clin Hypertens (Greenwich)* 2012 Jul;14(7):429-434 [FREE Full text] [doi: [10.1111/j.1751-7176.2012.00634.x](https://doi.org/10.1111/j.1751-7176.2012.00634.x)] [Medline: [22747615](https://pubmed.ncbi.nlm.nih.gov/22747615/)]
195. Lee W, Ahn J, Kim J, Hong Y, Hong SK, Kim YT, et al. Reliability and validity of a self-reported measure of medication adherence in patients with type 2 diabetes mellitus in Korea. *J Int Med Res* 2013 Aug;41(4):1098-1110 [FREE Full text] [doi: [10.1177/0300060513484433](https://doi.org/10.1177/0300060513484433)] [Medline: [23860015](https://pubmed.ncbi.nlm.nih.gov/23860015/)]

196. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. *J Clin Hypertens (Greenwich)* 2008 May;10(5):348-354 [[FREE Full text](#)] [doi: [10.1111/j.1751-7176.2008.07572.x](https://doi.org/10.1111/j.1751-7176.2008.07572.x)] [Medline: [18453793](#)]
197. Nakhaeizadeh M, Khalooei A. Psychometric properties of Persian version of the 8-item Morisky medication adherence scale in type 2 diabetes patients. *J Clin Diagn Res* 2019. [doi: [10.7860/jcdr/2019/41408.13076](https://doi.org/10.7860/jcdr/2019/41408.13076)]
198. Sakthong P, Chabunthom R, Charoenvisuthiwongs R. Psychometric properties of the Thai version of the 8-item Morisky medication adherence scale in patients with type 2 diabetes. *Ann Pharmacother* 2009 May;43(5):950-957. [doi: [10.1345/aph.1L453](https://doi.org/10.1345/aph.1L453)] [Medline: [19366872](#)]
199. Shilbayeh SA, Almutairi WA, Alyahya SA, Alshammari NH, Shaheen E, Adam A. Validation of knowledge and adherence assessment tools among patients on warfarin therapy in a Saudi hospital anticoagulant clinic. *Int J Clin Pharm* 2018 Feb;40(1):56-66. [doi: [10.1007/s11096-017-0569-5](https://doi.org/10.1007/s11096-017-0569-5)] [Medline: [29189976](#)]
200. Tan C, Teng G, Chong K, Cheung P, Lim A, Wee H, et al. Utility of the Morisky medication adherence scale in gout: a prospective study. *Patient Prefer Adherence* 2016;10:2449-2457 [[FREE Full text](#)] [doi: [10.2147/PPA.S119719](https://doi.org/10.2147/PPA.S119719)] [Medline: [27980395](#)]
201. Wang Y, Kong M, Ko Y. Psychometric properties of the 8-item Morisky medication adherence scale in patients taking warfarin. *Thromb Haemost* 2012 Oct;108(4):789-795. [doi: [10.1160/TH12-05-0368](https://doi.org/10.1160/TH12-05-0368)] [Medline: [22836920](#)]
202. Zongo A, Guénette L, Moisan J, Guillaumie L, Lauzier S, Grégoire JP. Revisiting the internal consistency and factorial validity of the 8-item Morisky medication adherence scale. *SAGE Open Med* 2016;4:2050312116674850 [[FREE Full text](#)] [doi: [10.1177/2050312116674850](https://doi.org/10.1177/2050312116674850)] [Medline: [27895914](#)]
203. Osborn CY, Gonzalez JS. Measuring insulin adherence among adults with type 2 diabetes. *J Behav Med* 2016 Aug 9;39(4):633-641 [[FREE Full text](#)] [doi: [10.1007/s10865-016-9741-y](https://doi.org/10.1007/s10865-016-9741-y)] [Medline: [27062271](#)]
204. Arnet I, Metaxas C, Walter PN, Morisky DE, Hersberger KE. The 8-item Morisky medication adherence scale translated in German and validated against objective and subjective polypharmacy adherence measures in cardiovascular patients. *J Eval Clin Pract* 2015 Apr;21(2):271-277. [doi: [10.1111/jep.12303](https://doi.org/10.1111/jep.12303)] [Medline: [25558796](#)]
205. Shin D, Kim C. Psychometric evaluation of a Korean version of the 8-item medication adherence scale in rural older adults with hypertension. *Aust J Rural Health* 2013 Dec;21(6):336-342. [doi: [10.1111/ajr.12070](https://doi.org/10.1111/ajr.12070)] [Medline: [24299439](#)]
206. Okello S, Nasasira B, Muir AN, Muyingo A. Validity and reliability of a self-reported measure of antihypertensive medication adherence in Uganda. *PLoS One* 2016;11(7):e0158499 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0158499](https://doi.org/10.1371/journal.pone.0158499)] [Medline: [27367542](#)]
207. Reynolds K, Viswanathan HN, Muntner P, Harrison TN, Cheetham TC, Hsu JY, et al. Validation of the osteoporosis-specific Morisky medication adherence scale in long-term users of bisphosphonates. *Qual Life Res* 2014 Sep;23(7):2109-2120. [doi: [10.1007/s11136-014-0662-3](https://doi.org/10.1007/s11136-014-0662-3)] [Medline: [24604077](#)]
208. Reynolds K, Viswanathan HN, O'Malley CD, Muntner P, Harrison TN, Cheetham TC, et al. Psychometric properties of the osteoporosis-specific Morisky medication adherence scale in postmenopausal women with osteoporosis newly treated with bisphosphonates. *Ann Pharmacother* 2012 May;46(5):659-670. [doi: [10.1345/aph.1Q652](https://doi.org/10.1345/aph.1Q652)] [Medline: [22510666](#)]
209. Wickersham KE, Sereika SM, Kang H, Tamres LK, Erlen JA. Use of a self-report medication adherence scale for measuring adherence to antiretroviral therapy in patients with HIV/AIDS. *J Nurs Meas* 2018 Aug;26(2):E72-E88. [doi: [10.1891/1061-3749.26.2.E72](https://doi.org/10.1891/1061-3749.26.2.E72)] [Medline: [30567952](#)]
210. Athavale AS, Bentley JP, Banahan BF, McCaffrey DJ, Pace PF. Preliminary development of the medication nonpersistence scale. *J Am Pharm Assoc (2003)* 2017;57(3):389-94.e1. [doi: [10.1016/j.japh.2017.01.014](https://doi.org/10.1016/j.japh.2017.01.014)] [Medline: [28279599](#)]
211. Resnick B, Wehren L, Orwig D. Reliability and validity of the self-efficacy and outcome expectations for osteoporosis medication adherence scales. *Orthop Nurs* 2003;22(2):139-147. [doi: [10.1097/00006416-200303000-00012](https://doi.org/10.1097/00006416-200303000-00012)] [Medline: [12703398](#)]
212. Qi B, Resnick B. Reliability and validity of the Chinese versions of self-efficacy and outcome expectations for osteoporosis medication adherence scales in Chinese immigrants. *J Nurs Meas* 2014;22(3):472-488. [doi: [10.1891/1061-3749.22.3.472](https://doi.org/10.1891/1061-3749.22.3.472)] [Medline: [25608433](#)]
213. Almeida-Brasil CC, Nascimento ED, Silveira MR, Bonolo PD, Ceccato MD. New patient-reported outcome measure to assess perceived barriers to antiretroviral therapy adherence: the PEDIA scale. *Cad Saude Publica* 2019;35(5):e00184218 [[FREE Full text](#)] [doi: [10.1590/0102-311x00184218](https://doi.org/10.1590/0102-311x00184218)] [Medline: [31166419](#)]
214. Kalichman SC, Cain D, Fuhrel A, Eaton L, di Fonzo K, Ertl T. Assessing medication adherence self-efficacy among low-literacy patients: development of a pictographic visual analogue scale. *Health Educ Res* 2005 Feb;20(1):24-35. [doi: [10.1093/her/cyg106](https://doi.org/10.1093/her/cyg106)] [Medline: [15253999](#)]
215. Zschocke I, Mrowietz U, Lotzin A, Karakasili E, Reich K. Assessing adherence factors in patients under topical treatment: development of the topical therapy adherence questionnaire (TTAQ). *Arch Dermatol Res* 2014 Apr;306(3):287-297 [[FREE Full text](#)] [doi: [10.1007/s00403-014-1446-x](https://doi.org/10.1007/s00403-014-1446-x)] [Medline: [24509981](#)]
216. Arnsten J, Demas P, Farzadegan H, Grant R, Gourevitch M, Chang C, et al. Antiretroviral therapy adherence and viral suppression in HIV-infected drug users: comparison of self-report and electronic monitoring. *Clin Infect Dis* 2001 Oct 15;33(8):1417-1423 [[FREE Full text](#)] [doi: [10.1086/323201](https://doi.org/10.1086/323201)] [Medline: [11550118](#)]



217. Teixeira A, Oliveira C, Teixeira M, Gaio AR, Lobo JM, de Almeida IF, et al. Development and validation of a novel questionnaire for adherence with topical treatments in psoriasis (QATOP). *Am J Clin Dermatol* 2017 Aug;18(4):571-581. [doi: [10.1007/s40257-017-0272-2](https://doi.org/10.1007/s40257-017-0272-2)] [Medline: [28321796](https://pubmed.ncbi.nlm.nih.gov/28321796/)]
218. Walewski KM, Cicutto L, D'Urzo AD, Heslegrave RJ, Chapman KR. Evaluation of a questionnaire to assess compliance with anti-asthma medications. *J Asthma* 2004 Feb;41(1):77-83. [doi: [10.1081/jas-120026064](https://doi.org/10.1081/jas-120026064)] [Medline: [15046381](https://pubmed.ncbi.nlm.nih.gov/15046381/)]
219. Lewin AB, LaGreca AM, Geffken GR, Williams LB, Duke DC, Storch EA, et al. Validity and reliability of an adolescent and parent rating scale of type 1 diabetes adherence behaviors: the self-care inventory (SCI). *J Pediatr Psychol* 2009 Oct;34(9):999-1007 [FREE Full text] [doi: [10.1093/jpepsy/jsp032](https://doi.org/10.1093/jpepsy/jsp032)] [Medline: [19423660](https://pubmed.ncbi.nlm.nih.gov/19423660/)]
220. Mumtaz T, Haider SA, Malik JA, La Greca AM. Translation, validation and effectiveness of self-care inventory in assessing adherence to diabetes treatment. *J Pak Med Assoc* 2016 Jul;66(7):853-858 [FREE Full text] [Medline: [27427135](https://pubmed.ncbi.nlm.nih.gov/27427135/)]
221. Jansà M, Vidal, Gimenez, Conget, Gallindo M, Roca, Colungo, Esmatjes, Salamero. Psychometric analysis of the Spanish and Catalan versions of the Diabetes Self-Care Inventory-Revised version questionnaire. *PPA* 2013 Oct:997. [doi: [10.2147/ppa.s50271](https://doi.org/10.2147/ppa.s50271)]
222. Alhomoud F, Alhomoud F, Millar I. How effectively are your patients taking their medicines? A critical review of the strathclyde compliance risk assessment tool in relation to the 'MMAS' and 'MARS'. *J Eval Clin Pract* 2016 Jun;22(3):411-420. [doi: [10.1111/jep.12501](https://doi.org/10.1111/jep.12501)] [Medline: [26696012](https://pubmed.ncbi.nlm.nih.gov/26696012/)]
223. Lubinga SJ, Millar I, Babigumira JB. Pilot evaluation of the psychometric properties of a self-medication risk assessment tool among elderly patients in a community setting. *BMC Res Notes* 2011 Oct 11;4:398 [FREE Full text] [doi: [10.1186/1756-0500-4-398](https://doi.org/10.1186/1756-0500-4-398)] [Medline: [21989334](https://pubmed.ncbi.nlm.nih.gov/21989334/)]
224. Pedrosa RB, Rodrigues RC, Oliveira HC, Alexandre NM. Construct validity of the Brazilian version of the self-efficacy for appropriate medication adherence scale. *J Nurs Meas* 2016;24(1):E18-E31. [doi: [10.1891/1061-3749.24.1.E18](https://doi.org/10.1891/1061-3749.24.1.E18)] [Medline: [27103240](https://pubmed.ncbi.nlm.nih.gov/27103240/)]
225. Pedrosa RB, Rodrigues RC. Adaptation and evaluation of the measurement properties of the Brazilian version of the self-efficacy for appropriate medication adherence scale. *Rev Lat Am Enfermagem* 2016;24:e2692 [FREE Full text] [doi: [10.1590/1518-8345.0167.2692](https://doi.org/10.1590/1518-8345.0167.2692)] [Medline: [27192417](https://pubmed.ncbi.nlm.nih.gov/27192417/)]
226. Risser J, Jacobson TA, Kripalani S. Development and psychometric evaluation of the self-efficacy for appropriate medication use scale (SEAMS) in low-literacy patients with chronic disease. *J Nurs Meas* 2007;15(3):203-219. [doi: [10.1891/106137407783095757](https://doi.org/10.1891/106137407783095757)] [Medline: [18232619](https://pubmed.ncbi.nlm.nih.gov/18232619/)]
227. Najimi A, Mostafavi F, Sharifirad G, Golshiri P. Development and study of self-efficacy scale in medication adherence among Iranian patients with hypertension. *J Educ Health Promot* 2017;6:83 [FREE Full text] [doi: [10.4103/jehp.jehp\\_64\\_16](https://doi.org/10.4103/jehp.jehp_64_16)] [Medline: [29114551](https://pubmed.ncbi.nlm.nih.gov/29114551/)]
228. Blumberg EJ, Hovell MF, Kelley NJ, Vera AY, Sipan CL, Berg JP. Self-report INH adherence measures were reliable and valid in Latino adolescents with latent tuberculosis infection. *J Clin Epidemiol* 2005 Jun;58(6):645-648. [doi: [10.1016/j.jclinepi.2004.11.022](https://doi.org/10.1016/j.jclinepi.2004.11.022)] [Medline: [15878479](https://pubmed.ncbi.nlm.nih.gov/15878479/)]
229. Muñoz-Moreno JA, Fumaz CR, Ferrer MJ, Tuldrà A, Rovira T, Viladrich C, SERAD Validation Team. Assessing self-reported adherence to HIV therapy by questionnaire: the SERAD (self-reported adherence) study. *AIDS Res Hum Retroviruses* 2007 Oct;23(10):1166-1175. [doi: [10.1089/aid.2006.0120](https://doi.org/10.1089/aid.2006.0120)] [Medline: [17961100](https://pubmed.ncbi.nlm.nih.gov/17961100/)]
230. Rofail D, Abetz L, Viala M, Gait C, Baladi J, Payne K. Satisfaction and adherence in patients with iron overload receiving iron chelation therapy as assessed by a newly developed patient instrument. *Value Health* 2009;12(1):109-117 [FREE Full text] [doi: [10.1111/j.1524-4733.2008.00390.x](https://doi.org/10.1111/j.1524-4733.2008.00390.x)] [Medline: [18637142](https://pubmed.ncbi.nlm.nih.gov/18637142/)]
231. Knobel H, Alonso J, Casado JL, Collazos J, González J, Ruiz I, GEEMA Study Group. Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA study. *AIDS* 2002 Mar 8;16(4):605-613. [doi: [10.1097/00002030-200203080-00012](https://doi.org/10.1097/00002030-200203080-00012)] [Medline: [11873004](https://pubmed.ncbi.nlm.nih.gov/11873004/)]
232. Ortega Suárez FJ, Sánchez Plumed J, Pérez Valentín MA, Pereira Palomo P, Muñoz Cepeda MA, Lorenzo Aguiar D, Grupo de Estudio Vatreñ. Validation on the simplified medication adherence questionnaire (SMAQ) in renal transplant patients on tacrolimus. *Nefrologia* 2011;31(6):690-696 [FREE Full text] [doi: [10.3265/Nefrologia.pre2011.Aug.10973](https://doi.org/10.3265/Nefrologia.pre2011.Aug.10973)] [Medline: [22130285](https://pubmed.ncbi.nlm.nih.gov/22130285/)]
233. Rathbun RC, Farmer KC, Lockhart SM, Stephens JR. Validity of a stage of change instrument in assessing medication adherence in indigent patients with HIV infection. *Ann Pharmacother* 2007 Feb;41(2):208-214. [doi: [10.1345/aph.1H383](https://doi.org/10.1345/aph.1H383)] [Medline: [17213294](https://pubmed.ncbi.nlm.nih.gov/17213294/)]
234. Willey C, Redding C, Stafford J, Garfield F, Geletko S, Flanigan T, et al. Stages of change for adherence with medication regimens for chronic disease: development and validation of a measure. *Clin Ther* 2000 Jul;22(7):858-871. [doi: [10.1016/s0149-2918\(00\)80058-2](https://doi.org/10.1016/s0149-2918(00)80058-2)] [Medline: [10945512](https://pubmed.ncbi.nlm.nih.gov/10945512/)]
235. Julian FS, Martin P, Erickson SR. Validation of the special projects of national significance adherence tool in HIV/AIDS patients. *Ann Pharmacother* 2010 May 4;44(6):1003-1009. [doi: [10.1345/aph.1m690](https://doi.org/10.1345/aph.1m690)]
236. Lu M, Safren SA, Skolnik PR, Rogers WH, Coady W, Hardy H, et al. Optimal recall period and response task for self-reported HIV medication adherence. *AIDS Behav* 2008 Jan;12(1):86-94. [doi: [10.1007/s10461-007-9261-4](https://doi.org/10.1007/s10461-007-9261-4)] [Medline: [17577653](https://pubmed.ncbi.nlm.nih.gov/17577653/)]
237. Gutiérrez-Pereyra F, Plaza V, Fernández-Rodríguez C. Validation of the 'test of the adherence to inhalers' (TAI) for asthma and COPD patients. - 2020:-. [doi: [10.1183/13993003.congress-2015.pa5012](https://doi.org/10.1183/13993003.congress-2015.pa5012)]

238. Mancebo MC, Pinto A, Rasmussen SA, Eisen JL. Development of the treatment adherence survey-patient version (TAS-P) for OCD. *J Anxiety Disord* 2008;22(1):32-43 [FREE Full text] [doi: [10.1016/j.janxdis.2007.01.009](https://doi.org/10.1016/j.janxdis.2007.01.009)] [Medline: [17324553](https://pubmed.ncbi.nlm.nih.gov/17324553/)]
239. He W, Bonner A, Anderson D. Patient reported adherence to hypertension treatment: a revalidation study. *Eur J Cardiovasc Nurs* 2016 Apr;15(2):150-156. [doi: [10.1177/1474515115603902](https://doi.org/10.1177/1474515115603902)] [Medline: [26307159](https://pubmed.ncbi.nlm.nih.gov/26307159/)]
240. Giordano TP, Guzman D, Clark R, Charlebois ED, Bangsberg DR. Measuring adherence to antiretroviral therapy in a diverse population using a visual analogue scale. *HIV Clin Trials* 2004;5(2):74-79. [doi: [10.1310/JFXH-G3X2-EYM6-D6UG](https://doi.org/10.1310/JFXH-G3X2-EYM6-D6UG)] [Medline: [15116282](https://pubmed.ncbi.nlm.nih.gov/15116282/)]
241. Duncan N, Kronenberger W, Roberson C, Shapiro A. VERITAS-PRN: a new measure of adherence to episodic treatment regimens in haemophilia. *Haemophilia* 2010 Jan;16(1):47-53. [doi: [10.1111/j.1365-2516.2009.02094.x](https://doi.org/10.1111/j.1365-2516.2009.02094.x)] [Medline: [19754842](https://pubmed.ncbi.nlm.nih.gov/19754842/)]
242. Duncan N, Kronenberger W, Roberson C, Shapiro A. VERITAS-Pro: a new measure of adherence to prophylactic regimens in haemophilia. *Haemophilia* 2010 Mar;16(2):247-255. [doi: [10.1111/j.1365-2516.2009.02129.x](https://doi.org/10.1111/j.1365-2516.2009.02129.x)] [Medline: [19925631](https://pubmed.ncbi.nlm.nih.gov/19925631/)]
243. Liao YW, Cheow C, Leung KTY, Tan H, Low SF, Cheen HH, et al. A cultural adaptation and validation study of a self-report measure of the extent of and reasons for medication nonadherence among patients with diabetes in Singapore. *Patient Preference Adherence* 2019 Jul;Volume 13:1241-1252. [doi: [10.2147/ppa.s208736](https://doi.org/10.2147/ppa.s208736)]
244. Blalock DV, Zullig LL, Bosworth HB, Taylor SS, Voils CI. Self-reported medication nonadherence predicts cholesterol levels over time. *J Psychosom Res* 2019 Mar;118:49-55. [doi: [10.1016/j.jpsychores.2019.01.010](https://doi.org/10.1016/j.jpsychores.2019.01.010)] [Medline: [30782354](https://pubmed.ncbi.nlm.nih.gov/30782354/)]
245. Vreeman RC, Scanlon ML, Tu W, Slaven JE, McAteer CI, Kerr SJ, et al. Validation of a self-report adherence measurement tool among a multinational cohort of children living with HIV in Kenya, South Africa and Thailand. *J Int AIDS Soc* 2019 May;22(5):e25304 [FREE Full text] [doi: [10.1002/jia2.25304](https://doi.org/10.1002/jia2.25304)] [Medline: [31148372](https://pubmed.ncbi.nlm.nih.gov/31148372/)]
246. Vale FC, Santa-Helena ET, Santos MA, Carvalho WM, Menezes PR, Basso CR, et al. Development and validation of the webAd-Q questionnaire to monitor adherence to HIV therapy. *Rev Saude Publica* 2018;52:62 [FREE Full text] [doi: [10.11606/s1518-8787.2018052000337](https://doi.org/10.11606/s1518-8787.2018052000337)] [Medline: [29846437](https://pubmed.ncbi.nlm.nih.gov/29846437/)]
247. Phillips T, Brittain K, Mellins CA, Zerbe A, Remien RH, Abrams EJ, et al. A self-reported adherence measure to screen for elevated HIV viral load in pregnant and postpartum women on antiretroviral therapy. *AIDS Behav* 2017 Feb;21(2):450-461 [FREE Full text] [doi: [10.1007/s10461-016-1448-0](https://doi.org/10.1007/s10461-016-1448-0)] [Medline: [27278548](https://pubmed.ncbi.nlm.nih.gov/27278548/)]
248. Wilson IB, Lee Y, Michaud J, Fowler FJ, Rogers WH. Validation of a new three-item self-report measure for medication adherence. *AIDS Behav* 2016 Nov;20(11):2700-2708 [FREE Full text] [doi: [10.1007/s10461-016-1406-x](https://doi.org/10.1007/s10461-016-1406-x)] [Medline: [27098408](https://pubmed.ncbi.nlm.nih.gov/27098408/)]
249. Fredericksen R, Yang F, Gibbons L, Edwards T, Brown S, Fitzsimmons E, et al. Development and content validation of measures assessing adherence barriers and behaviors for use in clinical care. *Res Social Adm Pharm* 2019 Sep;15(9):1168-1176. [doi: [10.1016/j.sapharm.2018.10.001](https://doi.org/10.1016/j.sapharm.2018.10.001)] [Medline: [30327183](https://pubmed.ncbi.nlm.nih.gov/30327183/)]
250. Linn AJ, Vervloet M, van Dijk L, Smit EG, van Weert JC. Effects of eHealth interventions on medication adherence: a systematic review of the literature. *J Med Internet Res* 2011 Dec 5;13(4):e103 [FREE Full text] [doi: [10.2196/jmir.1738](https://doi.org/10.2196/jmir.1738)] [Medline: [22138112](https://pubmed.ncbi.nlm.nih.gov/22138112/)]
251. Muntner P, Joyce C, Holt E, He J, Morisky D, Webber LS, et al. Defining the minimal detectable change in scores on the eight-item Morisky medication adherence scale. *Ann Pharmacother* 2011 May;45(5):569-575. [doi: [10.1345/aph.1P677](https://doi.org/10.1345/aph.1P677)] [Medline: [21521862](https://pubmed.ncbi.nlm.nih.gov/21521862/)]
252. Stoll CR, Izadi S, Fowler S, Green P, Suls J, Colditz GA. The value of a second reviewer for study selection in systematic reviews. *Res Synth Methods* 2019 Dec;10(4):539-545. [doi: [10.1002/jrsm.1369](https://doi.org/10.1002/jrsm.1369)] [Medline: [31272125](https://pubmed.ncbi.nlm.nih.gov/31272125/)]

## Abbreviations

**AACTG:** Adult AIDS Clinical Trials Group  
**ASK-20:** Adherence Starts with Knowledge 20  
**COSMIN:** Consensus-based Standards for the selection of health Measurement INstruments  
**CQR:** Compliance Questionnaire-Rheumatology  
**GMAS:** General Medication Adherence Scale  
**ITBS:** Immunosuppressant Therapy Barrier Scale  
**MAR-Scale:** Medication Adherence Reasons Scale  
**MARS:** Medication Adherence Rating Scale  
**MARS-5:** five-item Medication Adherence Rating Scale  
**MARS-9:** nine-item Medication Adherence Rating Scale  
**MMAS-4:** four-item Morisky Medication Adherence Scale  
**MMAS-8:** eight-item Morisky Medication Adherence Scale  
**PROM:** patient-reported outcome measure  
**PULSES:** Population-based, Unified, Learning System for Enhances and Sustainable Health  
**SEAMS:** self-efficacy for appropriate medication adherence scale  
**SICT:** satisfaction with iron chelation therapy  
**TAI:** Test of Adherence to Inhalers

*Edited by CL Parra-Calderón; submitted 11.04.20; peer-reviewed by P Kardas, E Román-Villarán, S Wee; comments to author 14.06.20; revised version received 05.07.20; accepted 08.08.20; published 09.10.20.*

*Please cite as:*

Kwan YH, Weng SD, Loh DHF, Phang JK, Oo LJY, Blalock DV, Chew EH, Yap KZ, Tan CYK, Yoon S, Fong W, Østbye T, Low LL, Bosworth HB, Thumboo J

*Measurement Properties of Existing Patient-Reported Outcome Measures on Medication Adherence: Systematic Review*

*J Med Internet Res* 2020;22(10):e19179

URL: <https://www.jmir.org/2020/10/e19179>

doi: [10.2196/19179](https://doi.org/10.2196/19179)

PMID: [33034566](https://pubmed.ncbi.nlm.nih.gov/33034566/)

©Yu Heng Kwan, Si Dun Weng, Dionne Hui Fang Loh, Jie Kie Phang, Livia Jia Yi Oo, Dan V Blalock, Eng Hui Chew, Kai Zhen Yap, Corrinne Yong Koon Tan, Sungwon Yoon, Warren Fong, Truls Østbye, Lian Leng Low, Hayden Barry Bosworth, Julian Thumboo. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 09.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Web-Based Patient-Reported Outcome Measures for Personalized Treatment and Care (PROMPT-Care): Multicenter Pragmatic Nonrandomized Trial

Afaf Girgis<sup>1,2\*</sup>, BSc, PhD; Ivana Durcinoska<sup>1,2\*</sup>, BBiotech, MPH; Anthony Arnold<sup>1,3</sup>, BAppSci, MRS RT; Joseph Descallar<sup>1,2</sup>, BSc, MBIstats; Nasreen Kaadan<sup>1,4</sup>, BAppSci; Eng-Siew Koh<sup>1,4</sup>, MBBS; Andrew Miller<sup>1,3,5</sup>, BMed, MInfCommTech (Res); Weng Ng<sup>1,4,6</sup>, MBBS, PhD; Martin Carolan<sup>1,3</sup>, BSc (Hons), PhD; Stephen A Della-Fiorentina<sup>1,6,7</sup>, MBBS; Sandra Avery<sup>1,4</sup>, Grad Cert; Geoff P Delaney<sup>1,2,4</sup>, MD, PhD

<sup>1</sup>Centre for Oncology Education and Research Translation, Ingham Institute for Applied Medical Research, Sydney, Australia

<sup>2</sup>Faculty of Medicine, The University of New South Wales, Sydney, Australia

<sup>3</sup>Illawarra Cancer Care Centre, Wollongong Hospital, Wollongong, Australia

<sup>4</sup>Liverpool Cancer Therapy Centre, Liverpool Hospital, Sydney, Australia

<sup>5</sup>Centre for Oncology Informatics, University of Wollongong, Wollongong, Australia

<sup>6</sup>School of Medicine, Western Sydney University, Sydney, Australia

<sup>7</sup>Macarthur Cancer Therapy Centre, Campbelltown Hospital, Sydney, Australia

\* these authors contributed equally

**Corresponding Author:**

Afaf Girgis, BSc, PhD

Centre for Oncology Education and Research Translation

Ingham Institute for Applied Medical Research

1 Campbell St, Liverpool

Sydney, 2170

Australia

Phone: 61 0412142841

Email: [afaf.girgis@unsw.edu.au](mailto:afaf.girgis@unsw.edu.au)

## Abstract

**Background:** Despite the acceptability and efficacy of e-patient-reported outcome (ePRO) systems, implementation in routine clinical care remains challenging.

**Objective:** This pragmatic trial implemented the PROMPT-Care (Patient Reported Outcome Measures for Personalized Treatment and Care) web-based system into existing clinical workflows and evaluated its effectiveness among a diverse population of patients with cancer.

**Methods:** Adult patients with solid tumors receiving active treatment or follow-up care in four cancer centers were enrolled. The PROMPT-Care intervention supported patient management through (1) monthly off-site electronic PRO physical symptom and psychosocial well-being assessments, (2) automated electronic clinical alerts notifying the care team of unresolved clinical issues following two consecutive assessments, and (3) tailored online patient self-management resources. Propensity score matching was used to match controls with intervention patients in a 4:1 ratio for patient age, sex, and treatment status. The primary outcome was a reduction in emergency department presentations. Secondary outcomes were time spent on chemotherapy and the number of allied health service referrals.

**Results:** From April 2016 to October 2018, 328 patients from four public hospitals received the intervention. Matched controls (n=1312) comprised the general population of patients with cancer, seen at the participating hospitals during the study period. Emergency department visits were significantly reduced by 33% ( $P=.02$ ) among patients receiving the intervention compared with patients in the matched controls. No significant associations were found in allied health referrals or time to end of chemotherapy. At baseline, the most common patient reported outcomes (above-threshold) were fatigue (39%), tiredness (38.4%), worry (32.9%), general wellbeing (32.9%), and sleep (24.1%), aligning with the most frequently accessed self-management domain pages of physical well-being (36%) and emotional well-being (23%). The majority of clinical feedback reports were reviewed by nursing staff (729/893, 82%), largely in response to the automated clinical alerts (n=877).

**Conclusions:** Algorithm-supported web-based systems utilizing patient reported outcomes in clinical practice reduced emergency department presentations among a diverse population of patients with cancer. This study also highlighted the importance of (1) automated triggers for reviewing above-threshold results in patient reports, rather than passive manual review of patient records; (2) the instrumental role nurses play in managing alerts; and (3) providing patients with resources to support guided self-management, where appropriate. Together, these factors will inform the integration of web-based PRO systems into future models of routine cancer care.

**Trial Registration:** Australian New Zealand Clinical Trials Registry ACTRN12616000615482; <https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=370633>

**International Registered Report Identifier (IRRID):** RR2-10.1186/s12885-018-4729-3

(*J Med Internet Res* 2020;22(10):e19685) doi:[10.2196/19685](https://doi.org/10.2196/19685)

## KEYWORDS

patient-reported outcomes (PROs); eHealth; patient-centered care; electronic health record; nonrandomized controlled trial; emergency department presentations; pragmatic trial; symptom screening

## Introduction

Organizations delivering health care services increasingly incorporate patient-reported outcomes (PRO) to inform person-centered care and evaluate services. Well-implemented ePRO systems are associated with improved patient-provider communication, patient satisfaction [1], health-related quality of life [2,3], compliance with chemotherapy [3], earlier detection of relapse in patients with lung cancer [4], reduced emergency department (ED) presentations [5,6], and improved cancer survival [6,7]. However, implementation and effectiveness evaluations in real-world clinical practice settings are not well studied.

A systematic review of 6 reviews identified facilitators and barriers for the implementation of ePROs in health services, with 2 early stages in the implementation process being critical for organizational time and resource investment [8]. First, designing the processes for using PROs within an organization, with a focus on decisions about data use for clinical purposes, rather than just which PROs to collect and how to collect them. Second, preparing organizations and staff for using PROs, including highlighting their validity and value, training clinicians using them, and developing electronic systems that fit into the centers' patient workflow.

Our team developed PROMPT-Care (Patient Reported Outcome Measures for Personalized Treatment and Care), an ePRO system for routinely collecting PROs remotely from home for patients with cancer. PROMPT-Care provides real-time feedback of results to the cancer care teams to inform patient-centered care and deliver evidence-based self-management information to address patient-reported problems [9]. PROMPT-Care is fully integrated into the electronic oncology information system (OIS), which is acceptable to patients and oncology staff and feasible to implement clinically [10].

This multicenter, pragmatic [11] nonrandomized intervention study aimed to successfully implement PROMPT-Care Version 2.0 [12] into existing clinical workflows, evaluate its effectiveness in a diverse population of patients with cancer, and explore the system utility from patient and health care professional perspectives. We hypothesized that PROMPT-Care intervention patients would have significantly fewer ED

presentations during the study period compared to a usual care control group. The intervention was also expected to impact the time spent on chemotherapy and the number of health service referrals.

## Methods

### Study Design and Participants

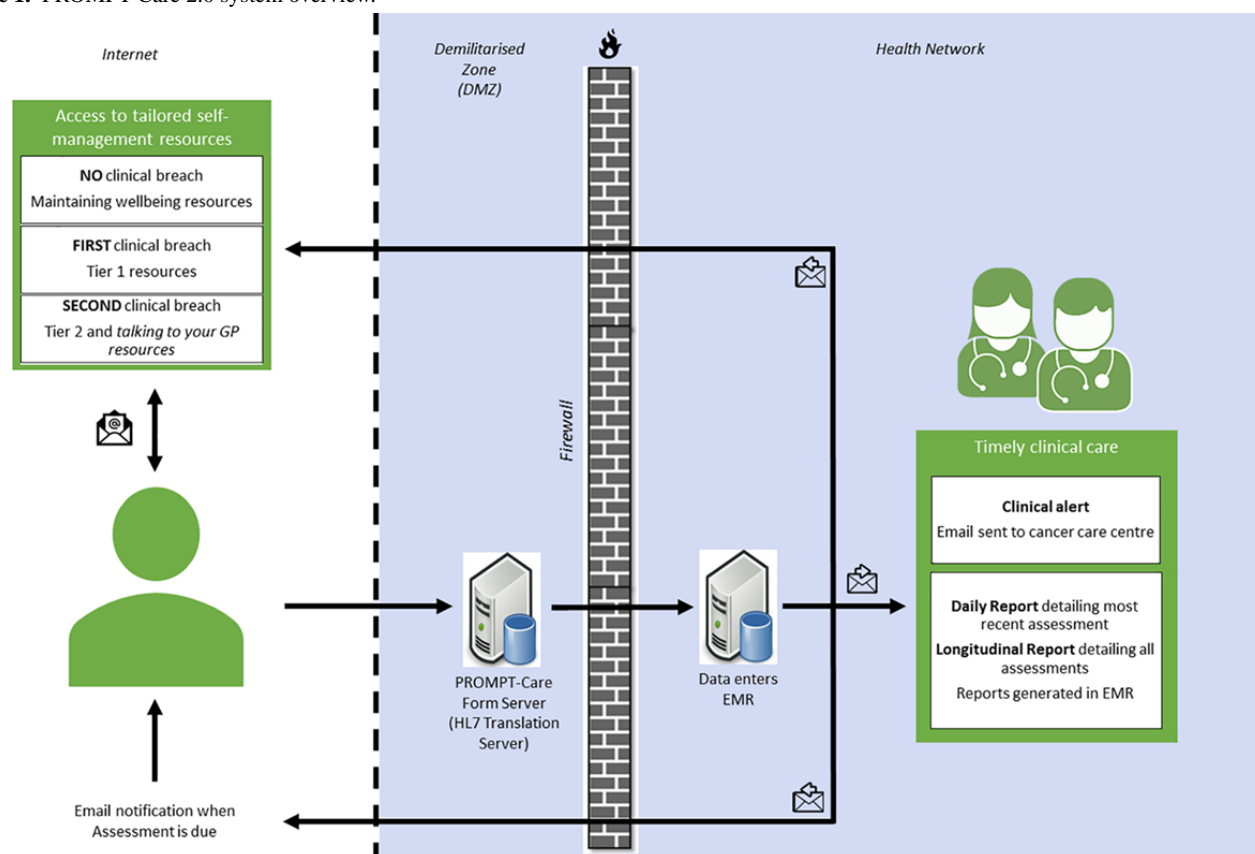
The trial protocol details the study design [12]. Briefly, we conducted a pragmatic, nonrandomized trial among patients with cancer throughout all stages of their cancer care trajectory, with different tumor types and receiving active treatment or follow-up care between April 2016 and October 2018 at four public hospitals in New South Wales, Australia. This study design [11] was chosen to inform external validity and determine this care model's suitability for broad cancer populations, including those in follow-up. Study participants were compared to the general population of patients with cancer receiving usual care at participating hospitals during the study period.

Eligible patients were adults with a confirmed diagnosis of a solid tumor and cognitively able to provide informed consent and complete the assessments in English. Patients without internet access outside the hospital were excluded. Patients with upcoming clinical appointments (treatment or follow-up) at participating hospitals were prescreened for eligibility by treating clinicians using lists extracted from the OIS every month and invited to participate by the nursing or research staff. Patients received written and verbal information about the study, provided their consent, and participated for a minimum of 6 months. Participants received monthly emails prompting them to complete the upcoming online assessment, with one reminder email sent a week later. The Human Research Ethics Committee of South Western Sydney and Illawarra Shoalhaven Local Health Districts (Reference No. HREC/15/LPOOL/287) granted ethics approval, and the trial was registered in the Australian New Zealand Clinical Trials Registry (ACTRN12616000615482).

### Intervention

Four key intervention components were standardized across participating hospitals: ePROs, clinical feedback reports, clinical alerts, and patient self-management (Figure 1).



**Figure 1.** PROMPT-Care 2.0 system overview.

### ePROs

Approximately once a month, patients were prompted via email to complete an online assessment of their physical and psychosocial well-being from home, which included the Distress Thermometer and checklist [13], the Edmonton Symptom Assessment Scale (ESAS) [14], and the Supportive Care Needs Survey-Screening Tool 9 [15]. Patients used an electronic device (eg, tablet, computer), with real-time electronic data transfer (using PROsaiq, DidymoDesigns) into the point-of-care OIS (MOSAIQ, Elekta Medical Systems), via automatic conversion into an Health Level Seven (HL7) message [16].

Patient privacy was maintained by ensuring the email did not contain identifiable patient information, in the event the email was intercepted or sent to the wrong email address. PROsaiq [16], which has restricted access, was located on a PROMPT-Care form server within the demilitarized zone of the health network. To complete an assessment, the patient opened the URL provided in the email, and the browser established a secure session to the PROMPT-Care system, where an SSL Certificate was installed. The patient was required to enter a unique medical record number and their surname to

access the system. The two identifiers were chosen to match the survey results to the correct patient in the OIS. Following the assessment completion, data were translated into a HL7 message by the OIS HL7 translation server, which also sat behind the health network firewall. If the medical record number and surname attached to the HL7 message matched to a patient in the OIS, the survey results were loaded into the OIS under the patient record. If a mismatch occurred, the data were not loaded and a failure was recorded in a log file. No patient information was stored on the PROMPT-Care form server.

### Clinical Feedback Reports

Any care team member could review the ePRO reports in real time with patients, or as required. The ePRO reports included (1) a 1-page summary of the most recent assessment results, including recommended clinical actions and referrals generated from algorithms to facilitate standardized care (Figure 2) [17,18], and (2) a longitudinal report of all assessments to date. Staff were oriented on clinical feedback report use and access at the start of the study, with a periodic refresher training and emails with brief instructions on accessing and using reports in practice throughout the trial.

Figure 2. Sample clinical feedback report.

**Liverpool Cancer Therapy Centre**

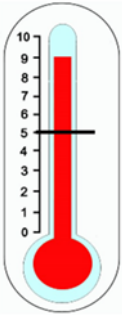
**PROMPT-Care**  
Patient Reported Outcome Measures for Personalised Treatment and Care

Patient Name:	MRN:	DOB	Date of Assessment:
---------------	------	-----	---------------------

THIS PATIENT HAS A HIGH DISTRESS SCORE. Review remainder of report for potential causes of this distress and action the specific recommendations as detailed

<b>Emotional</b>	<b>Recommendation</b>
Wellbeing	Consider reasons for concern and, if required, refer to social work OR psychology OR clinical/health services for further assessment and care
<b>Physical</b>	<b>Recommendation</b>
Breathing, Diarrhoea, Fatigue, Mouth sores, Nausea, Pain, Tiredness, Drowsy, Appetite, Shortness of breath, Lack of energy/tiredness, Getting around	Clinically address as appropriate OR refer to relevant medical specialist OR allied health professional for further assessment and care
Bathing/dressing	Consider reasons for concern and, if required, refer to a physiotherapist AND/OR occupational therapy for further assessment and care
<b>Practical</b>	<b>Recommendation</b>
Being informed about things you can do to help yourself get well	Address educational/informational needs and identify appropriate sources of support and information

**1. Distress (Distress Thermometer & Problem Checklist)**



<b>Practical problems:</b>	<b>Emotional problems:</b>	<b>Physical problems:</b>
Childcare	Depression	Appearance
Housing	Fears	Bathing/dressing
Insurance/financial	Nervousness	Breathing
Transportation	Sadness	Changes in urination
Work/School	Worry	Constipation
Treatment decisions	Loss of interest in usual activities	Diarrhoea
Other:	Coping with illness	Eating
Spiritual/Religious	Losing Hope	Fatigue
<b>Family problems:</b>		Feeling swollen
Dealing with partner	Dealing with Children	Fevers
Ability to have children	Lack of support from family & friends	Getting around
Family health issues		

**2. Symptoms (ESAS)**

**Needs Follow-up**

Pain	3
Tiredness	9
Nausea	9
Depressed	0
Anxious	0
Drowsy	7
Poor Appetite	9
Poor wellbeing	8
Shortness of breath	5

KEY  
This issue is problematic for the patient

**3. Unmet Needs (SCNS-ST9)**

<b>Need</b>	<b>Issue</b>
Psychological	Fears about the cancer spreading Uncertainty about the future
Health system and information	Being informed about your test results as soon as feasible Being informed about things you can do to help yourself to get well
Physical and daily living	Lack of energy/tiredness Not being able to do the things you used to do
Patient care and support	Reassurance by medical staff that the way you feel is normal Hospital staff acknowledging and show sensitivity to your feelings and emotional needs
Sexuality	Changes in sexual relationships

Print date and time: 25/10/2019 at 1:08:33PM Page 1 of 1 Version 3.0

### Clinical Alerts

An email alert was automatically generated whenever any individual ePRO item score breached a predefined threshold on two consecutive assessments. This email was generated by a SMTP server within the health network, where only approved health email addresses were able to receive this email alert ensuring that patient identifying data were unable to leave the organization. A designated member of the care team reviewed

emails at least once a day (Monday-Friday) and followed their cancer center's standard clinical care pathway.

### Patient Self-Management

Approved resources were hosted on a website to support patient self-management, [12,17], with domain-specific webpages—practical problems and emotional, physical, social and family well-being, and a “maintaining well-being” page to support general health. Immediately following ePRO

completion, patients received an email with links to pages related to the above-threshold ePROs (eg, a patient with a breached pain score received a link to the physical well-being page). Information resources, interactive resources (eg, videos, podcasts, self-help programs), and resources to facilitate effective communication with the general practitioner were provided. Patients with no breached ePROs received the “maintaining well-being” link. Resources were maintained on a public website, external to the health firewalls, and no identifiable information was included in the email.

### Selection of Controls

The “usual care” control group included all other patients receiving active treatment or follow-up care at participating centers during the study period. For comparability between the intervention and control samples, clinicians prescreened control patients as part of trial recruitment procedures; patients who did not meet the study eligibility criteria were excluded from the usual care group. Usual care consisted of standard clinical oncology practice, monitoring and managing patient issues, and management and OIS documentation of patient symptoms and issues as needed during clinical appointments.

### Outcomes

The primary outcome was ED presentations. Since patients could present at any hospital in their local health district, ED data was extracted from all hospitals with ED departments ( $n=8$ ) in the two local health districts, during the trial period. A presentation to ED may or may not have required an admission. Secondary outcomes were (1) total time receiving chemotherapy during the study period and (2) referral to in-hospital allied health services (eg, psychology, social work, nutrition and dietetics, occupational therapy, physiotherapy). Primary and secondary outcome data were extracted from patients’ OISs.

Receipt of clinical alerts was automatically logged via the PROMPT-Care system email monitoring and OIS records of clinician notes. Responses to email alerts were recorded in the OIS and research staff recorded clinical actions (eg, referrals, information provision, issue already being managed). The number of clinical feedback reports opened in response to patient assessment completion was tabulated based on medical record logs.

PROMPT-Care compliance was monitored by calculating the proportion of patients completing assessments at expected time points within the first 6 months. Patient views of self-management domain pages, monitored using Google Analytics, were summarized as counts and proportion of resources.

At enrolment, patients completed a demographics survey (sex, marital status, education, employment, and language spoken at home); clinical (cancer site and stage) and treatment (chemotherapy, radiotherapy treatments, active treatment, and follow-up care) details were extracted from the OIS.

Socioeconomic status was determined from the Index of Relative Socio-economic Disadvantage [19], a continuous score split into quintiles (1=most disadvantaged to 5=least disadvantaged).

### Sample Size

Propensity score matching was used to match control patients with intervention patients in a 4:1 ratio with regard to patient age, sex, and treatment status [20]. The study was powered to detect a minimum 14% between-group difference in the primary outcome—ED presentations. Assuming a 4:1 allocation to the control group versus intervention group and based on an assumed 1.4 ED presentations per patient during the study period for the control group [21], a minimum sample of 1760 patients (intervention:  $n=352$ ; control:  $n=1408$ ) was required to achieve 80% power and a two-sided statistical significance of 0.05.

### Statistical Analyses

Demographic and clinical characteristics were described as frequencies, mean scores, and percentages, with between-group comparisons using Chi-square tests or  $t$  tests. Intention-to-treat analyses were conducted in line with a prespecified analysis plan [12]. Multivariable negative binomial regression was used to identify between-group differences in ED presentation rates. To account for between-group differences in distribution of demographic and disease characteristics, we adjusted for disease stage, socioeconomic disadvantage, hospital site, and waiting time (time from diagnosis to start of PROMPT-Care). Allied health referrals were analyzed similarly. Multivariable Cox proportional hazards model was used to analyze length of time from start to end of chemotherapy, adjusting for stage of disease, treatment status, and socioeconomic disadvantage. A delayed entry model was used to specify how many days after the start of chemotherapy a patient started PROMPT-Care.

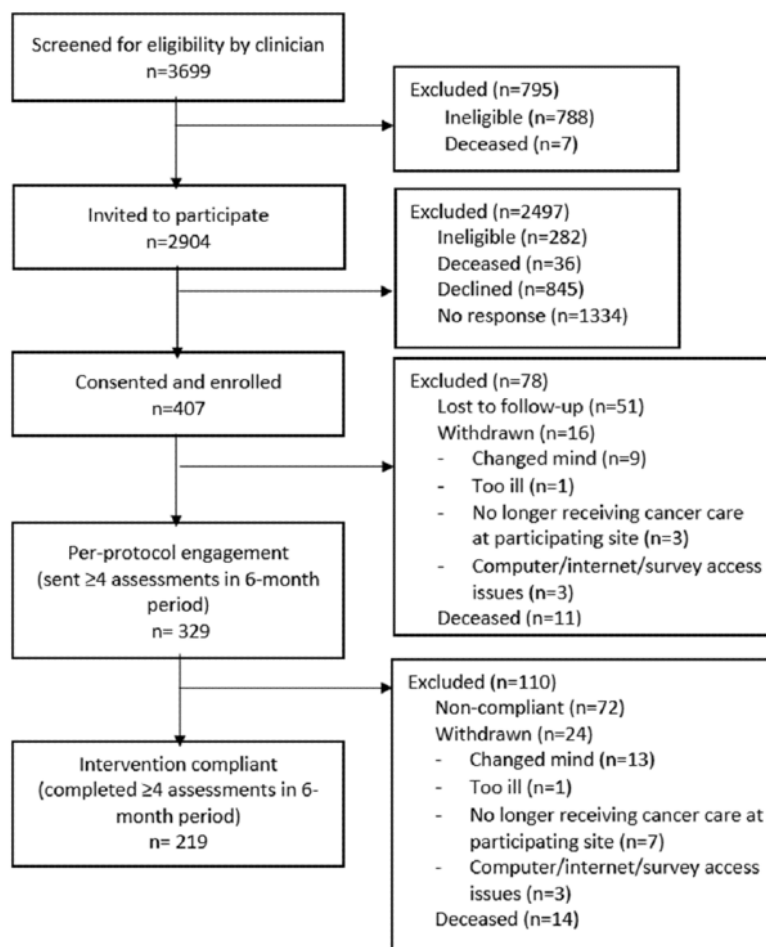
### Data Sharing

De-identified data will be available on request after all primary and secondary endpoints have been analyzed and published, and after signing of an agreement with the PROMPT-Care program. Requests for data sharing can be made to the corresponding author, including a research proposal that must be approved by the chief investigator team.

## Results

### Study Population

Between April 2016 and April 2018, clinicians prescreened 3699 patients against clinical and language exclusion criteria and invited 2904 (79%) to participate (Figure 3). A further 283 patients were ineligible, 36 were deceased, 845 declined, and 1334 did not respond. The remaining 406 patients were enrolled to the intervention group, with 328 (81%) receiving the “per-protocol intervention” (sent  $>4$  assessments within 6 months postenrolment), due to administrative staff error sending assessment emails.

**Figure 3.** CONSORT diagram.

### Propensity Score Matching

Propensity score matching was used to match age, sex, and treatment status. Prior to matching, there were no significant differences in age and sex between the treatment groups. However, there was a higher proportion of control patients on active treatment (1157/1911, 60.5%) compared with patients in the intervention group (139/328, 42.4%) ( $P<.001$ ). After

matching there were no differences in age, sex, and treatment groups (Table 1).

After matching, the study groups were not significantly different at baseline in the site of cancer and waiting time, but the control group included significantly more patients from the most socioeconomically disadvantaged quintile and at clinical stage IV ( $P=.01$ ; Table 1).

**Table 1.** Participant characteristics for intervention, control, and control participants matched for age, sex, and treatment status.

	PROMPT-Care <sup>a</sup> (n=328)	Control (n=1911)	<i>P</i> value	Matched control (n=1312)	<i>P</i> value
Age (years), mean (range)	62.4 (25-86)	62.7 (18-96)	.61	62.3	.90
<b>Sex, n (%)</b>			.33		.45
Male	133 (40.6)	720 (37.7)		502 (38.3)	
Female	195 (59.5)	1191 (62.3)		810 (61.7)	
<b>Site of cancer, n (%)</b>			.50		.18
Breast	132 (40.2)	854 (44.7)		620 (47.3)	
Prostate	51 (15.6)	295 (15.4)		199 (15.2)	
Colorectal	37 (11.3)	186 (9.7)		127 (9.7)	
Respiratory	29 (8.8)	127 (6.7)		83 (6.3)	
Gynaecological	16 (4.9)	83 (4.3)		43 (3.3)	
Upper gastrointestinal	15 (4.6)	78 (4.1)		51 (3.9)	
Skin	11 (3.4)	83 (4.8)		48 (3.7)	
Oral	10 (3.1)	34 (1.8)		21 (1.6)	
Other	27 (8.2)	171 (9)		120 (9.2)	
<b>Stage of disease<sup>b</sup>, n (%)</b>			.20		.01
0/I	66 (22.1)	415 (21.7)		303 (23.1)	
II	90 (27.4)	569 (29.8)		412 (31.4)	
III	57 (17.4)	366 (19.2)		248 (18.9)	
IV	80 (24.4)	357 (18.7)		209 (15.9)	
Missing	35 (10.7)	204 (10.7)		140 (10.7)	
<b>Treatment received, n (%)</b>					
Chemotherapy	105 (32)	561 (29.4)	.33	269 (20.5)	<.001
Radiotherapy	66 (20.1)	897 (46.9)	<.001	430 (32.8)	<.001
<b>Treatment status, n (%)</b>			<.001		>.99
Active treatment <sup>c</sup>	139 (42.46)	1157 (60.5)		558 (42.5)	
Follow-up care	189 (57.6)	754 (39.5)		754 (57.5)	
<b>Socioeconomic status (IRSD)<sup>d</sup>, n (%)</b>			<.001		<.001
1	54 (16.45)	609 (31.9)		403 (30.7)	
2	98 (29.9)	490 (25.6)		340 (25.9)	
3	52 (15.9)	275 (14.4)		207 (15.8)	
4	35 (10.7)	197 (10.3)		124 (9.5)	
5	90 (27.1)	340 (17.8)		238 (18.1)	
<b>Relationship status<sup>b,e</sup>, n (%)</b>			— <sup>f</sup>		—
Single	71 (23.1)	—		—	
Partnered	236 (76.9)	—		—	
<b>Education status<sup>b,e</sup>, n (%)</b>			—		—
High school or less	122 (39.7)	—		—	
Post-secondary education	185 (60.3)	—		—	
<b>Employment<sup>c,e</sup>, n (%)</b>			—		—
Employed	129 (42)	—		—	
Retired	155 (50.5)	—		—	



	PROMPT-Care <sup>a</sup> (n=328)	Control (n=1911)	<i>P</i> value	Matched control (n=1312)	<i>P</i> value
Other	23 (7.5)	—		—	
<b>Hospital site, n (%)</b>			—		<.001
1	146 (44.5)	628 (32.9)		447 (34.1)	
2	58 (17.7)	286 (15)		229 (17.5)	
3	88 (26.8)	867 (45.4)		550 (41.9)	
4	36 (11)	130 (6.8)		86 (6.6)	
Waiting time <sup>g</sup> , mean (range)	726.5 (1-5855)	662.2 (0-7458)	.22	785.1 (0-7458)	.31

<sup>a</sup>PROMPT-Care: Patient Reported Outcome Measures for Personalized Treatment and Care.

<sup>b</sup>Some level of missing data.

<sup>c</sup>Chemotherapy, radiotherapy, or both.

<sup>d</sup>IRSD: Index of relative socioeconomic disadvantage. 1=most disadvantaged; 5=least disadvantaged.

<sup>e</sup>Data extracted from the patient survey and are hence not available for the control group.

<sup>f</sup>Not available.

<sup>g</sup>Diagnosis date to PROMPT-Care start.

### Emergency Department Presentations

There were 314 ED visits from the 328 patients in the intervention group (0.96 ED visits per patient) and 1874 ED visits from the 1312 patients in the matched controls (1.4 ED visits per patient). After accounting for patient time in PROMPT-Care (intervention: 192,859 days; control: 1,006,956 days), the rates of ED visits were 16.2 per 100,000 patient days

in the intervention group and 18.6 per 100,000 patient days in the matched controls.

After adjustment for stage, socioeconomic disadvantage, recruitment site, and waiting time in the multivariable negative binomial regression model with an offset of time in PROMPT-Care (to account for maldistribution of these variables between the groups), ED visits were significantly lower by 33% ( $P=.02$ ) in the intervention group compared with the matched controls (Table 2).

**Table 2.** Comparison of emergency department presentations using negative binomial regression. Patients matched for age, sex, and treatment status.

	Univariate				Multivariable			
	RR <sup>a</sup>	Lower 95% CI	Upper 95% CI	P value	RR <sup>a</sup>	Lower 95% CI	Upper 95% CI	P value
<b>Group</b>								
Intervention	0.81	0.64	1.04	.10	0.75	0.60	0.95	.02
Control	Reference	— <sup>b</sup>	—	—	Reference	—	—	—
<b>Stage</b>				<.001				
0/I	Reference	—	—	—	Reference	—	—	—
II	1.50	1.16	1.96	.002	1.50	1.15	1.95	.002
III	2.93	2.20	3.91	<.001	2.56	1.92	3.43	<.001
IV	5.80	4.35	7.74	<.001	5.56	4.15	7.45	<.001
Missing	3.59	2.57	5.02	<.001	3.64	2.62	5.07	<.001
<b>Socioeconomic status (IRSD)<sup>c</sup></b>				.02				
1	1.81	1.25	2.61	.002	1.90	1.34	2.70	<.001
2	1.82	1.25	2.64	.002	1.38	0.96	1.98	.08
3	1.48	0.99	2.22	.06	1.77	1.21	2.59	.004
4	Reference	—	—	—	Reference	—	—	—
5	1.52	1.04	2.24	.03	1.40	0.96	2.03	.08
<b>Recruitment site</b>				<.001				
1	0.61	0.42	0.89	.01	0.91	0.63	1.32	.63
2	0.40	0.26	0.61	<.001	0.47	0.31	0.71	<.001
3	0.44	0.30	0.64	<.001	0.56	0.38	0.81	.002
4	Reference	—	—	—	Reference	—	—	—
Waiting time	0.9998	0.9997	1.00	<.001	0.9999	0.9998	1.0000	.018

<sup>a</sup>RR: Relative risk.<sup>b</sup>Not available.<sup>c</sup>IRSD: Index of relative socioeconomic disadvantage. 1=most disadvantaged; 5=least disadvantaged.

### Time on Chemotherapy Treatment and Allied Health Referrals

Time on chemotherapy did not differ between the intervention and control groups (*hazard ratio*=0.96; *P*=.71; see [Multimedia](#)

[Appendix 1](#)) after adjustment for stage, socioeconomic disadvantage, and recruitment site.

Allied health referrals were also not significantly different between intervention and control groups (*relative risk*=0.74; *P*=.20) after adjusting for stage, socioeconomic disadvantage, recruitment site, and waiting time ([Table 3](#)).

**Table 3.** Comparison of Allied Health referrals using negative binomial regression. Patients matched for age, sex, and treatment status.

	Univariate				Multivariable			
	RR <sup>a</sup>	Lower 95% CI	Upper 95% CI	P value	RR <sup>a</sup>	Lower 95% CI	Upper 95% CI	P value
<b>Group</b>								
Intervention	0.44	0.27	0.72	<.001	0.74	0.48	1.16	.20
Control	Reference	—	—	—	Reference	—	—	—
<b>Stage</b>				<.001				<.001
0/I	Reference	—	—	—	Reference	—	—	—
II	2.73	1.62	4.61	<.001	2.16	1.30	3.60	.003
III	7.29	4.14	12.86	<.001	6.10	3.52	10.56	<.001
IV	10.61	5.96	18.89	<.001	9.62	5.51	16.79	<.001
Missing	17.00	8.86	32.62	<.001	9.69	5.20	18.03	<.001
<b>Socioeconomic status (IRSD)<sup>b</sup></b>				<.001				.14
1	3.61	1.79	7.29	<.001	2.22	1.15	4.31	.02
2	2.26	1.11	4.58	.02	2.34	1.18	4.62	.01
3	1.12	0.51	2.42	.78	1.59	0.75	3.41	.23
4	Reference	—	—	—	Reference	—	—	—
5	1.85	0.88	3.87	.10	2.12	1.05	4.27	.04
<b>Recruitment site</b>				<.001				<.001
1	1.92	0.89	4.16	.10	2.58	1.23	5.41	.01
2	0.14	0.05	0.36	<.001	0.21	0.08	0.54	.001
3	3.52	1.63	7.57	.001	3.38	1.60	7.16	.001
4	Reference	—	—	—	Reference	—	—	—
Waiting time	0.9992	0.9990	0.9994	<.001	0.9993	0.9991	0.9995	<.001

<sup>a</sup>RR: Relative risk.<sup>b</sup>IRSD: Index of relative socioeconomic disadvantage. 1=most disadvantaged; 5=least disadvantaged.

## System Utility

Patients (n=328) completed 2746 PROMPT-Care assessments. At baseline, the most common PRO (above threshold) were fatigue (128/328, 39%), tiredness (126/328, 38.4%), worry (108/328, 32.9%), general well-being (108/328, 32.9%), and sleep (79/328, 24.1%), aligning with the most frequently accessed self-management domain pages of physical well-being (680/1867, 36%) and emotional well-being (429/1867, 23%). The majority of patients (218/328, 66.4%) used the system as intended, completing four or more assessments within the 6-month intervention period.

Overall, 32% (893/2751) of clinical feedback reports were reviewed, the vast majority (729/893, 82%) by nursing staff and 17% by oncologists (149/893).

In total, 71% (233/328) of intervention patients generated a clinical alert. A total of 877 clinical email alerts were generated, with a mean of 31 (range 2-78) alerts per month during the 30-month study. Overall, 44% (383/877) of clinical alerts were reviewed by designated nurse care coordinators, resulting in 496 actions: in-clinic or telephone patient follow-up (302/496, 61%), no further follow-up deemed necessary (83/496, 17%),

and telephone contact attempts but patient could not be reached (111/496, 22%). Issues were largely resolved through discussion (129/302, 43%) or information provision (98/302, 32%), with some health care professional referrals (75/302, 25%).

## Discussion

### Principal Findings

We investigated PROMPT-Care implementation into routine clinical practice among diverse populations of patients with cancer. For adult patients with cancer, receiving active treatment or in follow-up care, algorithm-supported web-based systems utilizing PROs in routine practice resulted in fewer ED presentations. Prespecified secondary analyses showed no statistically significant associations in allied health referrals or time on chemotherapy and were likely underpowered to detect any change. Another factor to consider is the multimodal nature of the intervention where, in addition to clinical follow-up for above-threshold PROs, patients also received targeted resources enabling patients to self-manage minor issues where clinically appropriate. Further research should explore multimodal interventions such as PROMPT-Care that combine ePRO clinical implementation and appropriate patient self-management, as

patients are responsible for managing their own care between hospital clinic visits.

The finding of reduced ED presentations observed in our study is also consistent with other web-based PRO studies. Basch et al [3,6] demonstrated that web-based symptom reporting with automated email alerts resulted in fewer ED visits. However, this was a single-site study, in a population with advanced disease receiving chemotherapy, with ED visits as a secondary outcome. Barbera et al [5] demonstrated that ESAS screening was associated with decreased ED visits, among patients with breast cancer; and Howell et al [22] showed reduced ED visits following ESAS screening in a prepost comparative cohort study. Additionally, to our knowledge, our study is the first to explore the impact of a multimodal intervention combining electronic PRO screening with clinical intervention and patient self-management.

### Integration and Clinical Use of ePRO Systems Into Routine Practice

In this pragmatic study, PROMPT-Care was implemented using available resources and workflows, with two important findings influencing future adoption. First, the automated clinical alerts prompted the majority of clinical report reviews, highlighting the importance of embedded triggers for reviewing above-threshold reports, which are otherwise passively accessible in the OIS. Second, nursing staff were instrumental in reviewing, triaging, and directly managing responses to clinical alerts, echoing many studies of nurse-led telephone navigator models of cancer care [23]. Additionally, with the relatively low number of clinical alerts generated each month, the automated alerts fit within existing workflows without creating onerous amounts of additional work. These findings will contribute to the evidence-based development and integration of ePRO systems into future models of routine care, to not only reduce the high demand on health services but also provide targeted systematic care to patients most in need.

### Strengths and Limitations

Our study addressed existing evidence gaps. We examined the impact on health service outcomes of an ePRO system implemented in routine practice settings, rather than a controlled research environment. We included a broad population of patients with cancer, enhancing generalizability. We monitored physical symptoms and psychosocial well-being and provided real-time feedback to care providers and patients. Intervention delivery to a broad cross-section of patients, across four centers providing comprehensive chemotherapy and radiotherapy treatment services, further enhances the study's external validity. Hence, our findings are potentially generalizable to other clinical settings in countries with similar health systems, highlighting the importance of informing ePRO system implementation more broadly.

Our study also has some limitations. In our study, the intervention did not reach 100% of patients. First, response rate was low, possibly contributing to recruitment bias in treatment status and socioeconomic disadvantage. Second, patients unable to complete assessments in English or without access to a device and internet outside the hospital were excluded. While we

acknowledge this limitation, our pilot study patients wanted remote electronic access to assessments [10], making this a critical component of our intervention design. Future interventions utilizing online or remote ePRO completion will likely be more accessible, with 86% of Australian households having internet access at home [24]. Third, due to administrative problems, only 81% of patients received the intervention per-protocol (sent  $\geq 4$  assessments in the first 6 months postrecruitment). Finally, there was limited follow-up (one email reminder) if monthly assessments were not completed, in contrast to nurses following up patients in the Basch et al [3] trial to ensure high adherence. Despite this, 67% of patients engaged with the system as intended, suggesting high intervention acceptability.

Clinically integrating PROs is challenging, even in centers with screening implemented for many years. Cancer Care Ontario's systematic ESAS distress screening commenced in 2007, with rates increasing steadily from approximately 20% in 2009 to 59% in 2015, but remaining below the provincial target of 70% screened (range 31% at lowest performing to 91% at highest performing centers). Chow et al [25] also found that patients completed a brief distress screener 75% of the time they received a text message, suggesting feasibility of remote ePRO screening, as per our PROMPT-Care model.

Our assessment and alert frequencies are other limitations. Monthly assessments were selected to accommodate the longest response timeframe for the selected scales (Supportive Care Needs Survey-Screening Tool 9 "in the past month..."), and inclusion of follow-up as well as on-treatment patients. The Clinical Advisory Group decided that clinical alerts should be generated following two consecutive breaches [17] to minimize false-positive alerts, since patients on treatment were in regular contact with the cancer service; hence, any additional concerning issues would be readily identified between assessments. However, increasing assessment frequency may identify more acute symptoms (eg, pain), which likely result in ED presentations. Finding the balance between screening burden and timely alerts is an ongoing challenge.

Retrospective interrogation of systematically collected data shows that with systematic clinical implementation of distress screening, distress levels significantly predicted service utilization and referral rates [26], particularly to social and psycho-oncology services [27]. A key component of our effectiveness evaluation was to observe nurse uptake into workflows as part of routine practice with minimal intervention. This likely resulted in the low observed opening rates of clinical alerts. Further research is needed to explore implementation strategies that would encourage and support clinical staff to embed ePRO review and action into routine workflows.

### Future Research

This study enhances our understanding of how PROs inform cancer care and patient self-management beyond a randomized controlled trial and raises priority research questions. We know very little about reaching underserved patient populations. In particular, the extent to which ePRO systems like PROMPT-Care are acceptable and feasible for (1) assessing PROs in languages other than English and (2) informing

patient-centered care in non-English subgroups is unknown. Furthermore, given the dearth of non-English resources to support patient self-management, applying cultural adaptation principles rather than simply translating existing resources into other languages [28] is a focus for future development.

## Conclusions

Although most previous research has evaluated ePRO systems with patients receiving adjuvant treatment, there is a compelling argument for eHealth systems like PROMPT-Care informing the care of the growing population of cancer survivors. Completing ePROs routinely can efficiently identify follow-up patients managing well, who can be supported with self-management resources rather than attending specialist follow-up appointments. ePROs can also detect issues of concern when patients do not have a scheduled appointment, prompting

timely clinical care if required to avoid escalation in severity of issues. Research into the acceptability and cost-effectiveness of this model of care is required, but our research supports its acceptability and feasibility to patients and oncology staff. Research to date has predominantly focused on testing ePRO intervention efficacy. We have purposefully undertaken a pragmatic trial to better understand the effectiveness of ePRO systems in real-world settings, demonstrating that ePROs are likely to be adopted in routine care when integrated into the patient OIS and existing clinical workflows, allowing easy access by the care team. However, significant barriers exist for many cancer centers to do this. It is imperative that future research explore implementation questions, focusing on evaluating the processes and outcomes of ePRO systems adopted as business-as-usual.

## Acknowledgments

We sincerely thank the following people who were integral to the development and delivery of the PROMPT-Care Program: Orlando Rincones, Kenneth Masters, Thomas T Tran, Martha Gerges, Tiffany Sandell, Kylie Smith, CheeFon Chang, and Janelle Levesque. We thank the patients who generously contributed their time and input to the development of PROMPT-Care. We gratefully acknowledge the significant input of the broader team, which includes care coordinators, clinicians and researchers from the South Western Sydney and Illawarra Shoalhaven Local Health Districts, and the Clinical and Technical Advisory Groups, which include more than 40 members from Local Health Districts across New South Wales as well as from the Cancer Institute New South Wales. AG is funded through Cancer Institute grants. This study was funded by the Cancer Institute New South Wales, Bupa Health Foundation, South Western Sydney Local Health Service, and Wollondilly Health Alliance.

## Authors' Contributions

AG and GD obtained study funding. AG was chief investigator; contributed to trial design, analysis plan, protocol development, data interpretation, and writing. GD and ID contributed to trial design, protocol development, analysis, data interpretation, and writing. AA, NK, ESK, AAM, WN, MGC, SDF, and SA contributed to the design and implementation of the intervention. JD contributed to development of analysis plan, conducted statistical analyses, and writing. All authors reviewed and approved the final manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Supplementary Table S1 - Time to end of chemotherapy.

[DOCX File, 16 KB - [jmir\\_v22i10e19685\\_app1.docx](#)]

## References

1. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res* 2013 Jun 11;13:211 [FREE Full text] [doi: [10.1186/1472-6963-13-211](#)] [Medline: [23758898](#)]
2. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. *J Clin Oncol* 2004 Feb 15;22(4):714-724. [doi: [10.1200/jco.2004.06.078](#)]
3. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. *J Clin Oncol* 2016 Feb 20;34(6):557-565. [doi: [10.1200/jco.2015.63.0830](#)]
4. Denis F, Viger L, Charron A, Voog E, Letellier C. Detecting lung cancer relapse using self-evaluation forms weekly filled at home: the sentinel follow-up. *Support Care Cancer* 2014 Jan;22(1):79-85. [doi: [10.1007/s00520-013-1954-9](#)] [Medline: [23995815](#)]



5. Barbera L, Sutradhar R, Howell D, Sussman J, Seow H, Dudgeon D, et al. Does routine symptom screening with ESAS decrease ED visits in breast cancer patients undergoing adjuvant chemotherapy? *Support Care Cancer* 2015 Oct;23(10):3025-3032. [doi: [10.1007/s00520-015-2671-3](https://doi.org/10.1007/s00520-015-2671-3)] [Medline: [25711657](https://pubmed.ncbi.nlm.nih.gov/25711657/)]
6. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA* 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: [10.1001/jama.2017.7156](https://doi.org/10.1001/jama.2017.7156)] [Medline: [28586821](https://pubmed.ncbi.nlm.nih.gov/28586821/)]
7. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. *J Natl Cancer Inst* 2017 Sep 01;109(9):dix029. [doi: [10.1093/jnci/dix029](https://doi.org/10.1093/jnci/dix029)] [Medline: [28423407](https://pubmed.ncbi.nlm.nih.gov/28423407/)]
8. Foster A, Croot L, Brazier J, Harris J, O'Cathain A. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. *J Patient Rep Outcomes* 2018 Dec;2:46 [FREE Full text] [doi: [10.1186/s41687-018-0072-3](https://doi.org/10.1186/s41687-018-0072-3)] [Medline: [30363333](https://pubmed.ncbi.nlm.nih.gov/30363333/)]
9. Girgis A, Delaney G, Miller A. Utilising ehealth to support survivorship care. *Cancer Forum* 2015;39(2):86-89 [FREE Full text]
10. Girgis A, Durcinoska I, Levesque JV, Gerges M, Sandell T, Arnold A, PROMPT-Care Program Group. eHealth system for collecting and utilizing patient reported outcome measures for personalized treatment and care (PROMPT-Care) among cancer patients: mixed methods approach to evaluate feasibility and acceptability. *J Med Internet Res* 2017 Oct 02;19(10):e330 [FREE Full text] [doi: [10.2196/jmir.8360](https://doi.org/10.2196/jmir.8360)] [Medline: [28970188](https://pubmed.ncbi.nlm.nih.gov/28970188/)]
11. Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998 Jan 24;316(7127):285. [doi: [10.1136/bmj.316.7127.285](https://doi.org/10.1136/bmj.316.7127.285)] [Medline: [9472515](https://pubmed.ncbi.nlm.nih.gov/9472515/)]
12. Girgis A, Durcinoska I, Gerges M, Kaadan N, Arnold A, Descallar J, et al. Study protocol for a controlled trial of an eHealth system utilising patient reported outcome measures for personalised treatment and care: PROMPT-Care 2.0. *BMC Cancer* 2018 Aug 23;18(1):845 [FREE Full text] [doi: [10.1186/s12885-018-4729-3](https://doi.org/10.1186/s12885-018-4729-3)] [Medline: [30139331](https://pubmed.ncbi.nlm.nih.gov/30139331/)]
13. Riba MB, Donovan KA, Andersen B, Braun I, Breitbart WS, Brewer BW, et al. Distress management, Version 3.2019, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2019 Oct 01;17(10):1229-1249. [doi: [10.6004/jnccn.2019.0048](https://doi.org/10.6004/jnccn.2019.0048)] [Medline: [31590149](https://pubmed.ncbi.nlm.nih.gov/31590149/)]
14. Richardson LA, Jones GW. A review of the reliability and validity of the Edmonton Symptom Assessment System. *Curr Oncol* 2009 Jan;16(1):55 [FREE Full text] [doi: [10.3747/co.v16i1.261](https://doi.org/10.3747/co.v16i1.261)] [Medline: [19229371](https://pubmed.ncbi.nlm.nih.gov/19229371/)]
15. Girgis A, Stojanovski E, Boyes A, King M, Lecathelinais C. The next generation of the supportive care needs survey: a brief screening tool for administration in the clinical oncology setting. *Psychooncology* 2012 Aug;21(8):827-835. [doi: [10.1002/pon.1973](https://doi.org/10.1002/pon.1973)] [Medline: [21484938](https://pubmed.ncbi.nlm.nih.gov/21484938/)]
16. Schuler T, Miller AA. PROsaic: A smart device-based and EMR-integrated system for patient-reported outcome measurement in routine cancer care. *Journal of Radiation Oncology Informatics* 2014;6(1):111-131. [doi: [10.5166/jroi-6-1-24](https://doi.org/10.5166/jroi-6-1-24)]
17. Girgis A, Durcinoska I, Koh E, Ng W, Arnold A, Delaney GP, PROMPT-Care Pathways Working Group. Development of health pathways to standardize cancer care pathways informed by patient-reported outcomes and clinical practice guidelines. *JCO Clin Cancer Inform* 2018 Dec;2:1-13. [doi: [10.1200/CCI.18.00024](https://doi.org/10.1200/CCI.18.00024)] [Medline: [30652587](https://pubmed.ncbi.nlm.nih.gov/30652587/)]
18. Girgis A, Durcinoska I, Arnold A, Delaney GP. Interpreting and acting on the PRO scores from the patient-reported outcomes for personalized treatment and care (PROMPT-Care) eHealth system. *Med Care* 2019 May;57 Suppl 5 Suppl 1:S85-S91. [doi: [10.1097/MLR.0000000000001094](https://doi.org/10.1097/MLR.0000000000001094)] [Medline: [30985601](https://pubmed.ncbi.nlm.nih.gov/30985601/)]
19. Australian Bureau of Statistics. Socio-Economic Indexes for Areas (SEIFA). Technical Paper. Canberra: Australian Bureau of Statistics; 2018.
20. Murphy B, Fraeman KH. A general SAS® macro to implement optimal N:1 propensity score matching within a maximum radius. 2017 Presented at: SAS Global Forum; April 2-5; Florida URL: <https://support.sas.com/resources/papers/proceedings17/0812-2017.pdf>
21. Bureau of Health Information. Spotlight on measurement: Emergency department utilisation by people with cancer. Sydney: Bureau of Health Information; 2014. URL: [http://www.bhi.nsw.gov.au/data/assets/pdf\\_file/0005/249692/CancerED\\_Spotlight.pdf](http://www.bhi.nsw.gov.au/data/assets/pdf_file/0005/249692/CancerED_Spotlight.pdf) [accessed 2020-09-23]
22. Howell D, Li M, Sutradhar R, Gu S, Iqbal J, O'Brien MA, et al. Integration of patient-reported outcomes (PROs) for personalized symptom management in. *Support Care Cancer* 2020 Oct;28(10):4933-4942. [doi: [10.1007/s00520-020-05313-3](https://doi.org/10.1007/s00520-020-05313-3)] [Medline: [32020357](https://pubmed.ncbi.nlm.nih.gov/32020357/)]
23. Cox K, Wilson E. Follow-up for people with cancer: nurse-led services and telephone interventions. *J Adv Nurs* 2003 Jul;43(1):51-61. [doi: [10.1046/j.1365-2648.2003.02672.x](https://doi.org/10.1046/j.1365-2648.2003.02672.x)] [Medline: [12801396](https://pubmed.ncbi.nlm.nih.gov/12801396/)]
24. Australian Bureau of Statistics. Household use of information technology, Australia, 2016-2017.: Australian Bureau of Statistics; 2018. URL: <https://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0> [accessed 2020-09-23]
25. Chow PI, Drago F, Kennedy EM, Chambers N, Sheffield C, Cohn WF. Examining the feasibility, acceptability, and potential utility of mobile distress screening in adult cancer patients. *Psychooncology* 2019 Sep;28(9):1887-1893. [doi: [10.1002/pon.5168](https://doi.org/10.1002/pon.5168)] [Medline: [31269306](https://pubmed.ncbi.nlm.nih.gov/31269306/)]
26. Hamilton J, Kroska EB. Distress predicts utilization of psychosocial health services in oncology patients. *Psychooncology* 2019 Jan;28(1):61-67. [doi: [10.1002/pon.4910](https://doi.org/10.1002/pon.4910)] [Medline: [30286522](https://pubmed.ncbi.nlm.nih.gov/30286522/)]

27. Götz A, Kröner A, Jenewein J, Spirig R. Evaluation of the adherence of distress screening with the distress thermometer in cancer patients 4 years after implementation. *Support Care Cancer* 2018 Dec 11;27(8):2799-2807. [doi: [10.1007/s00520-018-4579-1](https://doi.org/10.1007/s00520-018-4579-1)] [Medline: [30539312](https://pubmed.ncbi.nlm.nih.gov/30539312/)]
28. Resnicow K, Baranowski T, Ahluwalia JS, Braithwaite RL. Cultural sensitivity in public health: defined and demystified. *Ethn Dis* 1999;9(1):10-21. [Medline: [10355471](https://pubmed.ncbi.nlm.nih.gov/10355471/)]

## Abbreviations

**ED:** emergency department

**ESAS:** Edmonton Symptom Assessment Scale

**HL7:** Health Level Seven

**OIS:** oncology information system

**PRO:** patient-reported outcomes

**PROMPT-Care:** Patient Reported Outcome Measures for Personalized Treatment and Care

*Edited by G Eysenbach; submitted 28.04.20; peer-reviewed by C Foster, C Fincham, Y Chu; comments to author 08.06.20; revised version received 04.08.20; accepted 18.08.20; published 29.10.20.*

*Please cite as:*

Girgis A, Durcinoska I, Arnold A, Descallar J, Kaadan N, Koh ES, Miller A, Ng W, Carolan M, Della-Fiorentina SA, Avery S, Delaney GP

*Web-Based Patient-Reported Outcome Measures for Personalized Treatment and Care (PROMPT-Care): Multicenter Pragmatic Nonrandomized Trial*

*J Med Internet Res* 2020;22(10):e19685

URL: <http://www.jmir.org/2020/10/e19685/>

doi: [10.2196/19685](https://doi.org/10.2196/19685)

PMID: [33118954](https://pubmed.ncbi.nlm.nih.gov/33118954/)

©Afaf Girgis, Ivana Durcinoska, Anthony Arnold, Joseph Descallar, Nasreen Kaadan, Eng-Siew Koh, Andrew Miller, Weng Ng, Martin Carolan, Stephen A Della-Fiorentina, Sandra Avery, Geoff P Delaney. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Associations Between Patient Health Outcomes and Secure Message Content Exchanged Between Patients and Clinicians: Retrospective Cohort Study

Dawn M Heisey-Grove<sup>1,2</sup>, MPH, PhD; Laura E McClelland<sup>2</sup>, PhD; Cheryl Rathert<sup>3</sup>, PhD; Alexander Tartaglia<sup>2</sup>, DMin; Kevin Jackson<sup>4</sup>, RRT, PhD; Jonathan P DeShazo<sup>2</sup>, MPH, PhD

<sup>1</sup>MITRE Corporation, McLean, VA, United States

<sup>2</sup>College of Health Professions, Virginia Commonwealth University, Richmond, VA, United States

<sup>3</sup>Department of Health Management and Policy, Saint Louis University, St Louis, MO, United States

<sup>4</sup>Department of Nursing and Allied Health, College of Science, Engineering, and Technology, Norfolk State University, Norfolk, VA, United States

**Corresponding Author:**

Dawn M Heisey-Grove, MPH, PhD

MITRE Corporation

7525 Colshire Drive

McLean, VA, 22102

United States

Phone: 1 7035477389

Email: [heiseygroved@mitre.org](mailto:heiseygroved@mitre.org)

## Abstract

**Background:** The number of electronic messages securely exchanged between clinic staff and patients has risen dramatically over the last decade. A variety of studies explored whether the volume of messages sent by patients was associated with outcomes. None of these studies, however, examined whether message content itself was associated with outcomes. Because secure messaging is a significant form of communication between patients and clinic staff, it is critical to evaluate the context of the communication to best understand its impact on patient health outcomes.

**Objective:** To examine associations between patients' and clinicians' message content and changes in patients' health outcomes.

**Methods:** We applied a taxonomy developed specifically for secure messages to 14,394 patient- and clinic staff-generated messages derived from patient-initiated message threads. Our study population included 1602 patients, 50.94% (n=816) of whom initiated message threads. We conducted linear regression analyses to determine whether message codes were associated with changes in glycemic (A1C) levels in patients with diabetes and changes in systolic (SBP) and diastolic (DBP) blood pressure in patients with hypertension.

**Results:** Patients who initiated threads had larger declines in A1Cs ( $P=.01$ ) compared to patients who did not initiate threads. Clinic nonresponse was associated with decreased SBP ( $\beta=-.30$ ; 95% CI  $-0.56$  to  $-0.04$ ), as were staffs' action responses ( $\beta=-.30$ ; 95% CI  $-0.58$  to  $-0.02$ ). Increased DBP, SBP, and A1C levels were associated with patient-generated appreciation and praise messages and staff encouragement with effect sizes ranging from 0.51 (A1C) to 5.80 (SBP). We found improvements in SBP associated with patients' complaints ( $\beta=-4.03$ ; 95% CI  $-7.94$  to  $-0.12$ ). Deferred information sharing by clinic staff was associated with increased SBP ( $\beta=1.29$ ; 95% CI  $0.4$  to  $2.19$ ).

**Conclusions:** This is the first research to find associations between message content and patients' health outcomes. Our findings indicate mixed associations between patient message content and patient outcomes. Further research is needed to understand the implications of this work; in the meantime, health care providers should be aware that their message content may influence patient health outcomes.

(*J Med Internet Res* 2020;22(10):e19477) doi:[10.2196/19477](https://doi.org/10.2196/19477)

**KEYWORDS**

health information technology; electronic messaging; patient physician communication; diabetes; hypertension

## Introduction

### Background and Significance

The use of secure messaging—email messages exchanged between patients and clinical staff through a secure platform—has increased significantly over the last 2 decades as patients' access to the functionality increased [1-5]. Patients reported that secure messaging offered convenience, with the added benefit of documenting the conversation so that it could be referenced later [6-9]. Although clinicians cited challenging workflows as the biggest barrier to use [10], they noted that secure messaging improved communication between visits and boosted patient engagement, satisfaction, and trust [7,9].

Communication between patients and clinicians should include information exchange, uncertainty management, relationship development and fostering, and activities that enable decision making and health self-management [11]. According to Street et al [11], these communication functions lead to proximal and intermediate outcomes that eventually result in improved patient health outcomes. Prior research provides evidence that secure messaging is a significant modality for patient-clinician communication [4,12,13]. As such, message threads that include communication functions identified by Street et al [11] should be associated with better health outcomes.

To the authors' knowledge, there is no published research to date that linked secure message content and patients' health outcomes. Rather, researchers have explored whether the number of secure messages, or secure message use more generally, was associated with outcomes for a variety of conditions. An equal number of studies found that secure message use was associated with improvements in blood pressure control [14-16], or no association between them [12,16,17]. While a number of studies identified positive associations between secure message use and controlled glycemic levels [12,14,17-19], 2 studies identified no association [18,20], and 1 found inconsistent associations [16]. Even within studies, findings varied across conditions. For example, Price-Haywood et al [16] assessed performance on population-level measures for glycemic and blood pressure control and found that both improved among some patients with diabetes with evidence of a dose response. This did not hold, however, among patients with diabetes who had glycemic levels over 8%, nor for blood pressure changes among patients with hypertension. Similarly, Harris et al [18] found that the highest users of secure messaging had better glycemic control but did not identify similar patterns with blood pressure control. An analysis by Shimada et al [17] separated prescription refill requests and other types of secure message use and found that improvements in glycemic control after 2 years were associated with secure message use but not with prescription refills. Interestingly, they found the reverse was true for blood pressure control. These mixed findings, particularly the findings from Shimada et al [17], suggest that we must move beyond counting messages and begin classifying and quantifying message content types in order to better understand how secure messages, and more specifically, secure messaging content, impact important patient health outcomes. This research study directly attempts to address this need.

### Objective

We created and applied a theory-based taxonomy developed specifically for secure messaging to a large sample of patient- and clinician-generated messages and explored whether certain types of message content were associated with changes in glycemic levels among patients with diabetes and changes in blood pressure among patients with hypertension [21,22]. Our taxonomy provides taxa (ie, codes) for patient- and clinic staff-generated content. [Multimedia Appendix 1](#) includes a complete list of taxa (ie, codes), their definitions, and examples of each taxon. We included taxa (italicized throughout this paper) to classify patient-generated *Information seeking* and *Information sharing* (eg, *Self-reporting* and *Clinical updates*), and *Task-oriented requests* such as *Appointment scheduling*, *Prescription refill requests*, and *Other administrative requests*. We used different taxa to classify clinic staff-generated *Information sharing* and *Action responses* that indicate the degree of request fulfillment, and *Information seeking*. Finally, we classified *Social communication* that may be related to fostering relationships and trust among messages generated by either party.

Based on Mishel's Uncertainty in Illness Theory [23] and Street et al's [11] framework, we anticipated that actions indicative of self-care, such as *Task-oriented requests* not associated with uncertainty (eg, routine appointment requests, prescription refills) and *Self-reporting*, would be associated with improved health outcomes. Similarly, actions from the patients that might be indicative of trust between patient and clinician, such as *Information sharing* and positive *Social communication*, would be associated with improved health outcomes. Clinicians' *Information sharing* and *Encouragement* could be a mechanism to mitigate patients' uncertainty and improve trust, so we hypothesized that receipt of this message content would also result in improved health outcomes. Conversely, we anticipated that clinician responses that did not mitigate patients' uncertainty would result in poorer outcomes. We included in this category lack of a response from clinic staff to patient-initiated threads and clinic staffs' *Information sharing/Deferrals*. Finally, we expected that patients who expressed negative *Social communication* (eg, *Complaints*) would also experience poorer health outcomes.

We present this nascent research to provide early evidence that the content of some secure messages between patients and clinic staff may be predictive of outcomes for patients with chronic illness.

## Methods

### Study Population

Our study population included adult patients who registered with the patient portal of a large urban academic medical center. The portal allowed registered patients to send and receive secure messages with clinic staff, request appointments and prescription refills and renewals, view upcoming appointments and notes from prior visits, and find links for health-related educational materials and bill pay. Patients could access the portal through any device with a web browser.



Patients had to have at least one inpatient or two outpatient visits in 2016 with ICD-10-DM diagnosis codes for either diabetes (E11) or hypertension (I10), and one visit in 2018. We stratified the sample based on health condition (diabetes only, hypertension only, or both conditions) and whether patients initiated a message thread between January 1 and December 31, 2017, then randomly selected samples from each stratum. Patients who lacked baseline or endpoint values for the outcomes of interest were excluded from this study.

We included 2 different conditions in our research to control for disease condition, not to provide specific recommendations relative to condition. Our message sample included all patient-initiated threads generated by those sampled patients and saved to patients' charts between January 1 and December 31, 2017. We included only patient-initiated threads because we felt these were the best markers of patient uncertainty and self-management. This research received approval from the VCUHS Institutional Review Board under expedited review. We manually extracted messages from patients' electronic health records and redacted all identifiable information during the extraction process. We coded those deidentified messages, which were linked with a unique identifier not linked to the patients' medical record.

Our study sample consisted of 1602 patients (full population), of whom 50.94% (n=816) initiated at least one message thread (secure message-only population). We included patients with diabetes only (n=347), hypertension only (n=751), and both conditions (n=504). We coded 5844 message threads initiated by these patients, which included 8008 patient-generated messages and 6386 messages generated by 496 unique clinic staff. Our sampled population generated an average of 9.81 messages (median 5; max 117). Message responses to patient-initiated messages came from physicians, nursing staff, administrative staff, pharmacists, physician assistants, medical assistants, podiatrists, social workers, and medical technicians from departments and clinics across the medical center. Clinics employed a triaging system whereby administrative or nursing staff review incoming patient-generated messages and determine the best response approach. Decisions about which staff would serve as the triage point, and the triaging process itself, varied across clinics.

## Independent Variables

Table 1 lists the taxa, or codes, used in these analyses. Multimedia Appendix 1 provides the taxa definitions and examples for each. We created the taxonomy by leveraging common taxa reported in other published literature. Our 2 patient-generated *Information seeking* taxa were selected based on Mishel's Uncertainty in Illness Theory [23], which identifies the reasons patients might outreach to clinicians to manage their uncertainty around their illness. We included *Information seeking* and *Information sharing* taxa for both patients and clinicians in recognition that information exchange is a communication function on the pathway to improved patient outcomes [11]. In addition, we included other constructs from the Street et al [11] pathway, including *Task-oriented* patient-generated requests that might be markers of patient self-management, and *Social communication*. We leveraged the Taxonomy of Requests by Patients [24] for clinic staff-generated *Action responses* to those patient requests. We piloted our taxonomy with a small sample to ensure that no constructs were missing and that the appropriate level of granularity was present in the taxa [22].

We used the taxa to distinguish between different types of patient-generated and clinic staff-generated message content. For these analyses, we report findings for the individual taxa as well as the level 1 groupings of taxa for patient *Information seeking*, patient *Information sharing*, patient *Social communication*; patient *Task-oriented requests* reflective of self-management; other patient *Task-oriented requests*; staff *Information sharing*; and staff *Action responses*. We based our independent variables on counts of taxa either sent or received by patients between January 1 and December 31, 2017. Because we found a strong correlation between the likelihood of sending and receiving a taxon based on patients' thread volume, our independent variables measure taxa as a function of volume. Each taxon is represented in the linear regression models as a proportion of the total patient-generated or clinic staff-generated taxa they sent or received.

We also created an independent variable that measured clinic nonresponse, defined as a thread that included no messages sent from clinic staff. We measured clinic nonresponse as a proportion of the total threads initiated by the patient.



**Table 1.** Secure message taxonomy.

Patient or clinician generated taxon, Level 1, and Level 2 taxon	Level 3 taxon
<b>Patient generated</b>	
<b>Information seeking</b>	
Logistics	N/A <sup>a</sup>
Medical guidance	N/A
<b>Information sharing</b>	
Clinical update	N/A
Response to clinician's message	N/A
Self-reporting	N/A
<b>Task oriented</b>	
Prescription refills and requests	N/A
New or change prescription request	N/A
Other administrative	N/A
Referral requests	N/A
Scheduling request	Cancellation, Follow-up, Laboratory test or diagnostic procedure, New condition or symptom, Preventive care or physical examination, Reschedule
<b>Social communication</b>	
Appreciation or praise	N/A
Complaints	N/A
Life issues	N/A
<b>Clinic staff generated</b>	
<b>Action responses</b>	
Acknowledge	N/A
Denies	N/A
Fulfills request	N/A
Partially fulfills request	N/A
<b>Information seeking</b>	
N/A	N/A
<b>Information sharing</b>	
Deferred	N/A
Medical guidance	N/A
Orientation to procedures, treatments, or preventive behaviors	N/A
<b>Task oriented</b>	
Recommendation to schedule appointment	N/A
<b>Social communication</b>	
Encouragement	N/A

<sup>a</sup>N/A: Not applicable.

## Dependent Variables

We created 1 dependent variable for patients with diabetes and 2 for patients with hypertension. For patients with diabetes, we measured the change between the endpoint and baseline measures of glycemic control (A1C). For patients with hypertension, we included dependent variables that measured

changes between baseline and endpoint measures for systolic blood pressure (SBP) and diastolic blood pressure (DBP). We used the last recorded value in 2016 as the baseline measure. Our endpoint value was the first measured value obtained between January and June 2018. If multiple blood pressures were taken on the same day, we averaged available values.

## Covariates

Based on prior literature relating to patient–clinician communication and electronic communication practices [25–29], we expected differences in taxa use based on patient and clinician characteristics. We therefore controlled for patient age as of January 1, 2017; patient sex; race (Black, White, and other); payer type (public, private, uninsured, or other); rural home location as a bivariate derived from Rural–Urban Commuting Area codes [30]; health condition (diabetes only, hypertension only, or both conditions); the number of outpatient and inpatient visits during 2017; and the number of threads initiated during 2017. We also included baseline A1C and blood pressure values in models measuring change in glycemic control and blood pressure, respectively. For models that included only patients who initiated message threads (ie, secure message–only population), we included the average distance between zip code centroids of patients’ homes and the clinics to which they sent messages.

## Qualitative Analyses

We assigned taxa to all messages—those generated by patients and clinic staff—that were saved to the patient’s chart and part of patient-initiated threads created and completed between January 1 and December 31, 2017. Our context unit was the message thread. Coding units were no longer than a single message but could be shorter depending on the content in the message (eg, if multiple taxa are applied to the message). A given message was assigned as many taxa as there were concepts in the message; however, we limited the number of times a given taxon (ie, a single code) could be counted for each message to 1 per message.

A primary coder (DH-G) assigned taxa to all messages while a secondary coder (JDS) applied taxa to a random sample of messages (n=1908). The primary coder trained the secondary coder based on a set of definitions and sample coded text collected from a pilot study [22]; these samples and definitions were refined as the coding process continued. We conducted

the coding in batches and discrepancies were discussed and reconciled following the completion of each batch. The primary coder recoded each batch as appropriate based on those discussions. [Multimedia Appendix 2](#) lists the interrater and intrarater reliability coefficients for the last coded batch. Intrarater reliability ranged from fair to excellent. Three taxa received a poor kappa rating when comparing the results from the 2 coders (interrater reliability): clinician-generated *Action response/Denies*, *Recommendation to schedule*, and patient-generated *Information seeking/Logistics*.

## Quantitative Analyses

We estimated unadjusted differences by patient characteristics based on use of secure messaging by applying chi-square analyses for categorical variables and unpaired *t* test of means for continuous variables. We executed 2 linear regression analyses for each combination of taxon and dependent variable. The first model used the full population and the second used the secure message–only population. The comparison in the full population models included all patients who did not initiate a message thread and those patients who sent or received messages with the selected taxon. Models that included the secure message–only population compared patients who sent or received messages coded with the selected taxon to those who sent or received other types of messages. We report results as the unstandardized regression coefficient ( $\beta$  weight). We conducted all analyses using SAS version 9.4 (SAS Institute).

## Results

[Table 2](#) presents the descriptive statistics for our study population, comparing the populations who sent messages in 2017 to those who did not. Among patients with diabetes, we observed differences in the use of secure messaging by age, condition, insurance type, and sex. Except for age, we observed similar differences in use of secure messaging among patients with hypertension.

**Table 2.** Comparison of study population's characteristics by use of secure messaging in 2017.<sup>a,b</sup>

Characteristics	Patients with diabetes			Patients with hypertension		
	Sent messages (N=430)	Did not send messages (N=421)	<i>P</i> value	Sent messages (N=621)	Did not send messages (N=634)	<i>P</i> value
Age in years, mean	57.84	59.80	.02	59.97	58.65	.08
Distance between home and clinic in miles, mean	27.48	N/A <sup>c</sup>	N/A	33.47	N/A	N/A
<b>Conditions</b>						
Both, n (%)	235 (54.7)	269 (63.9)	.006	235 (37.8)	269 (42.4)	.10
Diabetes only, n (%)	195 (45.3)	152 (36.1)	.006	N/A	N/A	N/A
Hypertension only, n (%)	N/A	N/A	N/A	386 (62.2)	365 (57.6)	.10
<b>Home location</b>						
Rural, n (%)	9 (2.1)	17 (4.0)	.10	17 (2.7)	29 (4.6)	.08
Urban, n (%)	421 (97.9)	404 (96.0)	.10	604 (97.3)	605 (95.4)	.08
<b>Insurance</b>						
Other, n (%)	126 (29.3)	93 (22.1)	.02	160 (25.8)	160 (25.2)	.83
Private, n (%)	138 (32.1)	80 (19.0)	<.001	155 (25.0)	110 (17.4)	<.001
Public, n (%)	161 (37.4)	241 (57.2)	<.001	296 (47.7)	349 (55.0)	<.001
Uninsured, n (%)	5 (1.2)	7 (1.7)	.54	10 (1.6)	15 (2.4)	.34
<b>Race</b>						
Black, n (%)	193 (44.9)	214 (50.8)	.08	231 (37.2)	316 (49.8)	<.001
Other, n (%)	24 (5.6)	27 (6.4)	.61	23 (3.7)	26 (4.1)	.72
White, n (%)	213 (49.5)	180 (42.8)	.05	365 (58.8)	291 (45.9)	<.001
<b>Sex</b>						
Male, n (%)	134 (31.2)	182 (43.2)	<.001	235 (37.8)	279 (44.0)	.03
Female, n (%)	296 (68.8)	239 (56.8)	<.001	386 (62.2)	355 (56.0)	.03

<sup>a</sup>Percentages represent the proportion of the population with that characteristic.

<sup>b</sup>The *P* value is the unadjusted estimate of statistical difference between the populations who sent secure messages and those who did not.

<sup>c</sup>N/A: Not applicable.

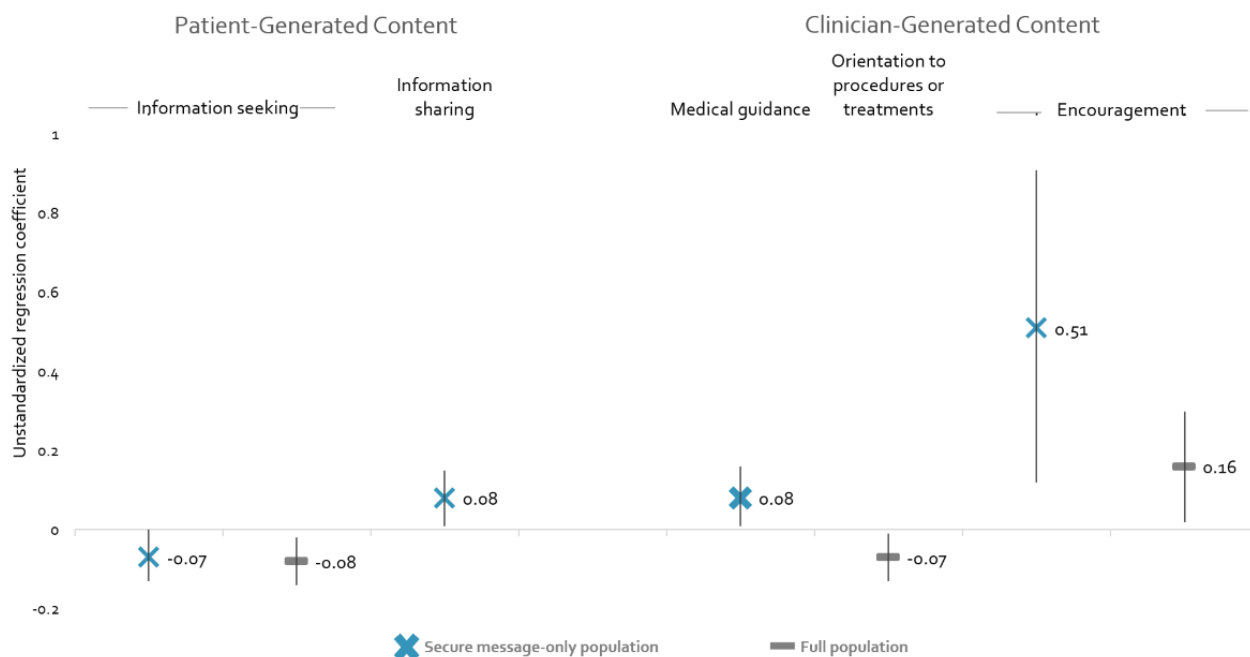
### Change in A1C Among Patients With Diabetes

Among patients who initiated threads, we observed a mean A1C decrease of  $-0.56$ , with the 2018 value being statistically lower on average than the 2016 value ( $P<.001$ ). The same was not true among patients who did not initiate threads ( $P=.20$ ). We observed differential changes between 2016 and 2018 when comparing patients with a baseline A1C indicating controlled diabetes ( $<7.0$ ) versus uncontrolled diabetes ( $>7.0$ ). Among patients who sent secure messages, we observed a mean increase in A1C of  $.04$  for patients with controlled diabetes ( $n=151$ ), compared to a mean decrease of  $-1.22$  among patients with uncontrolled diabetes ( $n=138$ ).

Figure 1 displays statistically significant associations between taxa and A1C changes for the taxa. Taxa not represented in the table were not associated with A1C changes at  $P<.05$  but are available in Multimedia Appendix 3, which also lists the *P*-value

for all covariates. Baseline A1C was the only covariate with statistical significance ( $P<.001$ ) across all analyses.

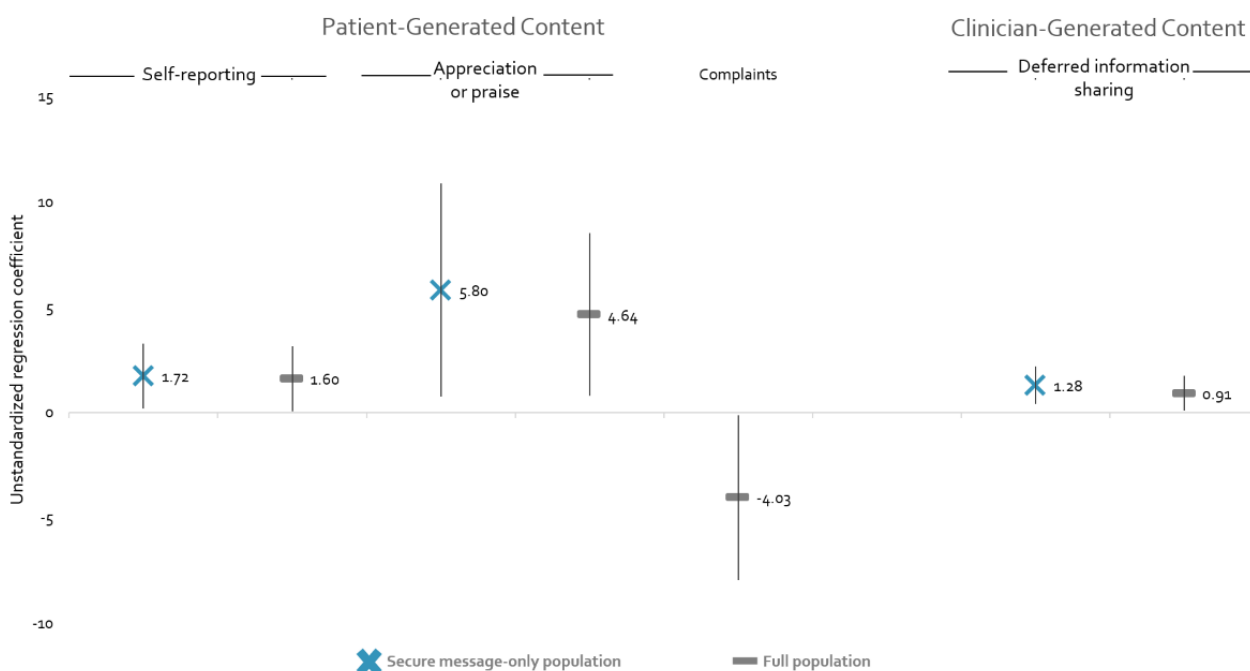
As their proportion of *Information seeking* increased, patients experienced greater declines in their A1C values. This was true when comparing patients within the full population and among the secure message-only population. Conversely, patients who shared information with their clinic staff experienced A1C increases, compared to those who sent other types of messages to clinic staff. Three clinician-generated subtaxa were associated with A1C changes. Patients who received *Orientation to procedures and treatments* had declines in A1C compared to patients who did not initiate threads ( $\beta=-.07$ ; 95% CI  $-0.13$  to  $-0.01$ ), whereas patients who received *Medical guidance* had increased A1C compared to patients who received other content from clinic staff ( $\beta=.08$ ; 95% CI  $0.01$ - $0.16$ ). Similarly, patients who received *Encouragement* from clinic staff experienced increased A1C values ( $\beta=.16$ ; 95% CI  $0.02$ - $0.03$ ) between 2016 and 2018.

**Figure 1.** Associations between taxa and A1C changes.

### Change in SBP Among Patients With Hypertension

Overall, we observed an unadjusted average increase in SBP between 2016 and 2018 among patients who initiated threads (3.41-point increase;  $P < .01$ ) and patients who did not initiate threads (2.45-point increase;  $P = .02$ ). Among patients who sent secure messages, those with controlled systolic blood pressure ( $< 120$  mmHg) experienced a mean increase in SBP of 16.41 ( $n = 149$ ), compared to a mean decrease of  $-0.71$  among those with uncontrolled SBP ( $n = 471$ ).

Figure 2 presents the taxa associated with SBP changes among these populations. Full analysis results are available in [Multimedia Appendix 3](#). Two patient-generated taxa (*Self-reporting* and *Appreciation or praise*) and the clinician-generated taxon for *Information sharing/Deferral* were associated with increased SBP. This was true in both population comparisons. By contrast, we observed decreased SBP among patients who sent *Complaints* compared to patients who did not initiate threads. Covariates of statistical significance included baseline SBP ( $P < .001$ ), race (Black vs White;  $P < .05$ ), and, for some analyses, age ( $P < .05$ ) and number of outpatient visits ( $P < .05$ ).

**Figure 2.** Associations between taxa and SBP changes. SBP: systolic blood pressure.

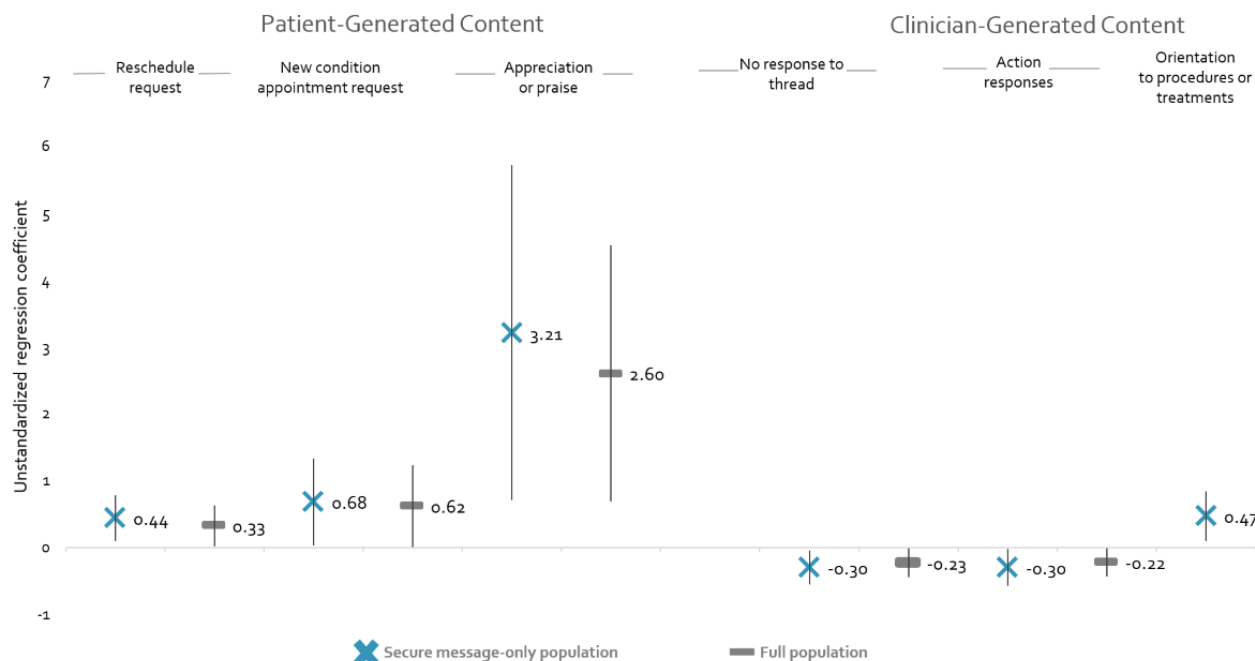
## Change in DBP Among Patients With Hypertension

Among patients who initiated threads and those who did not, we found statistically significant unadjusted increases in DBP between 2016 and 2018 of 1.71 and 1.67, respectively ( $P=.007$  and  $P=.003$ , respectively). Among patients who sent secure messages, patients with controlled DBP ( $n=351$ ) at baseline ( $<80$  mmHg) experienced a mean increase in DBP of 6.57,

compared to a mean decrease of  $-4.62$  among patients with uncontrolled DBP ( $n=269$ ) at baseline ( $>80$  mmHg).

Figure 3 presents the associations between taxa and changes in DBP that were statistically significant at  $P<.05$ . Full analysis results are available in [Multimedia Appendix 3](#). Three covariates were statistically significant across all analyses (age [ $P<.01$ ], baseline DBP [ $P<.001$ ], and number of outpatient visits [ $P<.05$ ]).

**Figure 3.** Associations between taxa and DBP changes. DBP: diastolic blood pressure.



Three patient-generated taxa were associated with increased DBP (*Schedule request/Reschedule*, *Schedule request/New symptoms or conditions*, and *Appreciation or praise*). Similar effect sizes were observed across the 2 models for each taxon. We also observed that as the prevalence of clinic staffs' nonresponse increased, patients experienced greater declines in DBP. Similarly, patients who received proportionally more *Action response*-related content (including *Acknowledgment*, *Fulfills request*, and *Partially fulfills request*) had greater DBP decreases. Patients who received *Orientation to procedures or treatments* had correspondingly increased DBP, compared to patients who did not receive those kinds of messages from clinic staff.

## Discussion

### Principal Findings

We present here the first findings exploring associations between secure message content and patient health outcomes. Our research found associations between selected message content and changes in patients' glycemic levels and blood pressures. As we anticipated, patients who sent *Information seeking* messages, or who received *Orientation to procedures or treatments* messages from clinic staff, experienced greater decreases in A1C. Consistent with other research [12,14,16-19], we found an overall association between secure messaging use and improved A1C. Also as expected, we observed DBP

decreases among patients who received confirmation of action, SBP increases in response to *Information sharing/Deferral*, and DBP increases associated with *Scheduling request/New condition*.

Counter to our hypotheses, however, we found that A1C increased among patients who shared information with clinic staff. We were only able to identify a statistically significant association in the level 1 *Information sharing* taxon ( $P=.02$ ). Associations in the *Information sharing* subtaxa were not statistically significant ( $P>.05$ ). Detecting an association at the grouped *Information sharing* level makes it challenging to interpret these results, as patients who provide *Clinical updates* may be receiving care from multiple providers and may have other health conditions that impact their glycemic levels. Patients who self-report to their clinicians, however, may be practicing self-management of their condition which, according to Street et al [11], should be associated with improved patient health outcomes. Similarly, information exchange, as represented in the *Information sharing/Response to clinician's message*, should be associated with improved outcomes. Our inability to detect a statistically significant association at the subtaxa level for *Information sharing* makes it difficult to interpret these results further. Future studies should consider expanding the sample size to improve the ability to detect associations at the subtaxa level.



We also observed DBP increases associated with certain types of patient-generated *Scheduling requests* and *Information sharing* by clinic staff. It could be that these patients had new procedures or other treatments that contributed to their DBP increases which could be a confounder to any possible influence on outcomes. Although our analyses controlled for patients' number of outpatient and inpatient visits, we did not control for severity of co-occurring conditions or types of procedures patients may have undergone during the study period. We did find that baseline DBP was a significant ( $P<.001$ ) predictor of the patients' change in DBP between baseline and endpoint.

We also observed an inverse association between DBP and clinic nonresponse: patients' DBP decreased as nonresponse prevalence increased. Our findings should not be taken as advocating for nonresponse to improve patients' health, because there are many unanswered questions about what is driving these findings. It will be important to assess when nonresponse may be appropriate and when it might have deleterious effects. For example, it is possible that certain types of patient-initiated threads do not require a clinic response, such as when a prescription refill request is completed, and patients are notified by their pharmacy that the prescription is ready rather than the clinic. Future research should explore the best communication modality for responses.

In the analyses that used only patients who initiated message threads, we found that as the prevalence of *Schedule request/Reschedule* increased, so did patients' DBP. We interpreted a reschedule request as a manifestation of self-care, following Mishel's Uncertainty in Illness Theory [23], because patients are taking charge of their health care visits by rescheduling to a time more convenient to them, thereby leading to less stress. Our findings indicate this is not the case. Instead, the poorer outcomes associated with *Schedule request/Reschedule* may be a manifestation of stress in patients' lives that required rescheduling medical appointments, a lack of or decline in activation, or a deficit in access to care (eg, transport issues). If these patients were not managing their stress and not maintaining their levels of self-care, their health outcomes might suffer as their patient activation threshold declined [31-33]. Consistent with this, we observed similar poor outcomes—with slightly larger effect sizes—associated with appointment requests for new conditions or symptoms.

Our research identified different effects by health condition on outcomes associated with staff sharing *Orientation to processes and treatments*. Patients with diabetes who received these messages had lower A1C levels in 2018 but patients with hypertension experienced increased DBP. Clinician-generated *Information sharing/Deferral* was associated with increased SBP among patients with hypertension, but we did not observe similar associations for A1C changes. It will be important to apply this taxonomy to other conditions to determine if other differences between outcomes and communication content exist by condition, to better improve communication between patients and clinic staff in ways that advance patients' health.

We included a number of covariates in our analyses, selected based on previous research findings that indicated their relevance to secure messaging use or patient health outcomes.

Not surprisingly given the differences we reported in the unadjusted changes in outcomes by patients' control status, baseline values for all outcome measures were statistically associated with the change between baseline and endpoint. In our adjusted analyses, we did not find a statistical relationship with health insurance type, number of inpatient visits, number of message threads, rural home location, or differences between Whites and other races. We were unable to include ethnicity in our analyses due to limitations in the source data. Consistent with other research that has found disparities in health outcomes between Black and White patients [34,35], we observed that Black patients experienced increases in SBP when compared with White patients after controlling for all other covariates in the analysis. Given that some message content was associated with patient outcomes and that race is also associated with these outcomes, future studies should explore whether the message content patients send and receive is associated with their race.

A recent literature review found that the majority of studies detected no association between clinicians' implicit bias and treatment recommendations when clinicians were asked to provide a diagnosis or treatment recommendation based on a written scenario [36]. Conversely, a different review noted that 5 of 6 observational and patient-reported measure-based studies found that physicians provided Black patients with less information than Whites [37]. Understanding how secure messaging communication varies by patient demographic characteristics and social determinants of health will be important to understanding how secure messaging might be used to improve patient outcomes in the future.

## Study Limitations

It is important to remember that this study examines the taxa in isolation; that is, a taxon is one component of the overall electronic conversation represented in each thread. From this research, we do not know what patient-generated messages preceded the staff response, so we cannot determine if, for example, *Orientation to procedures and treatments* responses answered patients' questions or were appropriate responses to the patient-initiated request. Analyses that explore the call-and-response nature of the message thread—that consider the initiating request, final response, and the pathway to get to that final response—should yield more insight into these results. For example, patients who requested an appointment but received *Orientation to procedures and treatments* may have poorer outcomes than patients whose request was partially or completely fulfilled. It may also be that the number of clinic staff involved in responding to a thread, or the time taken to respond, has an impact on patient outcomes by increasing uncertainty or reducing patients' trust [23]. Examining these factors might help explain why some of our findings do not align with our study hypotheses.

An important consideration for this research is that it demonstrates correlations and not causation. We hypothesized that *Self-reporting*, patients' *Appreciation and praise* for clinic staff, and staff *Encouragement* would be associated with improved outcomes but we found the opposite: poorer DBP and A1C values in 2018 were associated with the *Encouragement* and *Appreciation and praise* taxa and patients who self-reported

biometrics experienced increased SBP between the 2 years. Our outcomes were based on measurements obtained before and after the message collection period. If instead we obtained measurements in parallel to the secure messaging period, it is possible that we might have different results. For example, effects observed in 2018 may have less relevance to messages sent earlier in the calendar year (eg, patients only sent messages in the first quarter or half of the year). Another avenue of future study would include adding in more frequent measurements and exploring ways to identify any long-term impacts associated with certain taxa.

The Street et al [11] framework highlights intermediate outcomes on the pathway between communication functions and health outcomes. A proxy for the access to care construct, for example, might be overall health care utilization or whether the patients follow routine guidelines for care (eg, diabetic eye and foot examinations, or routine follow-up or preventive care appointments). Other constructs that could be measured with existing secondary data include self-care which might include the appropriate medication refill rates. These proximal outcomes also align to ones known to be associated with patient activation [38,39], further reinforcing the benefit to conducting these analyses. Future studies should explore associations between these taxa and those proximal and intermediate outcomes.

Responses to patients' messages are typically triaged by clinic staff such that only the most complex messages are shared with physicians [40-42]. Effective workflow design may be critical to effective and efficient responses. These workflows, however, may be very clinic and physician specific. Our study did not incorporate this aspect into our analysis, because we did not have data on workflow practices utilized by different clinics. We also know that communication behaviors vary by patient and clinician characteristics [25-29,43-48]. Our study controlled for patient characteristics, but not clinician characteristics. Although we captured the type of clinic staff who responded to individual messages, we did not control for that in our analyses because we aggregated counts based on patient and patients typically had responses from a variety of clinic staff types. Future studies should conduct analyses at the clinician and clinic levels to account for workflow and communication practice differences.

Among our patient population, we found patients with controlled glycemic levels or blood pressure had a mean increase over the study period, whereas patients with uncontrolled glycemic or blood pressure had a mean decrease. Our study was not powered sufficiently to identify associations based on whether patients' A1C or blood pressure was controlled at baseline [21]. It is likely, however, that patients' communication and clinic staff responses differ based on patients' current health status. We controlled for patients' baseline values for each health outcome measure and these values were significantly associated with all the outcomes ( $P<.001$ ), indicating a need for further research to better understand the differences in taxa use based on patients' baseline health status.

We used only messages saved to patients' charts. If clinic staff did not opt to save a message to the chart, it would not be captured in this study. We expect, therefore, that the numbers

presented in this paper underestimate the number of messages sent and received by patients. We also expect that the number of nonresponses was underrepresented because it seems likely that if clinic staff did not respond to a message, they would be less likely to save the message as well. It is also possible that messages we classified as nonresponse had a response that was not saved to patients' charts. If we assume that our sample underestimated the number of messages sent and received by patients, we would expect a bias toward the null and our results should therefore be viewed as conservative estimates of effect.

To our knowledge, only one other study quantified clinic nonresponse to patients' messages. Our study is the first to quantify nonresponse with a large number of messages and to link nonresponse to patient outcomes. The study by Lanham et al [49] study conducted chart reviews to determine if response occurred through other modalities and found that almost half of their 11 unanswered messages were resolved through other mechanisms. Our study did not assess responses by other communication modalities nor did it determine whether a response was warranted, but if we extrapolate the Lanham et al [49] findings to our work, that implies that only about half of the threads lacking a message response would have received a response not accounted for in our research (eg, phone, discussion during appointment). To better understand our study findings, it will be important to account for these other response types in future studies.

It is possible that thread initiation may be an indication of patient activation and engagement and clinic nonresponse may not inhibit patients' activation. Patient activation follows 4 stages: belief in the importance of engagement in the care processes; knowledge in what is needed to improve health; taking action to improve or maintain health; and finally, maintaining or persisting in those actions even when stressed [31]. Patients at higher stages of activation generally experience better outcomes, have lower health care costs, and higher rates of health screening and prevention activities [32,38,39,50]. Alexander et al [51] found that patients who communicated outside of office visits had higher patient activation rates. Consistent with their research, we found that patients who initiated threads experienced A1C improvements compared to patients who did not initiate threads.

Several of our taxa had poor interrater reliability scores and as a result, these findings should be viewed with caution. It is notable that none of these taxa were statistically associated with our outcomes, perhaps due to the potential lack of clarity in their definitions. Future studies using this taxonomy should seek to clarify the definitions for these taxa.

## Conclusion

This is the first study to explore associations between message content and patient health outcomes. We identified associations between certain patient- and clinic staff-generated taxa and changes in patients' glycemic levels and blood pressure. We also found that staff nonresponse was associated with improvements in patients' DBP, although the reasoning behind this association is unclear. Significantly more research is needed to better understand what we observed in our study, including exploring the context of the full electronic conversation and

outcomes, examining the temporal relationships between outcomes and message content, evaluating the impact of the potential confounder of patients' activation, exploring other intermediate outcomes that might be better measures of effect, and incorporating other communication modalities to capture

responses that occur outside of secure messaging. In the meantime, health care staff should be aware that message content is associated with patients' health outcomes when corresponding with patients through this medium.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Secure message taxonomy.

[DOCX File, 33 KB - [jmir\\_v22i10e19477\\_app1.docx](#)]

### Multimedia Appendix 2

Secure message taxa interrater and intrarater reliability.

[DOCX File, 19 KB - [jmir\\_v22i10e19477\\_app2.docx](#)]

### Multimedia Appendix 3

Taxa associations with health outcomes.

[DOCX File, 144 KB - [jmir\\_v22i10e19477\\_app3.docx](#)]

## References

1. Heisey-Grove D, Patel V, Searcy T. Physician electronic exchange of patient health information. Washington, DC: Office of the National Coordinator for Health Information Technology; 2015. URL: <https://dashboard.healthit.gov/evaluations/data-briefs/physician-electronic-exchange-patient-health-information.php> [accessed 2018-10-23]
2. Henry J, Pylypchuk Y, Patel V. Electronic capabilities for patient engagement among U.S. non-federal acute care hospitals 2012-2015. Washington, DC: Office of the National Coordinator for Health Information Technology; 2016. URL: <https://dashboard.healthit.gov/evaluations/data-briefs/hospitals-patient-engagement-electronic-capabilities-2015.php> [accessed 2016-08-21]
3. Tarver WL, Menser T, Hesse BW, Johnson TJ, Beckjord E, Ford EW, et al. Growth Dynamics of Patient-Provider Internet Communication: Trend Analysis Using the Health Information National Trends Survey (2003 to 2013). *J Med Internet Res* 2018 Mar 29;20(3):e109 [FREE Full text] [doi: [10.2196/jmir.7851](#)] [Medline: [29599107](#)]
4. Cronin R, Davis S, Shenson J, Chen Q, Rosenbloom S, Jackson G. Growth of Secure Messaging Through a Patient Portal as a Form of Outpatient Interaction across Clinical Specialties. *Appl Clin Inform* 2017 Dec 19;06(02):288-304. [doi: [10.4338/aci-2014-12-ra-0117](#)]
5. Shimada SL, Hogan TP, Rao SR, Allison JJ, Quill AL, Feng H, et al. Patient-Provider Secure Messaging in VA. *Medical Care* 2013;51:S21-S28. [doi: [10.1097/mlr.0b013e3182780917](#)]
6. Haun JN, Chavez M, Nazi K, Antinori N, Melillo C, Cotner BA, et al. Veterans' Preferences for Exchanging Information Using Veterans Affairs Health Information Technologies: Focus Group Results and Modeling Simulations. *J Med Internet Res* 2017 Oct 23;19(10):e359 [FREE Full text] [doi: [10.2196/jmir.8614](#)] [Medline: [29061553](#)]
7. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. *J Med Internet Res* 2013 Apr 04;15(4):e70 [FREE Full text] [doi: [10.2196/jmir.2443](#)] [Medline: [23557596](#)]
8. Rief JJ, Hamm ME, Zickmund SL, Nikolajski C, Lesky D, Hess R, et al. Using Health Information Technology to Foster Engagement: Patients' Experiences with an Active Patient Health Record. *Health Commun* 2017 Mar 25;32(3):310-319. [doi: [10.1080/10410236.2016.1138378](#)] [Medline: [27223684](#)]
9. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. *J Am Med Inform Assoc* 2013 May 01;20(3):519-525 [FREE Full text] [doi: [10.1136/amiajnl-2012-001253](#)] [Medline: [23242764](#)]
10. Hoonakker PLT, Carayon P, Cartmill RS. The impact of secure messaging on workflow in primary care: Results of a multiple-case, multiple-method study. *Int J Med Inform* 2017 Apr 23;100:63-76. [doi: [10.1016/j.ijmedinf.2017.01.004](#)] [Medline: [28241939](#)]
11. Street RLJ, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. *Patient Educ Couns* 2009 Mar 23;74(3):295-301. [doi: [10.1016/j.pec.2008.11.015](#)] [Medline: [19150199](#)]

12. Harris LT, Haneuse SJ, Martin DP, Ralston JD. Diabetes quality of care and outpatient utilization associated with electronic patient-provider messaging: a cross-sectional analysis. *Diabetes Care* 2009 Jul 14;32(7):1182-1187 [FREE Full text] [doi: [10.2337/dc08-1771](https://doi.org/10.2337/dc08-1771)] [Medline: [19366959](https://pubmed.ncbi.nlm.nih.gov/19366959/)]
13. North F, Luhman KE, Mallmann EA, Mallmann TJ, Tulledge-Scheitel SM, North EJ, et al. A Retrospective Analysis of Provider-to-Patient Secure Messages: How Much Are They Increasing, Who Is Doing the Work, and Is the Work Happening After Hours? *JMIR Med Inform* 2020 Jul 08;8(7):e16521 [FREE Full text] [doi: [10.2196/16521](https://doi.org/10.2196/16521)] [Medline: [32673238](https://pubmed.ncbi.nlm.nih.gov/32673238/)]
14. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. *Health Aff (Millwood)* 2010 Jul 20;29(7):1370-1375. [doi: [10.1377/hlthaff.2010.0048](https://doi.org/10.1377/hlthaff.2010.0048)] [Medline: [20606190](https://pubmed.ncbi.nlm.nih.gov/20606190/)]
15. Ralston J, Cook A, Anderson M, Catz S, Fishman P, Carlson J, et al. Home blood pressure monitoring, secure electronic messaging and medication intensification for improving hypertension control. *Appl Clin Inform* 2017 Dec 20;05(01):232-248. [doi: [10.4338/aci-2013-10-ra-0079](https://doi.org/10.4338/aci-2013-10-ra-0079)] [Medline: [29444256](https://pubmed.ncbi.nlm.nih.gov/29444256/)]
16. Price-Haywood EG, Luo Q, Monlezun D. Dose effect of patient-care team communication via secure portal messaging on glucose and blood pressure control. *J Am Med Inform Assoc* 2018 Jun 01;25(6):702-708. [doi: [10.1093/jamia/ocx161](https://doi.org/10.1093/jamia/ocx161)] [Medline: [29444256](https://pubmed.ncbi.nlm.nih.gov/29444256/)]
17. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures. *J Med Internet Res* 2016 Jul 01;18(7):e179 [FREE Full text] [doi: [10.2196/jmir.5663](https://doi.org/10.2196/jmir.5663)] [Medline: [27369696](https://pubmed.ncbi.nlm.nih.gov/27369696/)]
18. Harris LT, Koepsell TD, Haneuse SJ, Martin DP, Ralston JD. Glycemic control associated with secure patient-provider messaging within a shared electronic medical record: a longitudinal analysis. *Diabetes Care* 2013 Sep;36(9):2726-2733 [FREE Full text] [doi: [10.2337/dc12-2003](https://doi.org/10.2337/dc12-2003)] [Medline: [23628618](https://pubmed.ncbi.nlm.nih.gov/23628618/)]
19. Chung S, Panattoni L, Chi J, Palaniappan L. Can Secure Patient-Provider Messaging Improve Diabetes Care? *Diabetes Care* 2017 Oct;40(10):1342-1348. [doi: [10.2337/dc17-0140](https://doi.org/10.2337/dc17-0140)] [Medline: [28807977](https://pubmed.ncbi.nlm.nih.gov/28807977/)]
20. Greenwood DA, Hankins AI, Parise CA, Spier V, Olveda J, Buss KA. A Comparison of In-person, Telephone, and Secure Messaging for Type 2 Diabetes Self-Management Support. *Diabetes Educ* 2014 Jul;40(4):516-525. [doi: [10.1177/0145721714531337](https://doi.org/10.1177/0145721714531337)] [Medline: [24742540](https://pubmed.ncbi.nlm.nih.gov/24742540/)]
21. Heisey-Grove DM. What's in a message? The impact of patient-clinician email message content on patient health outcomes and healthcare utilization (Doctor of Philosophy Dissertation). 2020 Feb 19. URL: <https://scholarscompass.vcu.edu/etd/6141/> [accessed 2020-10-04]
22. Heisey-Grove DM, DeShazo JP. Look Who's Talking: Application of a Theory-Based Taxonomy to Patient-Clinician E-mail Messages. *Telemed J E Health* 2020 Feb 19:e1 (forthcoming). [doi: [10.1089/tmj.2019.0192](https://doi.org/10.1089/tmj.2019.0192)] [Medline: [32074474](https://pubmed.ncbi.nlm.nih.gov/32074474/)]
23. Mishel MH. Uncertainty in illness. *Image J Nurs Sch* 1988;20(4):225-232. [doi: [10.1111/j.1547-5069.1988.tb00082.x](https://doi.org/10.1111/j.1547-5069.1988.tb00082.x)] [Medline: [3203947](https://pubmed.ncbi.nlm.nih.gov/3203947/)]
24. Kravitz RL, Bell RA, Franz CE. A taxonomy of requests by patients (TORP): a new system for understanding clinical negotiation in office practice. *J Fam Pract* 1999 Nov;48(11):872-878. [Medline: [10907624](https://pubmed.ncbi.nlm.nih.gov/10907624/)]
25. Epstein R, Street R. Patient-Centered Communication in Cancer Care: Promoting Healing and Reducing Suffering. Bethesda, MD: National Cancer Institute; 2007.
26. Walther JB. Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior* 2007 Sep 03;23(5):2538-2557. [doi: [10.1016/j.chb.2006.05.002](https://doi.org/10.1016/j.chb.2006.05.002)]
27. Marlow S, Lacerenza C, Iwig C. The Influence of Textual Cues on First Impressions of an Email Sender. *Business and Professional Communication Quarterly* 2017 Aug 17;81(2):149-166. [doi: [10.1177/2329490617723115](https://doi.org/10.1177/2329490617723115)]
28. High AC, Solomon DH. Communication Channel, Sex, and the Immediate and Longitudinal Outcomes of Verbal Person-centered Support. *Communication Monographs* 2014 Jul 07;81(4):439-468. [doi: [10.1080/03637751.2014.933245](https://doi.org/10.1080/03637751.2014.933245)]
29. Spottswood E, Walther J, Holmstrom A, Ellison N. Person-Centered Emotional Support and Gender Attributions in Computer-Mediated Communication. *Hum Commun Res* 2013 May 22;39(3):295-316 [FREE Full text] [doi: [10.1111/hcre.12006](https://doi.org/10.1111/hcre.12006)]
30. United States Department of Agriculture Economic Research Service. 2010 Rural-Urban Commuting Area Codes. Washington, DC: United States Department of Agriculture Economic Research Service; 2019.
31. Hibbard J, Stockard J, Mahoney E, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. *Health Serv Res* 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text]
32. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. *Health Aff (Millwood)* 2015 Mar;34(3):431-437. [doi: [10.1377/hlthaff.2014.0452](https://doi.org/10.1377/hlthaff.2014.0452)] [Medline: [25732493](https://pubmed.ncbi.nlm.nih.gov/25732493/)]
33. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. *Health Aff (Millwood)* 2013 Feb;32(2):207-214. [doi: [10.1377/hlthaff.2012.1061](https://doi.org/10.1377/hlthaff.2012.1061)] [Medline: [23381511](https://pubmed.ncbi.nlm.nih.gov/23381511/)]
34. Centers for Disease Control and Prevention. Health, United States Spotlight: Racial and Ethnic Disparities in Heart Disease. 2019. URL: [https://www.cdc.gov/nchs/hus/spotlight/HeartDiseaseSpotlight\\_2019\\_0404.pdf](https://www.cdc.gov/nchs/hus/spotlight/HeartDiseaseSpotlight_2019_0404.pdf) [accessed 2020-10-04]
35. CMS OMH, NORC. Racial and Ethnic Disparities in Diabetes Prevalence, Self-Management, and Health Outcomes Among Medicare Beneficiaries (CMS OMH Data Highlight No. 6). Baltimore, MD: CMS OMH; 2017.



36. Maina IW, Belton TD, Ginzberg S, Singh A, Johnson TJ. A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. *Soc Sci Med* 2018 Feb 01;199(1):219-229. [doi: [10.1016/j.socscimed.2017.05.009](https://doi.org/10.1016/j.socscimed.2017.05.009)] [Medline: [28532892](https://pubmed.ncbi.nlm.nih.gov/28532892/)]
37. Shen MJ, Peterson EB, Costas-Muñiz EB, Hernandez MH, Jewell ST, Matsoukas K, et al. The Effects of Race and Racial Concordance on Patient-Physician Communication: A Systematic Review of the Literature. *J Racial Ethn Health Disparities* 2018 Feb;5(1):117-140 [FREE Full text] [doi: [10.1007/s40615-017-0350-4](https://doi.org/10.1007/s40615-017-0350-4)] [Medline: [28275996](https://pubmed.ncbi.nlm.nih.gov/28275996/)]
38. Rask KJ, Ziemer DC, Kohler SA, Hawley JN, Arinde FJ, Barnes CS. Patient activation is associated with healthy behaviors and ease in managing diabetes in an indigent population. *Diabetes Educ* 2009;35(4):622-630. [doi: [10.1177/0145721709335004](https://doi.org/10.1177/0145721709335004)] [Medline: [19419972](https://pubmed.ncbi.nlm.nih.gov/19419972/)]
39. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. *J Gen Intern Med* 2012 May;27(5):520-526 [FREE Full text] [doi: [10.1007/s11606-011-1931-2](https://doi.org/10.1007/s11606-011-1931-2)] [Medline: [22127797](https://pubmed.ncbi.nlm.nih.gov/22127797/)]
40. Heyworth L, Clark J, Marcello TB, Paquin AM, Stewart M, Archambeault C, et al. Aligning medication reconciliation and secure messaging: qualitative study of primary care providers' perspectives. *J Med Internet Res* 2013 Dec 02;15(12):e264 [FREE Full text] [doi: [10.2196/jmir.2793](https://doi.org/10.2196/jmir.2793)] [Medline: [24297865](https://pubmed.ncbi.nlm.nih.gov/24297865/)]
41. Ozkaynak M, Johnson S, Shimada S, Petrakis BA, Tulu B, Archambeault C, et al. Examining the Multi-level Fit between Work and Technology in a Secure Messaging Implementation. *AMIA Annu Symp Proc* 2014;2014:954-962 [FREE Full text] [Medline: [25954403](https://pubmed.ncbi.nlm.nih.gov/25954403/)]
42. Wooldridge A, Carayon P, Hoonakker P, Musa A, Bain P. Technology-Mediated Communication between Patients and Primary Care Clinicians and Staff. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 2016 Sep 15;60(1):556-560. [doi: [10.1177/1541931213601128](https://doi.org/10.1177/1541931213601128)]
43. Lafata JE, Morris HL, Dobie E, Heisler M, Werner RM, Dumenci L. Patient-reported use of collaborative goal setting and glycemic control among patients with diabetes. *Patient Educ Couns* 2013 Jul;92(1):94-99 [FREE Full text] [doi: [10.1016/j.pec.2013.01.016](https://doi.org/10.1016/j.pec.2013.01.016)] [Medline: [23433777](https://pubmed.ncbi.nlm.nih.gov/23433777/)]
44. Street RLJ. Information-giving in medical consultations: the influence of patients' communicative styles and personal characteristics. *Soc Sci Med* 1991;32(5):541-548. [doi: [10.1016/0277-9536\(91\)90288-n](https://doi.org/10.1016/0277-9536(91)90288-n)] [Medline: [2017721](https://pubmed.ncbi.nlm.nih.gov/2017721/)]
45. Song L, Weaver MA, Chen RC, Bensen JT, Fontham E, Mohler JL, et al. Associations between patient-provider communication and socio-cultural factors in prostate cancer patients: a cross-sectional evaluation of racial differences. *Patient Educ Couns* 2014 Dec;97(3):339-346 [FREE Full text] [doi: [10.1016/j.pec.2014.08.019](https://doi.org/10.1016/j.pec.2014.08.019)] [Medline: [25224313](https://pubmed.ncbi.nlm.nih.gov/25224313/)]
46. Street RLJ, Haidet P. How well do doctors know their patients? Factors affecting physician understanding of patients' health beliefs. *J Gen Intern Med* 2011 Jan;26(1):21-27 [FREE Full text] [doi: [10.1007/s11606-010-1453-3](https://doi.org/10.1007/s11606-010-1453-3)] [Medline: [20652759](https://pubmed.ncbi.nlm.nih.gov/20652759/)]
47. Sohl SJ, Borowski LA, Kent EE, Smith AW, Oakley-Girvan I, Rothman RL, et al. Cancer survivors' disclosure of complementary health approaches to physicians: the role of patient-centered communication. *Cancer* 2015 Mar 15;121(6):900-907 [FREE Full text] [doi: [10.1002/cncr.29138](https://doi.org/10.1002/cncr.29138)] [Medline: [25387799](https://pubmed.ncbi.nlm.nih.gov/25387799/)]
48. Street RL, Gordon HS, Ward MM, Krupat E, Kravitz RL. Patient participation in medical consultations: why some patients are more involved than others. *Med Care* 2005 Oct;43(10):960-969. [doi: [10.1097/01.mlr.0000178172.40344.70](https://doi.org/10.1097/01.mlr.0000178172.40344.70)] [Medline: [16166865](https://pubmed.ncbi.nlm.nih.gov/16166865/)]
49. Lanham HJ, Leykum LK, Pugh JA. Examining the Complexity of Patient-Outpatient Care Team Secure Message Communication: Qualitative Analysis. *J Med Internet Res* 2018 Jul 11;20(7):e218 [FREE Full text] [doi: [10.2196/jmir.9269](https://doi.org/10.2196/jmir.9269)] [Medline: [29997107](https://pubmed.ncbi.nlm.nih.gov/29997107/)]
50. Skolasky RL, Green AF, Scharfstein D, Boulton C, Reider L, Wegener ST. Psychometric properties of the patient activation measure among multimorbid older adults. *Health Serv Res* 2011 Apr;46(2):457-478 [FREE Full text] [doi: [10.1111/j.1475-6773.2010.01210.x](https://doi.org/10.1111/j.1475-6773.2010.01210.x)] [Medline: [21091470](https://pubmed.ncbi.nlm.nih.gov/21091470/)]
51. Alexander JA, Hearld LR, Mittler JN, Harvey J. Patient-physician role relationships and patient activation among individuals with chronic illness. *Health Serv Res* 2012 Jun;47(3 Pt 1):1201-1223 [FREE Full text] [doi: [10.1111/j.1475-6773.2011.01354.x](https://doi.org/10.1111/j.1475-6773.2011.01354.x)] [Medline: [22098418](https://pubmed.ncbi.nlm.nih.gov/22098418/)]

## Abbreviations

**DBP:** diastolic blood pressure

**SBP:** systolic blood pressure



*Edited by G Eysenbach; submitted 19.04.20; peer-reviewed by J Alpert, J Haun, A Linsky, E Bellei; comments to author 17.06.20; revised version received 05.09.20; accepted 07.09.20; published 29.10.20.*

*Please cite as:*

Heisey-Grove DM, McClelland LE, Rathert C, Tartaglia A, Jackson K, DeShazo JP

*Associations Between Patient Health Outcomes and Secure Message Content Exchanged Between Patients and Clinicians: Retrospective Cohort Study*

*J Med Internet Res* 2020;22(10):e19477

URL: <http://www.jmir.org/2020/10/e19477/>

doi: [10.2196/19477](https://doi.org/10.2196/19477)

PMID: [33118938](https://pubmed.ncbi.nlm.nih.gov/33118938/)

©Dawn M Heisey-Grove, Laura E McClelland, Cheryl Rathert, Alexander Tartaglia, Kevin Jackson, Jonathan P DeShazo. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Review

# The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review

Madison Milne-Ives<sup>1</sup>, BAS, MSc; Caroline de Cock<sup>1</sup>, BSc, MSc; Ernest Lim<sup>2,3</sup>, BSc, MBBS; Melissa Harper Shehadeh<sup>4</sup>, BSc, MSc, PhD; Nick de Pennington<sup>3,5</sup>, MA, BM BCh; Guy Mole<sup>3,5</sup>, BSc, MBBS, MSc; Eduardo Normando<sup>2</sup>, MD, PhD; Edward Meinert<sup>1,6,7</sup>, MA, MSc, MBA, MPA, PhD

<sup>1</sup>Digitally Enabled Preventative Health Research Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom

<sup>2</sup>Imperial College Healthcare NHS Trust, London, United Kingdom

<sup>3</sup>Ufonia Limited, Oxford, United Kingdom

<sup>4</sup>Institute of Global Health, University of Geneva, Geneva, Switzerland

<sup>5</sup>Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

<sup>6</sup>Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

<sup>7</sup>Centre for Health Technology, University of Plymouth, Plymouth, United Kingdom

**Corresponding Author:**

Edward Meinert, MA, MSc, MBA, MPA, PhD

Centre for Health Technology

University of Plymouth

8 Kirkby Place

Room 2

Plymouth, PL4 6DT

United Kingdom

Phone: 44 7824446808

Email: [edward.meinert@plymouth.ac.uk](mailto:edward.meinert@plymouth.ac.uk)

## Abstract

**Background:** The high demand for health care services and the growing capability of artificial intelligence have led to the development of conversational agents designed to support a variety of health-related activities, including behavior change, treatment support, health monitoring, training, triage, and screening support. Automation of these tasks could free clinicians to focus on more complex work and increase the accessibility to health care services for the public. An overarching assessment of the acceptability, usability, and effectiveness of these agents in health care is needed to collate the evidence so that future development can target areas for improvement and potential for sustainable adoption.

**Objective:** This systematic review aims to assess the effectiveness and usability of conversational agents in health care and identify the elements that users like and dislike to inform future research and development of these agents.

**Methods:** PubMed, Medline (Ovid), EMBASE (Excerpta Medica dataBASE), CINAHL (Cumulative Index to Nursing and Allied Health Literature), Web of Science, and the Association for Computing Machinery Digital Library were systematically searched for articles published since 2008 that evaluated unconstrained natural language processing conversational agents used in health care. EndNote (version X9, Clarivate Analytics) reference management software was used for initial screening, and full-text screening was conducted by 1 reviewer. Data were extracted, and the risk of bias was assessed by one reviewer and validated by another.

**Results:** A total of 31 studies were selected and included a variety of conversational agents, including 14 chatbots (2 of which were voice chatbots), 6 embodied conversational agents (3 of which were interactive voice response calls, virtual patients, and speech recognition screening systems), 1 contextual question-answering agent, and 1 voice recognition triage system. Overall, the evidence reported was mostly positive or mixed. Usability and satisfaction performed well (27/30 and 26/31), and positive or mixed effectiveness was found in three-quarters of the studies (23/30). However, there were several limitations of the agents highlighted in specific qualitative feedback.

**Conclusions:** The studies generally reported positive or mixed evidence for the effectiveness, usability, and satisfactoriness of the conversational agents investigated, but qualitative user perceptions were more mixed. The quality of many of the studies was limited, and improved study design and reporting are necessary to more accurately evaluate the usefulness of the agents in health

care and identify key areas for improvement. Further research should also analyze the cost-effectiveness, privacy, and security of the agents.

**International Registered Report Identifier (IRRID):** RR2-10.2196/16934

(*J Med Internet Res* 2020;22(10):e20346) doi:[10.2196/20346](https://doi.org/10.2196/20346)

## KEYWORDS

artificial intelligence; avatar; chatbot; conversational agent; digital health; intelligent assistant; speech recognition software; virtual assistant; virtual coach; virtual health care; virtual nursing; voice recognition software

## Introduction

### Background

Conversational agents are among the many digital technologies being introduced into the health sector to address current health care challenges, such as shortages of health care providers, which reduce the availability and accessibility of health care services [1-3]. Conversational agents use artificial intelligence (AI), including machine learning (a statistical means of training models with data so that they can make predictions based on a variety of features) and natural language processing (NLP; the ability to recognize and analyze verbal and written language) to interact with humans via speech, text, or other inputs and outputs on mobile, web-based, or audio-based platforms [1,4]. Many of these agents are designed to use NLP so that users can speak or write to the agent as they would to a human. The agent can then analyze the input and respond appropriately in a conversational manner [5].

Conversational agents first emerged as a tool in health care in 1966, with the development of a virtual psychotherapist (ELIZA) that could provide predetermined answers to text-based user input [6]. In the decades since, the capabilities of NLP have significantly progressed and aided the development of more advanced AI agents. Many different types of conversational agents that use NLP have been developed, including chatbots, embodied conversational agents (ECAs), and virtual patients, and are accessible by telephone, mobile phones, computers, and many other digital platforms [7-10]. The types of input that conversational agents can receive and interpret have also expanded, with some conversational agents capable of analyzing movements, such as gestures, facial expressions, and eye movements [11,12].

Conversational agents have been developed for many different aspects of the health sector to support health care professionals and the general public. Specific uses include screening for health conditions, triage, counseling, at-home health management support, and training for health care professionals [8,13-15]. With phone, mobile, and online platforms being widely accessible, conversational agents can support populations with limited access to health care or poor health literacy [16,17]. They also have the potential to be affordably scaled up to reach large proportions of a population [3]. Due to this accessibility, conversational agents are also a promising tool for the advancement of patient-centered care and can support users' involvement in the management of their own health [17,18]. Personalizable features have the potential to further improve usability and satisfaction, although more research is needed to

evaluate their effectiveness in achieving their stated health outcomes and reducing costs and to ensure that there are no negative consequences for decision making or privacy [10].

Despite the large body of research concerning the application of conversational agents in health care, most reviews have limited their focus to a particular health area, agent type, or function [10,19-22]. Although there are a few recent systematic reviews that have examined a more comprehensive scope, they have presented an overall synthesis of the body of knowledge. One review developed a taxonomy that described the architecture and functions of conversational agents in health care and the state of the field but did not evaluate the effectiveness, usability, or implications for users [5]. Another systematic review investigated the outcome measures of the studies of conversational agents but limited the inclusion criteria to agents that used natural language input and had been tested with human participants [2]. Additionally, their initial database searches only retrieved 1531 articles, which raises the concern that some relevant articles may have been overlooked [2]. Their search was updated in February 2018, but given the rapid pace of technological development, there is a need to provide an update and expansion to these previous systematic reviews.

For conversational agents to be successful in health care, it is crucial to understand the effectiveness of current agents in achieving their intended outcomes. However, it is just as important to understand how users feel about and relate to these agents because the adoption of new health technologies depends on user perceptions (eg, whether they trust the technology, find it easy to use, and feel that privacy and data security are respected) [23]. User-identified problems will need to be addressed if conversational agents are to have a significant impact on health care, because their impact depends on people being willing to use them and preferring to use them over alternatives. The information gathered in this review identifies the current issues with conversational agents that need to be overcome and can be used to help determine which elements of the agents are most likely to be successful and useful in various aspects of health care. As conversational agents are often touted as having the potential to reduce the burden on health care resources, evaluations of the implications of the agents for improved health care provision and reduced resource demand also need to be assessed.

### Objectives

The primary objectives of this review are to describe the scope of conversational agents currently being used for health care activities (by patients, health care providers, or the general public), examine the user perceptions of these agents, and

evaluate their effectiveness. We developed 3 main research questions to address these objectives. First, are the conversational agents investigated effective at achieving their intended health-related outcomes, and does the effectiveness vary depending on the type of agent? Second, how do users rate the usability and satisfactoriness of the conversational agents, and what specific elements of the agents do they like and dislike? Finally, what are the current limitations and gaps in the utility of conversational agents in health care? These objectives build on previous systematic reviews while widening the scope of included studies to update the body of knowledge on conversational agents in health care and to inform future research and development.

## Methods

### Database Search

The full methods for this review have been published in detail in a systematic review protocol [24]. The population, intervention, comparison, and outcome framework [25] was used to develop the search strategy, which was implemented following the PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analyses Protocols) checklist [26]. No study design filter was used; any type of study was eligible for inclusion. The search strategy was finalized and tailored to different databases in consultation with a medical librarian. PubMed, Medline (Ovid), EMBASE (Excerpta Medica dataBASE), CINAHL (Cumulative Index of Nursing and Allied Health Literature), Web of Science, and the Association for Computing Machinery Digital Library databases were searched. The search terms were grouped into 3 themes (conversational agents, health application, and outcome assessment) to capture all studies that fit the key inclusion criteria: evaluating conversational agents used in health care. These themes were subsequently searched with the structure: conversational agent (MeSH OR Keywords) AND health application (MeSH OR Keywords) AND outcome assessment (MeSH OR Keywords). The full search strategy can be found in [Multimedia Appendix 1](#). The search was completed on November 29, 2019.

### Inclusion and Exclusion Criteria

This systematic review aimed to assess conversational agents designed for health care purposes. Studies that evaluated at least 1 conversational agent were included. Studies targeting any population group, geographical location, and mental or physical health-related function (eg, screening, education, training, and self-management) were included. These broad inclusion criteria were established to enable an assessment of a wide range of applications of conversational agents. There were no restrictions on study type, as long as a conversational agent was evaluated, and intervention and observational studies such as cross-sectional surveys, cohort studies, and qualitative studies were included. Intervention studies were not required to have a specific comparator or any comparator.

During the screening process, studies of conversational agents that were not capable of interacting with human users via unconstrained NLP were excluded. These included conversational agents that only allowed users to select from predefined options or agents with prerecorded responses that

did not adapt to subsequent user responses. The basis for this exclusion is that, without the capability of using NLP, computational methods and technologies are rudimentary and do not advance the aims of AI for autonomous computational agents. As many studies did not explicitly state whether the investigated agent was capable of NLP, a description in the paper of the conversational agent allowing free-text or free-speech input was used as an indicator for NLP, and these studies were included. Studies that did not report the architecture of the agent were excluded.

Due to the number of conversational agents in development and/or those that did not progress to the evaluation stages of development, studies that were solely descriptive were excluded. Furthermore, because of the pace at which conversational agents have developed over recent decades, studies were limited to those published during or after 2008. In 2008, the first iPhone was released, and it marks an increase in the prevalence and capabilities of digital technology. Only studies published in English were included to ensure accurate interpretation by the authors. Conference publications were also excluded from the review of peer-reviewed literature.

### Outcomes

The primary objective of this review was to provide an overview of the use of NLP conversational agents in health care. Therefore, the primary outcomes evaluated were the effectiveness of conversational agents in achieving their intended health-related outcomes and user perceptions of the agents (including but not limited to acceptability, usability, satisfaction, and specific qualitative feedback). Secondary outcomes included improvement in health care provision and resource implications for the health care system.

### Screening and Study Selection

All studies retrieved from the databases were stored in the reference management software EndNote (version X9, Clarivate Analytics), which automatically eliminated duplicates. Due to time constraints, the EndNote search function was used to extract relevant studies before the screening of the citations against the inclusion and exclusion criteria by 2 independent reviewers. Where duplicates or publications from the same study were identified, the more recent publication or the one with the most detail was selected for inclusion in the review. All disagreements were discussed, and if a consensus was not reached, a third reviewer was consulted. Full EndNote search details are shown in [Multimedia Appendix 2](#).

The full texts of the articles that met the inclusion criteria were screened by one of the reviewers. Of the screened articles deemed eligible for inclusion, 58 were conference or meeting abstracts and did not have full texts available; therefore, they were excluded. This highlights the early developmental stages of many of these agents.

### Data Extraction

Data were extracted by 1 reviewer, and key data points from the studies, specified in the protocol and identified on further study of the publications, were recorded in a spreadsheet and validated by a second reviewer. The data extraction form was

based on the minimum requirements recommended by the Cochrane Handbook for Systematic Reviews [27]. The types

of data extracted from the studies are shown in Table 1.

**Table 1.** Data extracted from the studies.

Article information	Data extracted
General study information	<ul style="list-style-type: none"> <li>Title of publication</li> <li>Year of publication</li> <li>Authors</li> </ul>
Study characteristics	<ul style="list-style-type: none"> <li>Study design</li> <li>Country of study</li> <li>Study population</li> <li>Analyzed sample size</li> <li>Comparators</li> <li>Study duration</li> </ul>
Characteristics of the conversational agents	<ul style="list-style-type: none"> <li>Name of conversational agents</li> <li>Architecture</li> <li>Device or platform on which agent is accessed</li> <li>Intended user</li> <li>Primary purpose</li> </ul>
Intended outcomes of the conversational agents	<ul style="list-style-type: none"> <li>Health objective (general)</li> <li>Health objective (specific)</li> </ul>
Evaluation	<ul style="list-style-type: none"> <li>Effectiveness in achieving intended purpose</li> <li>Health literacy</li> <li>Improvement in health care provision</li> <li>Health care resource implications</li> <li>Usability</li> <li>Acceptability or satisfaction</li> <li>User perceptions qualitative feedback</li> <li>Conclusions</li> <li>Implications for future study</li> </ul>

### Risk-of-Bias and Quality Assessment

All quality assessments were conducted by 2 independent reviewers, with disagreements resolved by consensus. If this was not possible, the opinion of a third reviewer was sought. As there was a wide variety of study designs, the study types were classified by 1 reviewer and validated by a second

reviewer, with disagreements being resolved by discussion with a third reviewer. As the broad inclusion criteria were intended to capture all relevant studies, a few of the included studies used implementation models for artificial AI research that were beyond the scope of classic public health design methods. This resulted in some study designs being categorized as *other*.



The Cochrane Collaboration risk-of-bias tool was used to evaluate the risk of bias in randomized controlled trials (RCTs) [28]. The CASP (Critical Appraisal Skills Programme) tools for cohort and qualitative studies were used for the respective studies [29], and the Appraisal tool for Cross-Sectional Studies (AXIS) tool was used to assess the quality of cross-sectional survey studies [30]. Studies that were coded as *other* design types were also assessed using the AXIS tool, which was deemed to be the most rigorous and appropriate tool because it systematically evaluates elements of the introduction, methods, results, and discussion sections, and is not limited to the RCT-specific questions used in the risk-of-bias tool.

The results of the Cochrane Collaboration risk-of-bias tool were summarized using RevMan 5.3. CASP and AXIS scores were calculated using yes=1, no=0, and cannot tell or do not know=0 for each question. The scores for each question were summed to provide a score for each study, which were averaged according to study type and are presented in the results.

### Data Analysis and Synthesis

Due to the variability in populations, interventions, outcomes, and study designs, a meta-analysis of the studies was not possible. Therefore, we report a structured analysis of the findings to draw conclusions about the effectiveness and user perceptions of conversational agents in health care. For the purpose of this review, the agent was considered effective if there was a statistically significant ( $P<.05$ ) improvement in a given outcome as compared with a comparator or control, or over time. If no significance was reported or the difference was nonsignificant or significantly worse between groups or over time, the agent was considered to have no significant evidence supporting it. Limitations and future directions for research were also summarized.

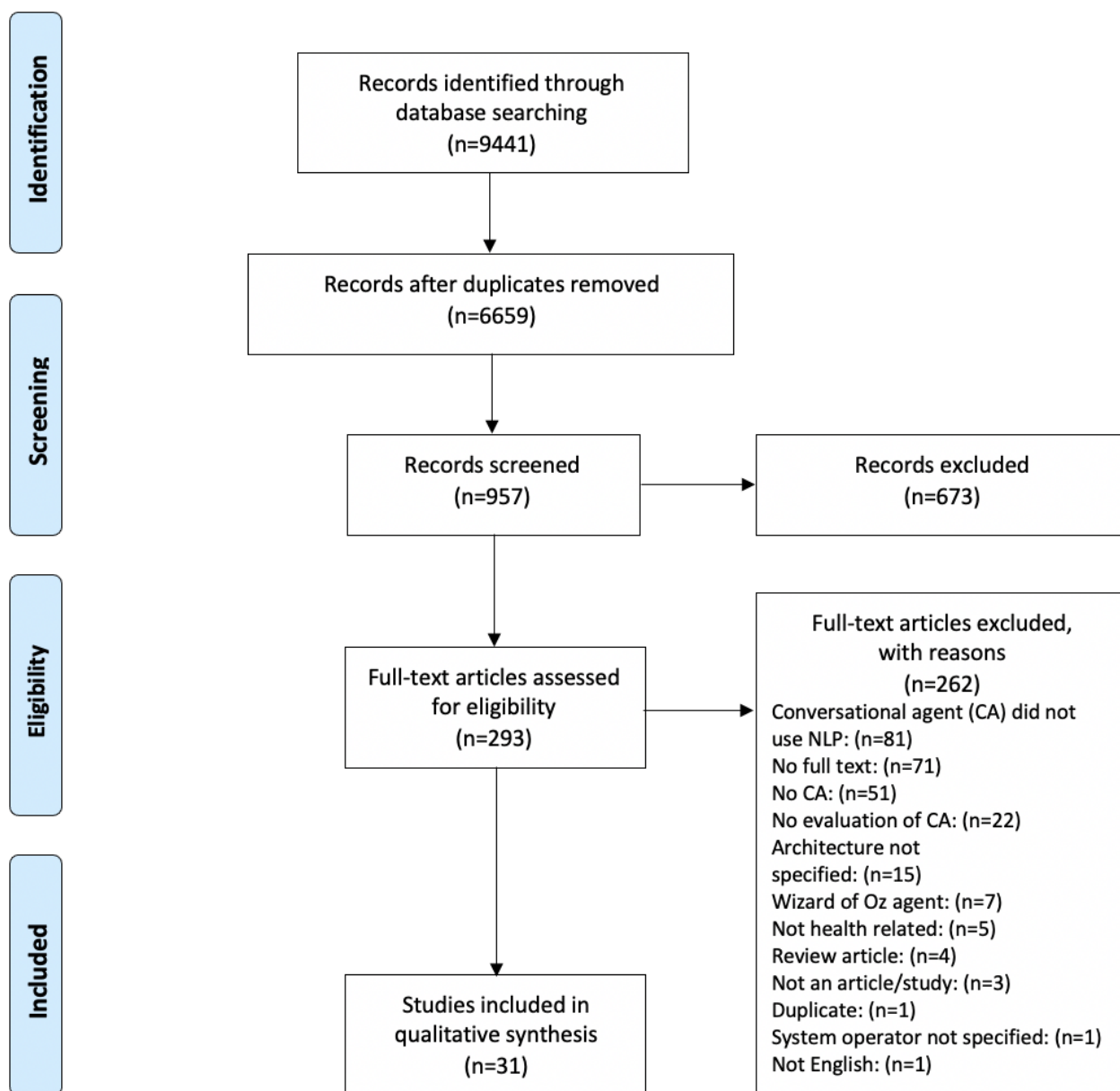
The synthesis framework for the assessment of health information technology (SF/HIT) was used to structure the evaluation of the studies because it included a whole system set of outcome variables [31]. These included effectiveness, satisfaction, and perceived ease of use or usefulness, among others. In accordance with the framework, evidence for each of the outcome variables was coded as *positive or mixed* or *neutral or negative*. If the study did not address the outcome in question, it was coded as *neutral or negative*.

Finally, where qualitative user feedback was reported by the studies, it was examined to extract common themes by extracting the sections of the original text that discussed the qualitative perceptions, reducing them to key themes, and then comparing those key themes across the different studies.

## Results

### Included Studies

Overall, 9441 studies were retrieved from the 6 databases, of which 2782 were duplicates. The reference management software EndNote was used for initial screening, with keywords based on the original search categories used to exclude studies that did not meet the criteria. After 6 passes, 957 citations remained for abstract screening. The primary reason for exclusion at the screening stage was that the study did not include an interactive, responsive conversational agent ( $n=470$ ), was a review paper ( $n=65$ ), was not health-related ( $n=48$ ), or did not report any evaluation of the conversational agent ( $n=46$ ). Of these 957 citations, 293 were selected for full-text review. In the final review, 31 papers were included. The reasons for exclusion after full-text review are detailed in Figure 1, with the most common reason being that the conversational agent did not use NLP ( $n=81$ ), the full text was not available ( $n=71$ ), or there was no conversational agent in the study ( $n=51$ ).

**Figure 1.** Preferred Reporting Items for Systematic Review and Meta-Analyses flow diagram. NLP: natural language processing.

## Study Characteristics

The characteristics of the 31 included studies are summarized in [Multimedia Appendix 3](#) [8,9,12-15,32-56]. Of these studies, 45% (14/31) evaluated conversational agents that had some type of audio or speech element. Of the agents, 45% (14/31) were chatbots (including 2 voice chatbots and 1 chatbot that also used a wizard), 19% (6/31) were ECAs (including 1 virtual doctor), and 10% (3/31) were interactive voice response (IVR) phone calls, virtual patients, and speech recognition screening systems. The final 2 comprised a contextual question-answering agent and a voice recognition triage system. In the 26 studies that reported the device that their conversational agent was used on; 35% (9/26) used computers, 27% (7/26) used web-based apps, 23% (6/26) used mobile phone apps, 15% (4/26) used telephone calls; 1 study used a tablet (the percentages do not add up to 100% because one agent could be used on a computer and also the telephone).

There were a wide variety of areas of health care targeted by the conversational agents of the included studies. The largest proportion of them (12/31, 39%) addressed mental health issues [13,32-42], with 19% (6/31) providing some form of clinical decision or triage support [8,12,40,42-44] and treatment support (including encouraging users to get screened) [9,45-49], 10% (3/31) being used to support training of health care students [15,41,50] and the screening or diagnosis of users [14,38,51], 7% (2/31) targeting physical health [52,53] and layperson medical education [54,55]; 1 agent was designed to help monitor users' speech [56]. The percentages do not add up to 100% because some of the studies that addressed mental health also fit into one of the other categories.

The study designs also varied widely, with 29% (9/31) using cross-sectional designs, 26% (8/31) using RCTs, 23% (7/31) using qualitative methods, 19% (6/31) using cohort studies, and 1 using a cluster crossover design. The full data extraction table is available in [Multimedia Appendix 4](#) [8,9,12-15,32-56].

### Overall Evaluation of Conversational Agents

Overall, about three-quarters of the studies (22/30, 73%) reported positive or mixed results for most of the outcomes. A total of 8 studies were coded as reporting positive or mixed evidence for 10 or more of the 11 outcomes specified in the SF/HIT; the analysis for this review was limited to the interpretation of impact as reported by study authors to reflect evaluation outcomes. Excluding 1 study, which was an acceptability study only and did not assess the other outcomes, the average number of outcomes that were coded as *positive or mixed* was 67% (7.4/11, SD 2.5). However, the number of outcomes met per study ranged from 1/11 to 11/11 (9-100%). Perceived ease of use or usefulness (27/30, 90%), the process

of service delivery or performance (26/30, 87%), appropriateness (24/30, 80%), and satisfaction (26/31, 84%) were the outcomes that had the most support from the studies. Just over three-quarters (23/30, 77%) of the studies also reported positive or mixed evidence of effectiveness.

However, very few studies discussed the cost-effectiveness (5/30, 17%, coded as *positive or mixed*) or safety, privacy, and security (14/30, 47%, coded as *positive or mixed*) outcomes for the agents being evaluated. About a quarter of studies (8/30, 27%) had neither positive nor mixed reported evidence for more than half of the SF/HIT outcomes. The evaluation of the SF/HIT outcomes is summarized in [Table 2 \[31\]](#).

**Table 2.** Summary of the studies based on the evaluation outcomes from the synthesis framework for the assessment of health information technology<sup>a</sup>.

First author (reference)	Preventive care	Adherence or attendance	Efficiency	Perceived ease of use or usefulness	Effectiveness	Performance	Safety or privacy or security	Acceptability	Cost-effectiveness	Appropriateness	Satisfaction	n (%)
Adams [9]	1	1	1	1	1	1	1	1	0	1	1	10 (91)
Bibault [46]	1	1	1	1	1	1	1	1	0	1	1	10 (91)
Borja-Harta [50]	0	1	1	1	1	1	1	0	0	1	0	7 (64)
Cameron [32]	0	0	1	1	0	1	0	1	0	0	1	5 (45)
Chaix [45]	1	0	1	1	1	1	1	0	0	1	1	8 (73)
Chang [8]	0	1	0	1	1	0	1	1	0	1	1	7 (64)
Crutzen [54]	0	1	1	1	1	1	1	1	0	1	1	9 (82)
Dimeff [42]	1	0	1	1	1	1	1	1	1	1	1	10 (91)
Elmasri [33]	0	0	0	1	0	1	1	0	0	1	1	5 (45)
Fitzpatrick [13]	1	1	1	1	1	1	1	1	0	1	1	10 (91)
Friederichs [53]	0	0	0	1	0	1	0	1	0	0	1	4 (36)
Fulmer [34]	1	1	0	0	1	1	1	0	0	0	1	6 (55)
Galescu [52]	0	0	1	1	0	1	0	0	0	0	0	3 (27)
Ghosh [44]	1	1	1	1	1	1	0	1	0	1	1	9 (82)
Havik [14]	1	1	1	1	1	1	0	1	1	1	1	10 (91)
Heyworth [47]	0	1	1	1	1	1	1	1	0	1	0	8 (73)
Hudlicka [35]	1	1	1	1	1	1	1	1	1	1	1	11 (100)
Inkster [36]	1	1	1	1	1	1	0	1	0	1	1	9 (82)
Ireland [56]											1	1 (100)
Isaza-Restrepo [15]	1	1	1	1	1	1	0	1	1	1	1	10 (91)
Ly [37]	0	1	0	1	0	1	0	0	0	1	1	5 (45)
Nakagawa [12]	1	0	1	1	1	1	0	0	0	1	1	7 (64)
Philip (2014) [51]	1	1	1	1	1	1	1	1	0	1	1	10 (91)
Philip (2017) [38]	1	1	1	1	1	1	0	1	0	1	1	9 (82)
Rhee [48]	1	1	1	1	1	1	0	1	0	1	1	9 (82)
Simon [49]	0	1	0	1	0	1	1	1	0	1	1	7 (64)
Spänig [43]	0	0	1	0	1	1	0	1	0	1	1	6 (55)
Washburn [41]	1	0	0	1	1	1	0	0	1	0	0	5 (45)
Wong [55]	0	0	0	1	0	0	0	0	0	0	0	1 (9)
Xu [40]	1	0	1	0	1	0	0	0	0	1	1	5 (45)
Yasavur [39]	0	1	1	1	1	0	0	1	0	1	1	7 (64)
n (%)	17 (57)	19 (63)	22 (73)	27 (90)	23 (77)	26 (87)	14 (47)	20 (67)	5 (17)	24 (80)	26 (84)	

<sup>a</sup>Positive or mixed results have been coded as 1, and neutral or negative results as 0.

When grouped by the agent's health care scope, studies of certain types of agents appear to do better than others (Table 3). Studies examining screening or diagnosis agents and treatment support agents had the highest average number of positive or mixed outcomes (mean 10, SD 0.6, and mean 9, SD 1.2, respectively). Treatment support agents had primary functions that included empowering patients to engage more

fully in clinical appointments, encouraging attending screenings for health care conditions, and supporting patient self-management. In contrast, mental health agents focused on addressing challenges related to depression, anxiety, and alcohol abuse, among others. However, given the small number of studies for each category of agents, these comparisons should be interpreted with caution.

**Table 3.** Summary of evaluation outcomes by the area of health care addressed by the conversational agent<sup>a</sup>.

Agent focus	Number of studies	Average number of outcomes coded positive or mixed, n (%)	Range of scores (SD)
Mental health [13,32-42]	12	7 (66)	5-11 (2.4)
Clinical decision or triage support [8,12,40,42-44]	6	7 (67)	5-10 (1.9)
Treatment support [9,45-49]	6	9 (79)	7-10 (1.2)
Health care training (students) [15,41,50]	3	7 (67)	5-10 (2.5)
Screening or diagnosis [14,38,51]	3	10 (88)	9-10 (0.6)
Health care education (laypeople) [54,55]	2	5 (45)	1-9 (5.7)
Physical health [52,53]	2	4 (32)	3-4 (0.7)

<sup>a</sup>The number of studies does not add up to 31 because some studies fit into 2 categories, and the study on monitoring speech was not included because it only addressed 1 of the 11 outcomes. The percentages associated with the average number of outcomes varied slightly because of rounding.

## Qualitative User Perceptions

A total of 18 of the 31 studies included more specific user feedback. The most frequently raised issue with conversational agents (9 studies) was poor understanding because of limited vocabulary, voice recognition accuracy, or error management of word inputs [13,32-37,41,52]. Related to this issue, as the conversational agents often had to ask questions more than once to be able to process the response, users in 3 studies noted disliking the repetitive conversations with the agents [13,36,37]. Both of these issues are key areas of improvement for future research and development of conversational agents because they represent limitations in the usability of the agents in a real-world context.

Feedback from users in 5 studies expressed a preference for interactivity, with users in 1 study noting that they liked the interactivity of the chatbot [35,37], and users in the other 4 studies expressed a desire for greater interactivity or relational skills in the conversational agent [14,32,34,53]. Similarly, users in 4 studies reported liking that the agent had a personality and/or showed empathy [13,32,34,42], whereas users in other studies reported disliking the lack of personal connection or had difficulty in empathizing with the agent [35,37,50] or reported disliking its limited conversation and responses [35,56].

Due to the wide variety of conversational agents, their aims and health care contexts, much of the qualitative user perception data concerned distinct aspects of the agents. However, several studies reported feedback concerned with customization or availability of feature options, with 2 studies commenting on it positively (eg, having both voice and touch modes to allow hands-free work and rapid data input on a triage system for nurses) [8,35], and 3 studies desiring more features and more control [33,37,48]. Additionally, users in 2 studies suggested that better integration of the agent with electronic health record (EHR) systems (for a virtual doctor [42]) or health care

providers (for an asthma self-management chatbot [48]) would be useful.

Other features of the agents that users reported liking were the reminders and assistance in forming routines [37,48] and that the agents provided accountability [13,34,48], facilitated learning [13,34,37], and were easy to learn and use [8,15]. In the included studies, 3 of the conversational agents were virtual patients, and users in all 3 studies reported liking that it provided a platform for risk-free learning because they were not practicing on real patients [15,41,50].

Several studies reported user feedback that was specific to that conversational agent. This included a preference for telephone IVR over web-based pediatric care guidance [9] and a simple avatar with a computer-generated voice over a more life-like agent with a recorded voice [42]. Users in 1 study reported liking that the agent initiated conversations [37]. There was opposite feedback in 2 studies about the format of the response, with users preferring preformatted options for one chatbot [36], whereas some users preferred the free-text responses for a diagnostic chatbot because it allowed them to provide contextual information. In contrast, others found it more difficult to know how to respond so the agent would understand [14].

Other agent-specific negative feedback was that the virtual doctor did not have the ability to go deep enough or provide access to other materials [42], that too much information was provided [13,33] or the interaction was too long [13], the use of nonverbal expressions by the avatar [35], and a lack of clarity regarding the aim of the chatbot [37]. Some students who used the virtual patients also reported that it was difficult to empathize [50] and that the agent did not sufficiently encompass real situational complexity [15]. The variety of specific feedback reports demonstrates the importance of examining usability for individual conversational agents and tailoring the design to the intended population. Although there were some preferences and



complaints that were frequently reported, much of the feedback was agent dependent. A summary of the thematic analysis is included in [Multimedia Appendix 5](#).

### Implications for Health Care Provision and Resources

Unfortunately, only a few of the studies discussed any improvement in health care provision or implications for resources; 2 of the studies that suggested improvement in health care provision were evaluating virtual patients [41,50], and students in 1 study reported significantly increased confidence in their clinical skills and ability to interview patients. Over 80% of users also reported that the agents helped them follow their treatment more effectively [45] and be more prepared for pediatric visits [9]. In a study of an ECA for sleep disorder screening, 65% of users reported thinking that the agent could provide significant assistance to physicians [51]. Regarding resource implications, the study of a preparatory IVR phone call before pediatric visits found that visit time was significantly reduced in the IVR group compared with the control group [9]. The use of an ECA to screen for depression [38] and a virtual doctor for suicidal patients in emergency departments (EDs) [42] were suggested by the authors to save physicians' time and reduce the costs associated with ED visits for suicidal ideation, but these outcomes were not evaluated. Similarly, another study suggested that mindfulness meditation could be of more use with more cost-effective training made available via a virtual coach [35].

Suggestions such as this, that conversational agents have the potential to improve health care provision, save health care providers' time, and reduce costs, were frequently made in the studies. However, as demonstrated above, very few studies quantified these claims and even fewer measured these outcomes with objective measures. This is a limitation of the studies as a whole. Although many were in the early stages of testing, claims about the potential value to the health care system in terms of time or money should be substantiated. However, as evidenced by the number of *neutral or negative* coding in the evaluation, many of the studies did not consider whole system implementation outcomes. It will be important for the future development of conversational agents to consider outcomes such as these from the beginning so that agents that are not only acceptable and usable but also provide value to the health care system can be built.

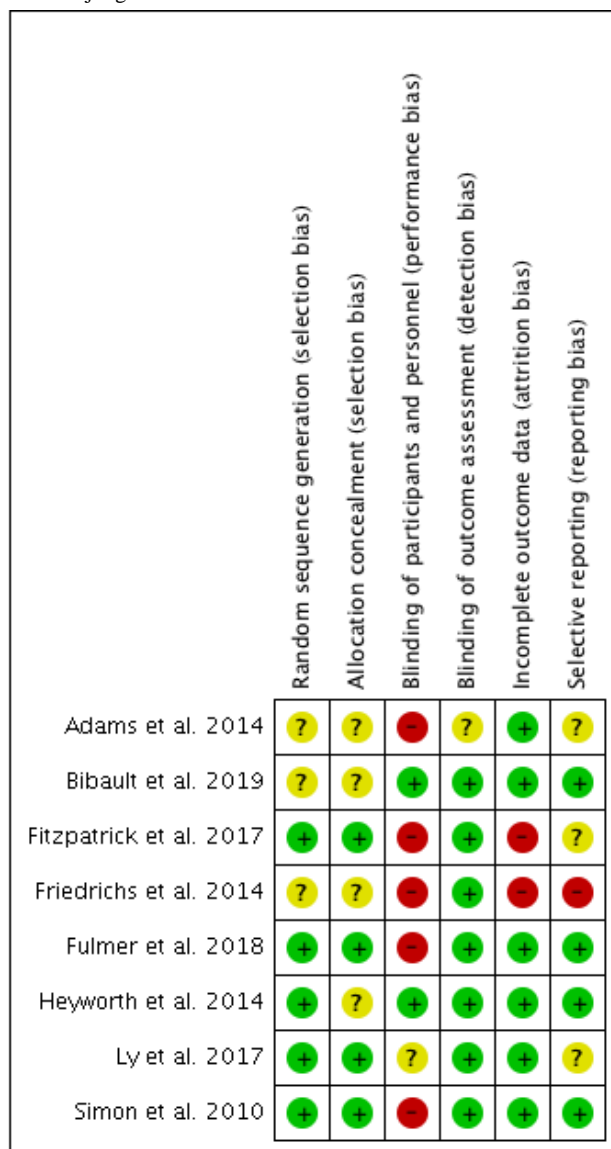
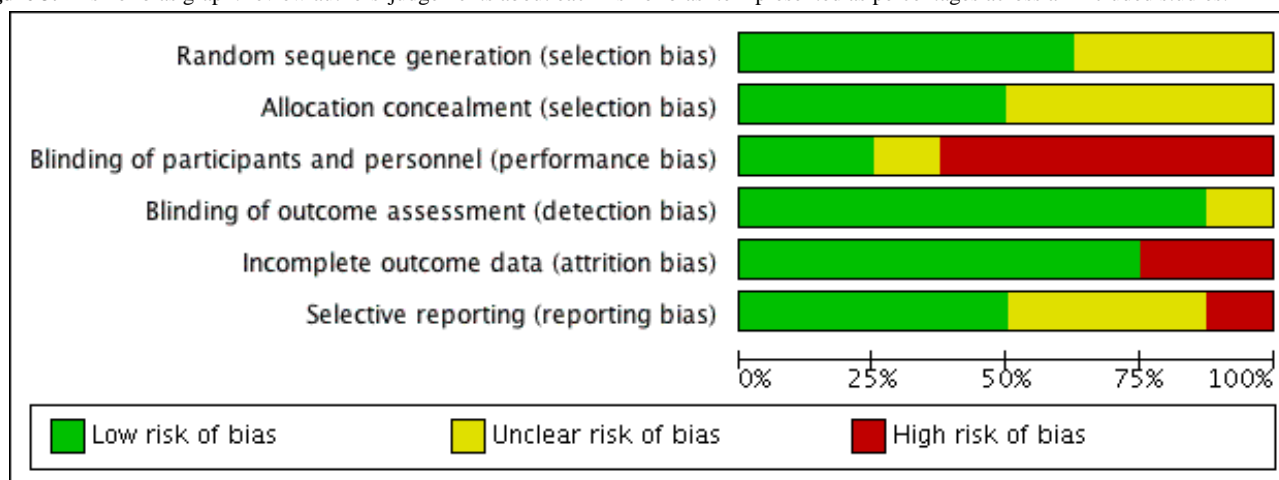
### Risk-of-Bias and Quality Assessments

There were a variety of study types included in this review; so several different quality assessment tools were used to assess the risk of bias and quality of the 31 included studies. A total of 6 studies could not be classified as RCTs, cohort, qualitative, or cross-sectional studies, and their study design was coded as *other* [12,39,40,44,52,55]. Most of these studies were papers describing the development and initial evaluation of conversational agents, and half of them did not have participants [40,44,55]. Initially, studies that did not have an explicit design were classified as qualitative or interpretative studies. However,

on further analysis, many of the studies did not fit the criteria for qualitative studies - evaluating subjective, thematic, and non-numerical data - because they evaluated performance metrics such as word error rates [52], accuracy [12,39,40,52,55], precision [44], and user experience quantified on Likert scales [39]. Therefore, these studies were coded as *other* and assessed using the AXIS tool for cross-sectional studies, which was deemed to provide the most systematic evaluation of the various elements of the studies [30]. The quality of these studies was assessed as best as possible; however, the judgments should be considered in the context of these limitations.

Overall, the quality of the studies was poor to moderate. On average, RCTs [9,13,34,37,46,47,49,53] and qualitative studies [41,48,56] evaluated were generally determined to have the highest quality and lowest risk of bias, with none of the other 3 study types meeting more than half the criteria for quality assessment. The evaluation of the risk of bias for the 8 RCTs ([Figure 2](#)) was carried out using the Cochrane Collaboration risk-of-bias tool [28], and the results were summarized using RevMan 5.3 software (Cochrane) [57]. Overall, the RCTs performed fairly well in the risk-of-bias assessment ([Figure 3](#)). About half the studies were assessed as having a low risk of selection bias because of proper random sequence generation (5/8) and allocation concealment (4/8), and a low risk of reporting bias (4/8), as outcomes reported could be compared with a priori protocols or trial registrations. Most studies reported blinding of outcome assessors (7/8) and a low risk of attrition bias because of low or equal dropout across groups or the use of intention-to-treat analyses (6/8). Most of the studies (5/8) had a high risk of performance bias, but this was predominantly because blinding was not possible given the nature of the intervention.

The cohort (n=9) and qualitative (n=3) studies assessed using the CASP checklists met, on average, 5/12 (range 1-10) and 7/10 (range 4-9) criteria, respectively [29]. Of the cohort studies, the questions with the best performance were, "Did the study address a clearly focused issue?" (8/9 yes), "Was the follow up long enough?" (8/9 yes), and "Do the results of this study fit with other available evidence?" (6/9 yes). Studies performed the worst, either by failing to meet the criteria or failing to report it, on questions about cohort recruitment (1/9 yes), identifying and accounting for confounding factors (1/9 yes), accurate exposure and outcome measurement (2/9 and 3/9 yes, respectively), and the applicability of results to the local population (3/9 yes). The qualitative studies, on the other hand, performed best on the questions about whether the qualitative methodology was appropriate, the consideration of ethical issues, clear statements of findings, and whether the results would help locally (3/3 yes for each). None of the 3 studies reported any consideration of the relationship between researcher and participant. They also performed poorly on questions about sample recruitment, data collection, and data analysis (1/3 yes for each).

**Figure 2.** Risk of bias summary: review authors' judgements about each risk of bias item for each included study.**Figure 3.** Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

The cross-sectional (n=5) and *other* (n=6) studies assessed using the AXIS tool met, on average, 50% (range 26-80%) and 42% (range 29-70%) of the criteria, respectively [30]. Percentages are reported instead of the exact number of criteria because

several of the questions were not applicable to the studies; so the total number of criteria assessed per study was not the same (averages 19 and 16; ranges 18-20, and 10-19, respectively). Overall, the cross-sectional studies performed best on questions

about the clarity of aims (5/5 yes), appropriate outcome variables for the aims (5/5 yes), internal consistency (5/5 yes), and adequate description of basic data (4/5 yes). They performed worst on questions about sample selection—whether it was taken from an appropriate base to represent the population (1/5 yes) and whether the process was likely to select a representative sample (0/5 yes)—the use of appropriate outcome measures (previously assessed; 0/5 yes), whether the methods were adequately described for replication (1/5 yes), and conflicts of interest (1/5 no, most did not report).

The *other* studies performed best on the questions about whether the study design was appropriate for the aims and whether the conclusions were justified by the results (6/6 yes for both). They also did well, overall, on the appropriate choice of outcome variables and internal consistency (5/6 yes for both). However, all the *other* studies for which the questions were applicable performed poorly on questions about the justification of sample size (0/5 yes), whether the selection process was likely to get a representative sample (0/5 yes), addressing nonresponders (0/2 yes), adequate description of basic data (0/4 yes), concerns about nonresponse bias (0/3 no), the presentation of results for all the analyses described in the methods (0/6 yes, although this was mostly because analyses were not adequately described in the methods), and conflicts of interest (0/6 no, again because nothing was reported). Furthermore, only 1 study adequately addressed the questions about the use of previously assessed outcome measures (1/5 yes), sufficient description of the methods for replication (1/6 yes), and discussion of study limitations (1/6 yes). It should be noted that the AXIS tool used to assess the *other* studies was designed for cross-sectional studies and does not fit exactly with the designs of these studies. Therefore, it is possible that these studies would perform better when assessed by a tool specific to their study type. Tables depicting the judgments for each question of the CASP cohort and qualitative checklists and the AXIS tool for the cross-sectional and *other* studies are included in [Multimedia Appendices 6-9 \[8,12,14,15,32,33,35,36,38-45,48,50-52,54-56\]](#).

## Discussion

### Principal Findings

In this systematic review, we examined 31 studies that evaluated the effectiveness and usability of conversational agents in health care. Overall, studies reported a moderate amount of evidence supporting the effectiveness, usability, and positive user perceptions of the agents. On average, two-thirds of the studies (67%) reported positive or mixed evidence for each evaluation outcome. However, this ranged significantly, with usability, agent performance, and satisfaction having the most support across the studies, and cost-effectiveness receiving hardly any support. It should also be noted that the definitions of *effectiveness* were highly varied and, as evidenced by the methodological limitations identified in the quality assessment, rarely evaluated with the scrutiny expected for medical devices. Although the results reported are promising for the use of conversational agents in health care, there are a number of limitations in both the studies analyzed and the structure of this review that questions the validity of this finding.

With regard to qualitative user perceptions of the agents, specific feedback was very mixed. Users highlighted many positive factors of the agents, particularly their personality and ability to provide empathy and emotional support, that they support learning, they are easy to use and access, and they help them be accountable, all of which support the generally positive evaluations of usability and satisfaction outcomes. However, there were a number of limitations of the agents that were consistently raised across the studies that reported qualitative feedback. These included the following: the agents had difficulty understanding them, the agents were repetitive and not sufficiently interactive, and the users had difficulty forming personal connections with the agents. This suggests that despite the generally positive usability reported by the studies, there are a number of barriers to the successful use of conversational agents in health care that will need to be addressed before they can achieve the greatest impact. It should be noted that this review only included studies of conversational agents that used NLP and that free-text inputs are likely to present greater difficulties for comprehension.

The results of this systematic review are largely consistent with the literature, particularly the previous systematic review evaluating conversational agents in health care [2]. They also found a limited quality of design and evidence in the included studies, with inconsistent reporting of study methods (including methods of selection, attrition, and a lack of validated outcome measures) and conflicts of interest [2]. The previous systematic review identified that high-quality evidence of effectiveness and patient safety was limited, which was also observed in this review. Similarly, it noted that high overall satisfaction was generally reported by the studies, but that the most common issues with conversational agents related to language understanding or poor dialogue management, which is consistent with our findings [2]. Some of this similarity in results is likely because of the overlap in included studies; 7 of their 17 included studies were also included in our review [2].

### Quality of the Evidence

As noted in a previous systematic review [2], there were significant issues with the quality of many of the included studies. One of the consistent issues among many of them was a high risk of selection bias. A large proportion of the studies relied on volunteers for the study, many of whom were recruited via self-selection means such as flyers and emails or by downloading the app being studied. The risk with self-selection recruitment is that participants who elect to take part in the study are already more positively predisposed to new technologies than those who do not participate, and would tend to evaluate the technology more positively. To make matters worse, several of the studies also did not sufficiently report their recruitment strategies, and so their potential selection bias cannot be accurately evaluated. In research such as this, where user perceptions are a main outcome, this is a serious concern. Future studies should take care to implement recruitment strategies that minimize this risk of selection bias or balance the potential bias in evaluations by actively recruiting participants who are less inclined toward new technology.

Another limitation of many of the studies was the small sample size. Almost two-thirds of the studies (19/31) used samples of less than 100 participants or items of analysis (eg, voice clips and clinical scenarios) with a median sample size of 48 across all the studies. Many also did not sufficiently report demographic data or whether their sample was representative of their target population. Although many of these studies were early feasibility and usability trials, this is an important issue to address in future research testing these agents to determine whether an agent will be used and used effectively by its target population.

## Limitations

The validity of the evidence extracted from the included studies was also affected by limitations in the structure of this review. The SF/HIT was used to provide a structured set of whole system implementation outcomes to evaluate the conversational agents [31]. However, an issue with the use of this framework, which was discovered during analysis, was that many of the included studies were describing system innovation. Therefore, they did not address or provide evidence for many of the outcomes described by the SF/HIT. Additionally, as the included data indicated a self-reported impact in the studies of effectiveness, the study effectiveness is biased favorably toward the authors' reporting of impact.

This limitation in the use of the framework for this review also highlights a limitation in many of these studies, namely, that they do not think about whole system implementation from the early stages of agent design, development, and testing. It is possible that the lack of evaluation of the implications of the agents for health care provision and resources was because of an emphasis on technology development and evaluation, rather than system integration. This is a pervasive issue in technological innovation, so much that it drove the development of the nonadoption, abandonment, scale-up, spread, and sustainability framework as a means of predicting and assessing the success of new health technologies [58] and the development and evaluation of new conversational agents to ensure that these later-stage implications of health care provision, cost-effectiveness, and privacy and security are sufficiently considered from the early stages of innovation. They must also be properly evaluated with a large sample of users, rather than be simply presented as unsubstantiated claims that the agent will reduce costs and save health care providers' time.

Additionally, in accordance with the SF/HIT framework, the impact of outcomes on each outcome was coded as *positive or mixed* or *neutral or negative*. However, this combination of positive and mixed outcomes reduces the granularity of the results. During the coding process, several outcomes were distinctly coded as *positive or mixed*, and collating the 2 outcome impacts into 1 reduces the precision of the information presented to the readers. Additionally, studies that did not assess the outcome in question were coded as *neutral or negative* because they did not provide explicit support for the outcome. In the analysis, outcomes were initially coded separately as positive, mixed, positive or mixed (for studies that reported a positive outcome but did not provide sufficient statistical evidence), and neutral or negative. This table is available in

[Multimedia Appendix 10](#). Positive and mixed outcomes were combined for the final presentation of the data in line with the framework. However, it might be more useful to distinguish between studies that attempted to find significant evidence for an outcome but did not and those that did not attempt it. This would provide a clearer picture of which outcomes are not being supported by the evidence and should be targeted for improvement, and which outcomes still need to be examined. In the future, it would be worth evaluating whether the coding system should be adjusted to provide a more detailed and informative summary of the evidence.

Further limitations of this review are that we limited the focus to include only unconstrained NLP and interaction. This was chosen as a focus because of the advantages NLP offers for simulating human-to-human interaction. However, it may have excluded studies of relevant conversational agents that could be satisfactory, useful, and effective in addressing current health care challenges. Additionally, no spidering searches were used to identify potentially relevant studies in the references of the included studies that were missed in the initial search. The exclusion of conference abstracts might also have caused relevant papers that were classified as abstracts to be missed; however, a previous systematic review that included conference abstracts in their search only had 1 included in their final selection [2]. The inclusion of only studies published in English is also likely to exclude relevant research on conversational agents conducted in other countries. These limitations should be addressed in future studies to ensure that the full body of relevant literature is examined.

## Future Directions

Future reviews of conversational agents in health care could be extended to include constrained NLP and non-NLP conversational agents. A synthesis of the evidence identified here with other types of conversational agents in health care, perhaps structured according to the taxonomy suggested by Montenegro et al [5], could be used to examine overall trends and provide a better picture of what is being used, what works, and what does not, to further guide the development of conversational agents that are most likely to be successful.

Future research should also include more qualitative evaluations of the features that users like and dislike. Only half (18/31) of the studies included in this review reported specific user feedback, despite the fact that 7 of the remaining 13 studies included some measure of usability or user perceptions. It will be important to identify all of the structural, physical, and psychological barriers to use if conversational agents are to achieve their potential for improving health care provision and reducing the strain on health care resources. To this end, it would be useful for future studies to structure their evaluation of conversational agents around a behavioral change framework (eg, the Behavior Change Wheel framework [59]). This is important not only when evaluating the effectiveness of behavior change-focused conversational agents, but also when determining whether and how the adoption of new conversational agent technology will be successful.

It will be important for future studies of conversational agents to take care to properly structure and report their studies to



improve the quality of the evidence. Without high-quality evidence, it is difficult to assess the current state of conversational agents in health care - what is working, and what needs to be improved to make them a more useful tool. Similarly, there is a gap in the evidence regarding the health economics of these agents. Very few studies in this review even discussed the cost analysis of the agent in questions, let alone provide substantive evidence about its cost-effectiveness. The evaluation of costs and outcomes of new technologies and their privacy, security, and interoperability will be necessary to advance value-based health care [60]. However, there is very little evidence to suggest that the conversational agents examined in this review considered or addressed these concerns. User feedback on 2 of the studies even noted that better interoperability between the agent and EHRs or health care providers would improve its usefulness.

## Conclusions

The objective of this systematic review was to synthesize evidence of conversational agents' usability, effectiveness, and

satisfaction in health care. Although the studies generally reported positive outcomes relating to the agents' usability and effectiveness, the quality of the evidence was not sufficient to provide strong evidence to support these claims. This study extended the literature by expanding its summary to examine a whole system set of evaluation outcomes, including cost-effectiveness, privacy, and security, which have not been systematically examined in previous reviews. In addition, it provides a distinct contribution by conducting a thematic analysis of the qualitative user perceptions of the agents. Further research is needed to examine the cost-effectiveness and value of these agents in health care, both in their current and potential states. Higher-quality studies—with more consistent reporting of design methods and better sample selection—are also needed to more accurately assess the usefulness and identify the key areas of improvement for current conversational agents. A more holistic approach to the design, development, and evaluation of conversational agents will help drive innovation and improve their value in health care.

## Acknowledgments

The authors would like to thank the outreach librarians Liz Callow (University of Oxford) and Kirsten Elliot (Imperial College London), for their assistance in developing search terms and reviewing search strategies. Specific funding for this work has not been acquired. EM's work on digital health solutions is currently supported by the Sir David Cooksey Fellowship in Healthcare Translation at the University of Oxford. The conclusions drawn in this paper were made by the authors and are not necessarily supported by the University of Oxford. The funding body had no role in the design, execution, or analysis of this systematic review.

## Authors' Contributions

CC and EM conceived the study topic and designed the review protocol. CC and MMI screened the studies. CC conducted the data extraction, which was validated by MMI, and MMI conducted the risk-of-bias and quality assessments, which were validated by EM. MMI and EM analyzed the extracted data. The methods section was drafted by CC, and the rest of the review was written by MMI with revisions from EM. MHS, EL, NP, EN and GM provided feedback on the final drafted text. EM supervised the study execution. The authors confirm that they have followed all the appropriate research reporting guidelines. The PRISMA checklist for systematic reviews has been uploaded as [Multimedia Appendix 11](#) along with other relevant materials.

## Conflicts of Interest

EL, NP, and GM are all employees of Ufonia Limited, a voice AI company. However, the paper was funded by the Sir David Cooksey Fellowship in Healthcare Translation at the University of Oxford, and Ufonia had no editorial influence on the final drafting. Their contribution was limited to feedback, given their applied voice AI expertise; therefore, no conflict of interest is identified.

### Multimedia Appendix 1

Search queries and number of results for each database.

[[DOCX File , 16 KB - jmir\\_v22i10e20346\\_app1.docx](#) ]

### Multimedia Appendix 2

EndNote search details.

[[DOCX File , 12 KB - jmir\\_v22i10e20346\\_app2.docx](#) ]

### Multimedia Appendix 3

Summary of study characteristics.

[[DOCX File , 27 KB - jmir\\_v22i10e20346\\_app3.docx](#) ]

### Multimedia Appendix 4



Data extraction table.

[[XLSX File \(Microsoft Excel File\), 166 KB - jmir\\_v22i10e20346\\_app4.xlsx](#)]

#### Multimedia Appendix 5

Summary of the thematic analysis of qualitative user feedback.

[[XLSX File \(Microsoft Excel File\), 112 KB - jmir\\_v22i10e20346\\_app5.xlsx](#)]

#### Multimedia Appendix 6

Summary of the quality assessment and judgments of cohort studies using the CASP (Critical Appraisal Skills Programme) Cohort Study Checklist.

[[XLSX File \(Microsoft Excel File\), 17 KB - jmir\\_v22i10e20346\\_app6.xlsx](#)]

#### Multimedia Appendix 7

Summary of the quality assessment and judgments of qualitative studies using the CASP (Critical Appraisal Skills Programme) Qualitative Study Checklist.

[[XLSX File \(Microsoft Excel File\), 12 KB - jmir\\_v22i10e20346\\_app7.xlsx](#)]

#### Multimedia Appendix 8

Summary of the quality assessment and judgments of the cross-sectional studies using the Appraisal tool for Cross-Sectional Studies tool.

[[XLSX File \(Microsoft Excel File\), 14 KB - jmir\\_v22i10e20346\\_app8.xlsx](#)]

#### Multimedia Appendix 9

Summary of the quality assessment and judgments of the 'other' studies using the Appraisal tool for Cross-Sectional Studies tool.

[[XLSX File \(Microsoft Excel File\), 13 KB - jmir\\_v22i10e20346\\_app9.xlsx](#)]

#### Multimedia Appendix 10

Summary of the studies based on the evaluation outcomes from the synthesis framework for the assessment of health information technology differentiating between positive and mixed outcomes.

[[XLSX File \(Microsoft Excel File\), 82 KB - jmir\\_v22i10e20346\\_app10.xlsx](#)]

#### Multimedia Appendix 11

PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) checklist.

[[DOC File , 64 KB - jmir\\_v22i10e20346\\_app11.doc](#)]

## References

1. Bibault J, Chaix B, Nectoux P, Pienkowsky A, Guillemasse A, Brouard B. Healthcare ex Machina: are conversational agents ready for prime time in oncology? Clin Transl Radiat Oncol 2019 May;16:55-59 [FREE Full text] [doi: [10.1016/j.ctro.2019.04.002](#)] [Medline: [31008379](#)]
2. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: [10.1093/jamia/ocy072](#)] [Medline: [30010941](#)]
3. Luxton DD. Ethical implications of conversational agents in global public health. Bull World Health Organ 2020 Apr 1;98(4):285-287 [FREE Full text] [doi: [10.2471/BLT.19.237636](#)] [Medline: [32284654](#)]
4. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: [10.7861/futurehosp.6-2-94](#)] [Medline: [31363513](#)]
5. Montenegro JL, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67 [FREE Full text] [doi: [10.1016/j.eswa.2019.03.054](#)]
6. Weizenbaum J. ELIZA — a computer program for the study of natural language communication between man and machine. Commun ACM 1983 Jan;26(1):23-28 [FREE Full text] [doi: [10.1145/357980.357991](#)]
7. Campillos-Llanos L, Thomas C, Bilinski ?, Zweigenbaum P, Rosset S. Designing a virtual patient dialogue system based on terminology-rich resources: challenges and evaluation. Nat Lang Eng 2019 Jul 15:1-38 [FREE Full text] [doi: [10.1017/s1351324919000329](#)]
8. Chang P, Sheng Y, Sang Y, Wang D. Developing a wireless speech- and touch-based intelligent comprehensive triage support system. Comput Inform Nurs 2008;26(1):31-38. [doi: [10.1097/01.NCN.0000304754.49116.b4](#)] [Medline: [18091619](#)]

9. Adams WG, Phillips BD, Bacic JD, Walsh KE, Shanahan CW, Paasche-Orlow MK. Automated conversation system before pediatric primary care visits: a randomized trial. *Pediatrics* 2014 Sep;134(3):e691-e699. [doi: [10.1542/peds.2013-3759](https://doi.org/10.1542/peds.2013-3759)] [Medline: [25092938](https://pubmed.ncbi.nlm.nih.gov/25092938/)]
10. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. *J Med Internet Res* 2019 Nov 7;21(11):e15360 [FREE Full text] [doi: [10.2196/15360](https://doi.org/10.2196/15360)] [Medline: [31697237](https://pubmed.ncbi.nlm.nih.gov/31697237/)]
11. Sun R, Aldunate R, Ratnam R, Jain S, Morrow D, Sosnoff J. Validity and usability of an automated fall risk assessment tool for older adults internet. *Innov Aging* 2018;362. [doi: [10.1093/geroni/igy023.1338](https://doi.org/10.1093/geroni/igy023.1338)]
12. Nakagawa S, Enomoto D, Yonekura S, Kanazawa H, Kuniyoshi Y. A Telecare System that Estimates Quality of Life through Communication. In: International Conference on Cloud Computing and Intelligence Systems. 2018 Presented at: CCIS'18; November 23-25, 2018; Nanjing, China. [doi: [10.1109/ccis.2018.8691360](https://doi.org/10.1109/ccis.2018.8691360)]
13. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (WOEBOT): a randomized controlled trial. *JMIR Ment Health* 2017 Jun 6;4(2):e19 [FREE Full text] [doi: [10.2196/mental.7785](https://doi.org/10.2196/mental.7785)] [Medline: [28588005](https://pubmed.ncbi.nlm.nih.gov/28588005/)]
14. Håvik R, Wake J, Flobak E, Lundervold A, Guribye F. A conversational interface for self-screening for ADHD in adults. *Internet Sci* 2019;144. [doi: [10.1007/978-3-030-17705-8\\_12](https://doi.org/10.1007/978-3-030-17705-8_12)]
15. Isaza-Restrepo A, Gómez MT, Cifuentes G, Argüello A. The virtual patient as a learning tool: a mixed quantitative qualitative study. *BMC Med Educ* 2018 Dec 6;18(1):297 [FREE Full text] [doi: [10.1186/s12909-018-1395-8](https://doi.org/10.1186/s12909-018-1395-8)] [Medline: [30522478](https://pubmed.ncbi.nlm.nih.gov/30522478/)]
16. van Heerden A, Ntinga X, Vilakazi K. The Potential of Conversational Agents to Provide a Rapid HIV Counseling and Testing Services. In: International Conference on the Frontiers and Advances in Data Science. 2017 Presented at: FADS'17; October 23-25, 2017; Xi'an, China. [doi: [10.1109/fads.2017.8253198](https://doi.org/10.1109/fads.2017.8253198)]
17. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. *J Health Commun* 2010;15(Suppl 2):197-210. [doi: [10.1080/10810730.2010.499991](https://doi.org/10.1080/10810730.2010.499991)] [Medline: [20845204](https://pubmed.ncbi.nlm.nih.gov/20845204/)]
18. Zhang Z, Bickmore T. Medical Shared Decision Making with a Virtual Agent. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. 2018 Presented at: IVA'18; November 5-8, 2018; Sydney, NSW, Australia. [doi: [10.1145/3267851.3267883](https://doi.org/10.1145/3267851.3267883)]
19. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can J Psychiatry* 2019 Jul;64(7):456-464 [FREE Full text] [doi: [10.1177/0706743719828977](https://doi.org/10.1177/0706743719828977)] [Medline: [30897957](https://pubmed.ncbi.nlm.nih.gov/30897957/)]
20. Russo A, D'Onofrio G, Gangemi A, Giuliani F, Mongiovi M, Ricciardi F, et al. Dialogue systems and conversational agents for patients with dementia: the human-robot interaction. *Rejuvenation Res* 2019 Apr;22(2):109-120. [doi: [10.1089/rej.2018.2075](https://doi.org/10.1089/rej.2018.2075)] [Medline: [30033861](https://pubmed.ncbi.nlm.nih.gov/30033861/)]
21. Xing Z, Yu F, Qanir YA, Guan T, Walker J, Song L. Intelligent conversational agents in patient self-management: a systematic survey using multi data sources. *Stud Health Technol Inform* 2019 Aug 21;264:1813-1814. [doi: [10.3233/SHTI190661](https://doi.org/10.3233/SHTI190661)] [Medline: [31438357](https://pubmed.ncbi.nlm.nih.gov/31438357/)]
22. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. *J Med Internet Res* 2017 May 9;19(5):e151 [FREE Full text] [doi: [10.2196/jmir.6553](https://doi.org/10.2196/jmir.6553)] [Medline: [28487267](https://pubmed.ncbi.nlm.nih.gov/28487267/)]
23. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine: qualitative study. *JMIR Res Protoc* 2018 Dec 4;7(12):e11072 [FREE Full text] [doi: [10.2196/11072](https://doi.org/10.2196/11072)] [Medline: [30514693](https://pubmed.ncbi.nlm.nih.gov/30514693/)]
24. de Cock C, Milne-Ives M, van Velthoven MH, Alturkistani A, Lam C, Meinert E. Effectiveness of conversational agents (virtual assistants) in health care: protocol for a systematic review. *JMIR Res Protoc* 2020 Mar 9;9(3):e16934 [FREE Full text] [doi: [10.2196/16934](https://doi.org/10.2196/16934)] [Medline: [32149717](https://pubmed.ncbi.nlm.nih.gov/32149717/)]
25. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* 2007 Jun 15;7:16 [FREE Full text] [doi: [10.1186/1472-6947-7-16](https://doi.org/10.1186/1472-6947-7-16)] [Medline: [17573961](https://pubmed.ncbi.nlm.nih.gov/17573961/)]
26. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *Br Med J* 2015 Jan 2;350:g7647 [FREE Full text] [doi: [10.1136/bmj.g7647](https://doi.org/10.1136/bmj.g7647)] [Medline: [25555855](https://pubmed.ncbi.nlm.nih.gov/25555855/)]
27. Higgins J. Cochrane Handbook for Systematic Reviews of Interventions. 2019. ISBN 2019:9781119536628.
28. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The cochrane collaboration's tool for assessing risk of bias in randomised trials. *Br Med J* 2011 Oct 18;343:d5928 [FREE Full text] [doi: [10.1136/bmj.d5928](https://doi.org/10.1136/bmj.d5928)] [Medline: [22008217](https://pubmed.ncbi.nlm.nih.gov/22008217/)]
29. CASP Checklists. Critical Appraisal Skills Programme: CASP. URL: <https://casp-uk.net/casp-tools-checklists/> [accessed 2020-09-11]
30. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open* 2016 Dec 8;6(12):e011458 [FREE Full text] [doi: [10.1136/bmjopen-2016-011458](https://doi.org/10.1136/bmjopen-2016-011458)] [Medline: [27932337](https://pubmed.ncbi.nlm.nih.gov/27932337/)]

31. Christopoulou SC, Kotsilieris T, Anagnostopoulos I. Assessment of health information technology interventions in evidence-based medicine: a systematic review by adopting a methodological evaluation framework. *Healthcare (Basel)* 2018 Aug 31;6(3):- [[FREE Full text](#)] [doi: [10.3390/healthcare6030109](https://doi.org/10.3390/healthcare6030109)] [Medline: [30200307](#)]
32. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Assessing the Usability of a Chatbot for Mental Health Care. In: Bodrunova S. *Internet Science.*, editor. *Lecture Notes in Computer Science*, vol 11551 Springer, Cham; 2019.
33. Elmasri D, Maeder A. A Conversational Agent for an Online Mental Health Intervention Internet. *Brain Informatics and Health.*? 2016;251. [doi: [10.1007/978-3-319-47103-7\\_24](https://doi.org/10.1007/978-3-319-47103-7_24)]
34. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. A conversational agent for an online mental health intervention internetusing psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment Health* 2018 Dec 13;5(4):e64 [[FREE Full text](#)] [doi: [10.2196/mental.9782](https://doi.org/10.2196/mental.9782)] [Medline: [30545815](#)]
35. Hudlicka E. Virtual training and coaching of health behavior: example from mindfulness meditation training. *Patient Educ Couns* 2013 Aug;92(2):160-166 [[FREE Full text](#)] [doi: [10.1016/j.pec.2013.05.007](https://doi.org/10.1016/j.pec.2013.05.007)] [Medline: [23809167](#)]
36. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth* 2018 Nov 23;6(11):e12106 [[FREE Full text](#)] [doi: [10.2196/12106](https://doi.org/10.2196/12106)] [Medline: [30470676](#)]
37. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv* 2017 Dec;10:39-46 [[FREE Full text](#)] [doi: [10.1016/j.invent.2017.10.002](https://doi.org/10.1016/j.invent.2017.10.002)] [Medline: [30135751](#)]
38. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Sci Rep* 2017 Feb 16;7:42656 [[FREE Full text](#)] [doi: [10.1038/srep42656](https://doi.org/10.1038/srep42656)] [Medline: [28205601](#)]
39. Yasavur U, Lisetti C, Rishe N. Let's talk! speaking virtual counselor offers you a brief intervention. *J Multimodal User Interfaces* 2014 Sep 5;8(4):381-398. [doi: [10.1007/s12193-014-0169-9](https://doi.org/10.1007/s12193-014-0169-9)]
40. Xu R, Mei G, Zhang G, Gao P, Judkins T, Cannizzaro M, et al. A voice-based automated system for PTSD screening and monitoring. *Stud Health Technol Inform* 2012;173:552-558. [Medline: [22357057](#)]
41. Washburn M, Bordnick P, Rizzo AS. A pilot feasibility study of virtual patient simulation to enhance social work students' brief mental health assessment skills. *Soc Work Health Care* 2016 Oct;55(9):675-693. [doi: [10.1080/00981389.2016.1210715](https://doi.org/10.1080/00981389.2016.1210715)] [Medline: [27552646](#)]
42. Dimeff LA, Jobes DA, Chalker SA, Piehl BM, Duvivier LL, Lok BC, et al. A novel engagement of suicidality in the emergency department: virtual collaborative assessment and management of suicidality. *Gen Hosp Psychiatry* 2020;63:119-126. [doi: [10.1016/j.genhosppsych.2018.05.005](https://doi.org/10.1016/j.genhosppsych.2018.05.005)] [Medline: [29934033](#)]
43. Spänig S, Emberger-Klein A, Sowa J, Canbay A, Menrad K, Heider D. The virtual doctor: an interactive clinical-decision-support system based on deep learning for non-invasive prediction of diabetes. *Artif Intell Med* 2019 Sep;100:101706. [doi: [10.1016/j.artmed.2019.101706](https://doi.org/10.1016/j.artmed.2019.101706)] [Medline: [31607340](#)]
44. Ghosh S, Bhatia S, Bhatia A. Quro: facilitating user symptom check using a personalised chatbot-oriented dialogue system. *Stud Health Technol Inform* 2018;252:51-56. [Medline: [30040682](#)]
45. Chaix B, Bibault J, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. *JMIR Cancer* 2019 May 2;5(1):e12856 [[FREE Full text](#)] [doi: [10.2196/12856](https://doi.org/10.2196/12856)] [Medline: [31045505](#)]
46. Bibault J, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. *J Med Internet Res* 2019 Nov 27;21(11):e15787 [[FREE Full text](#)] [doi: [10.2196/15787](https://doi.org/10.2196/15787)] [Medline: [31774408](#)]
47. Heyworth L, Kleinman K, Oddleifson S, Bernstein L, Frampton J, Lehrer M, et al. Comparison of interactive voice response, patient mailing, and mailed registry to encourage screening for osteoporosis: a randomized controlled trial. *Osteoporos Int* 2014 May;25(5):1519-1526. [doi: [10.1007/s00198-014-2629-1](https://doi.org/10.1007/s00198-014-2629-1)] [Medline: [24566584](#)]
48. Rhee H, Allen J, Mammen J, Swift M. Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study. *Patient Prefer Adherence* 2014;8:63-72 [[FREE Full text](#)] [doi: [10.2147/PPA.S53504](https://doi.org/10.2147/PPA.S53504)] [Medline: [24470755](#)]
49. Simon SR, Zhang F, Soumerai SB, Ensroth A, Bernstein L, Fletcher RH, et al. Failure of automated telephone outreach with speech recognition to improve colorectal cancer screening: a randomized controlled trial. *Arch Intern Med* 2010 Feb 8;170(3):264-270. [doi: [10.1001/archinternmed.2009.522](https://doi.org/10.1001/archinternmed.2009.522)] [Medline: [20142572](#)]
50. Borja-Hart NL, Spivey CA, George CM. Use of virtual patient software to assess student confidence and ability in communication skills and virtual patient impression: a mixed-methods approach. *Curr Pharm Teach Learn* 2019 Jul;11(7):710-718. [doi: [10.1016/j.cptl.2019.03.009](https://doi.org/10.1016/j.cptl.2019.03.009)] [Medline: [31227094](#)]
51. Philip P, Bioulac S, Sauteraud A, Chaufton C, Olive J. Could a virtual human be used to explore excessive daytime sleepiness in patients? *Presence* 2014 Nov 1;23(4):369-376. [doi: [10.1162/pres\\_a.00197](https://doi.org/10.1162/pres_a.00197)]
52. Galescu L, Allen J, Ferguson G, Quinn J, Swift M. Speech Recognition in a Dialog System for Patient Health Monitoring. In: *International Conference on Bioinformatics and Biomedicine Workshop*. 2009 Presented at: BIBMW'09; November 1-4, 2009; Washington, DC. [doi: [10.1109/bibmw.2009.5332111](https://doi.org/10.1109/bibmw.2009.5332111)]

53. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a web-based physical activity intervention with an avatar: randomized controlled trial. *J Med Internet Res* 2014 Feb 13;16(2):e48 [FREE Full text] [doi: [10.2196/jmir.2974](https://doi.org/10.2196/jmir.2974)] [Medline: [24550153](https://pubmed.ncbi.nlm.nih.gov/24550153/)]
54. Crutzen R, Peters GY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. *J Adolesc Health* 2011 May;48(5):514-519. [doi: [10.1016/j.jadohealth.2010.09.002](https://doi.org/10.1016/j.jadohealth.2010.09.002)] [Medline: [21501812](https://pubmed.ncbi.nlm.nih.gov/21501812/)]
55. Wong W, Thangarajah J, Padgham L. Contextual question answering for the health domain. *J Am Soc Inf Sci Tec* 2012 Oct 30;63(11):2313-2327 [FREE Full text] [doi: [10.1002/asi.22733](https://doi.org/10.1002/asi.22733)]
56. Ireland D, Atay C, Liddle J, Bradford D, Lee H, Rushin O, et al. Hello Harlie: enabling speech monitoring through chat-bot conversations. *Stud Health Technol Inform* 2016;227:55-60. [Medline: [27440289](https://pubmed.ncbi.nlm.nih.gov/27440289/)]
57. Copenhagen: The Nordic Cochrane Centre. RevMan. URL: <https://community.cochrane.org/help/tools-and-software/revman-5> [accessed 2020-09-11]
58. Greenhalgh T, Wherton J, Papoutsis C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J Med Internet Res* 2017 Nov 1;19(11):e367 [FREE Full text] [doi: [10.2196/jmir.8775](https://doi.org/10.2196/jmir.8775)] [Medline: [29092808](https://pubmed.ncbi.nlm.nih.gov/29092808/)]
59. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011 Apr 23;6:42 [FREE Full text] [doi: [10.1186/1748-5908-6-42](https://doi.org/10.1186/1748-5908-6-42)] [Medline: [21513547](https://pubmed.ncbi.nlm.nih.gov/21513547/)]
60. Meinert E, Alturkistani A, Brindley D, Knight P, Wells G, Pennington ND. The technological imperative for value-based health care. *Br J Hosp Med (Lond)* 2018 Jun 2;79(6):328-332. [doi: [10.12968/hmed.2018.79.6.328](https://doi.org/10.12968/hmed.2018.79.6.328)] [Medline: [29894248](https://pubmed.ncbi.nlm.nih.gov/29894248/)]

## Abbreviations

**AI:** artificial intelligence  
**AXIS:** Appraisal tool for Cross-Sectional Studies  
**CASP:** Critical Appraisal Skills Programme  
**ECA:** embodied conversational agent  
**ED:** emergency department  
**EHR:** electronic health record  
**IVR:** interactive voice response  
**NLP:** natural language processing  
**PRISMA:** Preferred Reporting Items for Systematic Review and Meta-Analyses  
**RCT:** randomized controlled trial  
**SF/HIT:** synthesis framework for the assessment of health information technology

*Edited by G Eysenbach; submitted 17.05.20; peer-reviewed by S McRoy, X Huang; comments to author 08.06.20; revised version received 12.06.20; accepted 02.09.20; published 22.10.20.*

### *Please cite as:*

Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, Normando E, Meinert E  
*The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review*  
*J Med Internet Res* 2020;22(10):e20346  
URL: <http://www.jmir.org/2020/10/e20346/>  
doi: [10.2196/20346](https://doi.org/10.2196/20346)  
PMID: [33090118](https://pubmed.ncbi.nlm.nih.gov/33090118/)

©Madison Milne-Ives, Caroline de Cock, Ernest Lim, Melissa Harper Shehadeh, Nick de Pennington, Guy Mole, Eduardo Normando, Edward Meinert. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Federated Learning on Clinical Benchmark Data: Performance Assessment

Geun Hyeong Lee<sup>1</sup>, MS; Soo-Yong Shin<sup>1,2,3</sup>, PhD

<sup>1</sup>Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea

<sup>2</sup>Big Data Research Center, Samsung Medical Center, Seoul, Republic of Korea

<sup>3</sup>Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon, Republic of Korea

**Corresponding Author:**

Soo-Yong Shin, PhD

Department of Digital Health

Samsung Advanced Institute for Health Sciences & Technology

Sungkyunkwan University

115 Irwon-ro

Gangnam-gu

Seoul, 06355

Republic of Korea

Phone: 82 2 3410 1449

Email: [sy.shin@skku.edu](mailto:sy.shin@skku.edu)

## Abstract

**Background:** Federated learning (FL) is a newly proposed machine-learning method that uses a decentralized dataset. Since data transfer is not necessary for the learning process in FL, there is a significant advantage in protecting personal privacy. Therefore, many studies are being actively conducted in the applications of FL for diverse areas.

**Objective:** The aim of this study was to evaluate the reliability and performance of FL using three benchmark datasets, including a clinical benchmark dataset.

**Methods:** To evaluate FL in a realistic setting, we implemented FL using a client-server architecture with Python. The implemented client-server version of the FL software was deployed to Amazon Web Services. Modified National Institute of Standards and Technology (MNIST), Medical Information Mart for Intensive Care-III (MIMIC-III), and electrocardiogram (ECG) datasets were used to evaluate the performance of FL. To test FL in a realistic setting, the MNIST dataset was split into 10 different clients, with one digit for each client. In addition, we conducted four different experiments according to basic, imbalanced, skewed, and a combination of imbalanced and skewed data distributions. We also compared the performance of FL to that of the state-of-the-art method with respect to in-hospital mortality using the MIMIC-III dataset. Likewise, we conducted experiments comparing basic and imbalanced data distributions using MIMIC-III and ECG data.

**Results:** FL on the basic MNIST dataset with 10 clients achieved an area under the receiver operating characteristic curve (AUROC) of 0.997 and an F1-score of 0.946. The experiment with the imbalanced MNIST dataset achieved an AUROC of 0.995 and an F1-score of 0.921. The experiment with the skewed MNIST dataset achieved an AUROC of 0.992 and an F1-score of 0.905. Finally, the combined imbalanced and skewed experiment achieved an AUROC of 0.990 and an F1-score of 0.891. The basic FL on in-hospital mortality using MIMIC-III data achieved an AUROC of 0.850 and an F1-score of 0.944, while the experiment with the imbalanced MIMIC-III dataset achieved an AUROC of 0.850 and an F1-score of 0.943. For ECG classification, the basic FL achieved an AUROC of 0.938 and an F1-score of 0.807, and the imbalanced ECG dataset achieved an AUROC of 0.943 and an F1-score of 0.807.

**Conclusions:** FL demonstrated comparative performance on different benchmark datasets. In addition, FL demonstrated reliable performance in cases where the distribution was imbalanced, skewed, and extreme, reflecting the real-life scenario in which data distributions from various hospitals are different. FL can achieve high performance while maintaining privacy protection because there is no requirement to centralize the data.

(*J Med Internet Res* 2020;22(10):e20891) doi:[10.2196/20891](https://doi.org/10.2196/20891)



**KEYWORDS**

federated learning; medical data; privacy protection; machine learning; deep learning

## Introduction

### Background

Traditional machine learning and deep learning require a centralized dataset to train a model. Therefore, such methods not only require data transfer to collect data from many devices, people, or institutions but also have a high computational cost because they must be trained on large datasets. When collecting privacy-sensitive data such as medical data, privacy protection is a major hurdle. Centralized databases are the main targets of hacking attacks, and therefore the risk of a data breach is severely increased [1,2]. Moreover, data centralization increases the risk of reidentification of deidentified data because of the increased data size [3].

To reduce the computational cost, Google proposed a method known as federated learning (FL), which uses the computational cores in mobile devices [4-6]. In FL, training is performed at the individual client level, and then the local weights of each client are sent to the server. The server collects the updated local weights and calculates the new global weights. Subsequently, the client downloads the global weights from the server and continues the training process. Since its first use in mobile apps [7-9], many researchers have been studying and improving FL in various fields [10-14]. In particular, studies on heterogeneity of data [4,15], robust optimization [16-20], and security methods such as differential privacy and secure multiparty computation have also been conducted with an FL approach [12,21,22]. Research on FL has also been conducted in the medical field [10,13,19]. In particular, studies have been conducted using electronic medical records and brain tumor data [23-25]. However, the application of FL to real medical data has not been sufficiently studied.

FL can be used to resolve privacy issues and mitigate the risk of a data breach in clinical information, since transfer and centralization of data are not required. Privacy protection is particularly beneficial for medical data analysis, since medical data represent some of the most sensitive types of personal data. To protect patients' privacy, deidentification methods have typically been applied [26-28]. However, data centralization is required for both deidentifying data and evaluating the risk of reidentification. If the data are centralized, the risk of a data breach is increased. Moreover, when deidentifying the dataset, the direct or indirect identifiers in the medical data must be determined. This is challenging because of the lack of clear guidelines. The Health Insurance Portability and Accountability Act in the United States provides clear deidentification guidance; it defines 18 types of protected health information to be removed [29]. However, many researchers and social activists claim that this guidance should be revised to enhance privacy protection [30]. In contrast, FL does not require the centralization of raw

data. As a result, even the FL developers cannot access the raw data. Therefore, FL can solve privacy or deidentification issues that occur when using clinical data.

### Objectives

The aim of this study was to assess the performance of FL on three benchmark datasets: the Modified National Institute of Standards and Technology (MNIST) dataset, Medical Information Mart for Intensive Care-III (MIMIC-III) dataset, and PhysioNet Electrocardiogram (ECG) dataset. We also verified FL in environments that simulate real-world data distributions by modifying the MNIST, MIMIC-III, and ECG datasets.

## Methods

### FL Code and Server

FL is supported by several open-source projects, including TensorFlow Federated in TensorFlow 2.0 [31], PySyft [32,33], and Federated AI Technology Enabler [34,35]. However, there are limitations in using these libraries. First, most of these libraries only support a single server and not a network environment. Therefore, there is no control process for data communication. Second, as a prototype, the necessary features were not fully implemented to handle a complex dataset. For future research using real clinical data from hospitals, we implemented our own client-server version of FL using Python. The implemented server code is available on the FL\_Server repository [36] and the client code is available on the FL\_Client repository [37]. The MNIST dataset analyzed during the current study is available in the Keras package in the TensorFlow framework. Additionally, the original code used to generate and preprocess the MIMIC-III experiment used in this study referred to the mimic3-benchmarks repository [38]. The original MIMIC-III dataset analyzed during this study is available on the PhysioNet repository [39]. The ECG dataset analyzed during this study is available on the 2017 PhysioNet/CinC Challenge website [40]. The model and environment assessed in this study refer to Hannun et al [41].

The FL server was developed using the Django framework and Python in Amazon Web Services (AWS). The server provides several application programming interfaces (APIs) for communication with a client, as shown in Table 1, and performs federated averaging (FedAVG) [4], which calculates the weighted averages. FedAVG is a widely used optimization algorithm that calculates the average value when the local weights collected from the client reach a specific level. The implemented code was deployed and managed in AWS Beanstalk, which was continuously monitored during the training process.

**Table 1.** Application programming interface calls provided by the server.

Method	URL	Parameter	Description	Return
GET	/round	N/A <sup>a</sup>	Request current round	Number
GET	/weight	N/A	Request global weight	List
PUT	/weight	List	Update local weight	N/A

<sup>a</sup>N/A: Not applicable.

## Client

The client consists of three components. The first is the local learning component, which builds a suitable model for the dataset during the learning phase. The second is the communication component, which updates local weights according to the results of local training (the first component) on the server and downloads the global weights from the server. The third is the performance measure component, in which the performance of each client is measured using the downloaded global weights. The implemented code was deployed on an AWS EC2 instance. We used the specifications of g4dn.xlarge with the NVIDIA T4 Tensor core GPU for the Amazon instance.

## Communications

Client–server communication for FL was implemented based on the process described by McMahan et al [42]. However, the implemented code exhibits some differences. The communication assumes that all clients (hospitals) are always powered (as is the case for a typical computer but not for a mobile device) and that their online status is maintained by a wired network connection. In addition, rather than selecting clients via an eligibility criterion from multiple client pools (thousands or millions), the code was implemented to manage a predefined fixed number of clients. In other words, all clients could participate in each round.

A schematic diagram of the FL client–server communication is shown in [Multimedia Appendix 1](#). In brief, the client decides whether to participate in the current round through the API. If it has already participated (sending local weights to the server), it waits to participate in the next round. The server waits for the client's weight updates and ensures that no clients are eventually dropped. All communications are performed through the API provided by the server. The monitoring system is used to continuously observe system abnormalities.

## Datasets

### MNIST

The MNIST dataset, which consists of digit handwriting images, contained 70,000 samples (including 60,000 for training and 10,000 for testing). The basic model was a simple artificial neural network with an input layer, one hidden layer with 128 units with a rectified linear unit activation function, and an output layer. The hyperparameters for training were set as follows: batch size 32, maximum 1000 epochs, and early stopping. Stochastic gradient descent was used as an optimizer [43].

For FL, we used 10 individual clients to best mimic a real environment. We modified the datasets and hyperparameters

of the learning algorithms. The datasets were modified considering differences in the distribution of medical data between hospitals. Hyperparameters were adjusted for training in each client. The proposed approach was evaluated on the MNIST dataset in four different experiments.

We first evaluated the basic performance of the FL. Ten clients randomly selected 600 images from the basic dataset. We continued the process for up to 500 rounds and observed the results. For the imbalanced FL experiment, each client used different sizes of randomly selected data, ranging from 1 to 600, for training (ie, one client used 36 data points and another client used 537 data points). However, other environments such as hyperparameters and the number of rounds were the same as set in the basic FL experiment. In addition, the MNIST dataset was split into single-digit groups, ranging from 0 to 9. Each of the 10 numbers was assigned to 10 different clients. Consequently, each client had a single digit instead of 10. This modified MNIST simulated an extremely skewed data distribution. Each client randomly selected 600 images from a dataset with a single digit for training. The simple artificial neural network used in the basic model was also used in these experiments. The hyperparameters were set as follows: 5 epochs and a batch size of 10. We continued the process for up to 3000 rounds and observed the results. For evaluation, a model was created with the latest updated global weights using 10,000 test samples. Finally, we conducted an extension of the modified MNIST FL that represents a skewed distribution. Each client was trained on data with an imbalanced and skewed distribution. Hence, each client was trained only on a single digit using a randomly selected sample.

### MIMIC-III

The MIMIC-III dataset is a clinical dataset related to human health information, including demographics, vital signs, laboratory tests, and medications from intensive care units. MIMIC-III data were preprocessed using a state-of-the-art (SOTA) benchmark [44]. In this case, FL experiments with three individual clients were performed to predict in-hospital mortality, which is a classification problem that predicts death within the first 48 hours of an intensive care unit stay. After preprocessing the MIMIC-III dataset using the method described by Harutyunyan et al [44], the dataset contained 21,139 samples (including 17,903 for training and 3236 for testing). The basic model was a standard long short-term memory (LSTM) with reference to the benchmark [44]. The LSTM was chosen with 16 hidden units, depth 2, dropout 0.3, time step 1.0, batch size 8, and an adaptive moment estimation (ADAM) optimizer.

For FL, randomly chosen samples from the original dataset were divided into 3 datasets without duplication and assigned

to each client. This simulates having data from three different institutions. The same basic LSTM was used, and hyperparameters were set as follows: 2 epochs and a batch size of 4. We continued the process for up to 30 rounds and observed the results.

For the basic FL experiment, each client was trained on a subset of data that were split into three parts with the same data size without duplication. For the imbalanced FL experiment, all data were split into 50%, 30%, and 20% without duplication, and one subset was assigned to each client.

## ECG

The 2017 PhysioNet/CinC Challenge ECG dataset was used in this study [40]. This target problem is a multiclassification problem that classifies four signals: atrial fibrillation, normal sinus rhythm, alternative rhythm, and noisy using a single short ECG signal. The total data size is 8528 single-lead ECG data points. The dataset was divided into 90% training data (7676) and 10% test data (852). For traditional learning, a convolution neural network with 34 layers based on Hannun et al [41] was applied to the ECG dataset. The hyperparameters were chosen with a batch size of 32 and an ADAM optimizer.

For FL, randomly chosen samples from the original dataset were divided into 3 datasets without duplication and assigned to 3 clients. The same model was used, and hyperparameters were set as follows: 3 epochs and a batch size of 16. We continued the process for up to 30 rounds and observed the results.

For the basic FL experiment, each client was trained on a subset of data that were split into three parts with the same data size without duplication. For the imbalanced FL experiment, all data were split into 50%, 30%, and 20% without duplication, and a subset was assigned to each client.

## Evaluation

During training, we monitored the FL accuracy to evaluate performance. If the accuracy did not improve during the round, we completed the FL. Finally, we chose the best model and conducted bootstrapping to determine if there were significant differences between the experiments.

In all experiments, the area under the receiver operating characteristic curve (AUROC) score and F1-score were used as performance metrics. In addition, we evaluated the confusion matrix, precision, recall, or area under the precision recall curve (AUPRC) for comparison with the performance of the SOTA method. We calculated the 95% CIs and resampled the test set  $K$  times (for MNIST and ECG,  $K$  was 100, whereas for MIMIC-III,  $K$  was 10,000).

## Results

### MINST

The proposed approach was evaluated on the MNIST dataset for five different cases (as described in the Methods). Table 2 presents the values of the AUROC and F1-score for each case, and Multimedia Appendix 2 presents the confusion matrix for each case.

**Table 2.** Comparison of the experimental results for the five different MNIST cases described in the Methods.<sup>a</sup>

Experiments	AUROC <sup>b</sup> (95% CI)	F1-score (95% CI)	Precision (95% CI)	Recall (95% CI)
CML <sup>c</sup>	0.999 (0.999-0.999)	0.981 (0.978-0.983)	0.981 (0.972-0.989)	0.981 (0.971-0.989)
Basic FL <sup>d</sup>	0.997 (0.996-0.998)	0.946 (0.941-0.950)	0.945 (0.929-0.959)	0.945 (0.930-0.959)
Imbalanced FL	0.995 (0.994-0.995)	0.921 (0.917-0.927)	0.920 (0.904-0.937)	0.920 (0.903-0.937)
Skewed FL	0.992 (0.991-0.993)	0.905 (0.899-0.911)	0.905 (0.885-0.922)	0.904 (0.885-0.920)
Imbalanced and skewed FL	0.990 (0.989-0.991)	0.891 (0.884-0.896)	0.890 (0.869-0.909)	0.889 (0.868-0.908)

<sup>a</sup>All experiments used the same model and hyperparameters. All results are presented with a 95% CI by resampling the validation task 100 times.

<sup>b</sup>AUROC: area under the receiver operating characteristic curve.

<sup>c</sup>CML: centralized traditional machine-learning method.

<sup>d</sup>FL: federated learning.

Centralized machine learning (CML) is a baseline training method that was used as a control group. CML achieved an AUROC of 0.999 and an F1-score of 0.981. For basic FL, the AUROC and F1-score were 0.997 and 0.946, respectively. The initial performance of the basic FL was fairly high, with an accuracy of approximately 0.800, which continually improved (Multimedia Appendix 3A).

Imbalanced FL was designed to reflect a realistic clinical data distribution. As described in the Methods section, each client had a different training data size. Interestingly, the performance of imbalanced FL was significantly superior, with an AUROC and F1-score of 0.995 and 0.921, respectively. The initial

performance was rather poor, as expected. However, after several rounds of processing, the performance rapidly improved to reach an accuracy of 0.900, after which the performance improvement was slow (Multimedia Appendix 3B).

Skewed FL assumed an extreme case. Each client had only one digit from 0 to 9, thereby simulating a situation in which each hospital has a unique subpopulation of patients without overlaps. The final AUROC and F1-score were 0.992 and 0.905, respectively. As expected, the initial performance was poor; however, it rapidly improved after the initial rounds (Multimedia Appendix 3C).

The most extreme case was designed by combining an imbalanced and a skewed dataset. In this experiment, the AUROC and F1-score were 0.990 and 0.891, respectively. Similar to the skewed FL, the initial performance was very poor, but it rapidly improved after the initial rounds ([Multimedia Appendix 3D](#)).

Additionally, the precision and recall results for each digit class classification in each experiment are presented in [Multimedia Appendices 4-8](#).

### MIMIC-III

The proposed approach was evaluated on the MIMIC-III dataset in two different cases to compare the performance with a

reported benchmark. FL experiments were performed on three individual clients. Apart from the AUROC and F1-score, we also refer to the AUPRC, which is reported in the benchmark [44]. The results are presented in [Table 3](#) and in [Multimedia Appendices 9 and 10](#).

SOTA performance was achieved by executing the codes provided in Harutyunyan

[38]. FL achieved an AUROC, F1-score, and AUPROC comparable with those of the SOTA method. The imbalanced FL experiment, as an extension of the basic MIMIC-III FL, also achieved AUROC, F1-score, and AUPRC comparable with those of SOTA ([Table 3](#)).

**Table 3.** Comparison results of MIMIC-III.<sup>a</sup>

Experiments	AUROC <sup>b</sup> (95% CI)	F1-score (95% CI)	AUPRC <sup>c</sup> (95% CI)	Precision (95% CI)	Recall (95% CI)
SOTA <sup>d</sup>	0.857 (0.837-0.875)	0.944 (0.938-0.950)	0.505 (0.451-0.558)	0.973 (0.967-0.979)	0.773 (0.907-0.927)
Basic FL <sup>e</sup>	0.850 (0.830-0.869)	0.944 (0.938-0.950)	0.483 (0.427-0.537)	0.975 (0.969-0.980)	0.797 (0.906-0.926)
Imbalanced FL	0.850 (0.829-0.869)	0.943 (0.937-0.949)	0.481 (0.426-0.535)	0.981 (0.976-0.986)	0.714 (0.897-0.918)

<sup>a</sup>All results are presented with a 95% CI by resampling 10,000 times.

<sup>b</sup>AUROC: area under the receiver operating characteristic curve.

<sup>c</sup>AUPRC: area under the precision-recall curve.

<sup>d</sup>SOTA: state of the art.

<sup>e</sup>FL: federated learning.

### ECG

The proposed approach was evaluated on the ECG database using two different methods to compare the performance with a reported benchmark [41]. The results are presented in [Table 4](#) and [Multimedia Appendices 11-14](#).

Benchmark results were achieved using the code available on github [45]. The AUROC and F1-score of both basic and imbalanced FL were comparable with those of the benchmark ([Table 4](#)).

**Table 4.** Comparison results for the electrocardiogram dataset.<sup>a</sup>

Experiments	AUROC <sup>b</sup> (95% CI)	F1-score (95% CI)	Precision (95% CI)	Recall (95% CI)
Benchmark	0.954 (0.930-0.978)	0.814 (0.655-0.910)	0.820 (0.672-0.943)	0.814 (0.640-0.936)
Basic FL <sup>c</sup>	0.938 (0.860-0.978)	0.807 (0.651-0.931)	0.823 (0.645-0.942)	0.795 (0.660-0.925)
Imbalanced FL	0.943 (0.883-0.977)	0.807 (0.635-0.902)	0.830 (0.650-0.935)	0.788 (0.626-0.905)

<sup>a</sup>All results are presented with a 95% CI by resampling 100 times.

<sup>b</sup>AUROC: area under the receiver operating characteristic curve.

<sup>c</sup>FL: federated learning.

## Discussion

### Principal Findings

When comparing the performances of CML and FL in basic MNIST experiments, both the AUROC and F1-score were high. Unexpectedly, when using an imbalanced dataset, FL delivered good performance with only small differences (AUROC and F1-score of 0.003 and 0.035, respectively). When using a skewed dataset, FL also yielded remarkable results with respect to both the AUROC and F1-score. When comparing the confusion matrices for experiments with four datasets (ie, normal, imbalanced, skewed, and a combination of two distributions), FL showed some deterioration in performance

for visually similar numbers (eg, 3 vs 5; 4 vs 9). Even in the basic MNIST classification, the performance was relatively poor in these cases. However, this problem was not related to the small sizes of the training datasets. When we monitored the size of the training datasets for each client, the dataset for class 5 was not small. Moreover, depending on the experiment, the datasets for class 1 or 7 could be small, but superior classification performance was nevertheless achieved. This trend was maintained in the experiments with basic FL and imbalanced FL using the MIMIC-III dataset.

The FL experiments using MIMIC-III also exhibited good and competitive performance compared to a benchmark that has been trained on CML. The experimental results of in-hospital



mortality using the MIMIC-III dataset, which is a well-known dataset with real clinical data, also showed good performance. This experiment was performed by splitting the randomly selected MIMIC-III data into three parts (ie, from the perspective of each institution, learning one-third of the total data). However, the performances of FL and CML were almost the same, with only a 0.005 difference in AUROC detected compared with the SOTA performance reported by Harutyunyan et al [44]. Before the experiments, we expected that the performance of FL would be slightly inferior to that of CML because FL uses a distributed dataset instead of a centralized dataset. Nevertheless, no significant difference was found in well-known evaluation indicators such as accuracy, sensitivity, precision, and F1-score (except for AUROC). Experimental results with an imbalanced dataset were very similar to those of basic FL. Therefore, an individual client may only use a small amount of data for training in FL, and the results will be similar to those achieved when all available data are used for training.

The FL experiments using ECG data also exhibited good and competitive performance compared to CML. This experiment was performed by splitting the randomly selected ECG data into three parts (ie, from the perspective of each institution, learning on one-third of the total data) with each using different data distributions.

However, the performance of FL and CML was not significantly different. Experimental results with an imbalanced dataset were very similar to those for basic FL. As shown in [Multimedia Appendices 12-14](#), the noisy case was shown to have relatively low performance in precision and recall. This is because the data size for training was only 3% of the total size. However, the other classes performed well, such as atrial fibrillation, normal sinus rhythm, and alternative rhythm.

The performance of FL was verified using three datasets with changed data distributions: imbalanced (with disproportionately represented classes) and skewed (the distribution of the target variable was different) to imitate real-world medical data. As a result, FL was comparable to CML. During the initial rounds, only a relatively small amount of data was used on each client instead of an ensemble; therefore, the performance of FL was significantly inferior to that of CML. However, in the subsequent rounds, the performance of FL (with respect to AUROC and F1-score) became similar to that of CML. Typically, medical centers have datasets with very different distributions, and our results demonstrate that FL is suitable for real-world medical datasets without requiring data centralization.

One reason for the comparable performance of FL might be that the weight updates and the process of FedAVG could have a similar effect in mini-batches [46-48] and ensembles [49]. In FL, each client trains on a relatively small dataset and then transfers the local weights to the server. The server then collects

the local weights and updates the global weights that reflect all of the data through FedAVG. Subsequently, the round is repeated to improve the global weights. Hence, individual clients are an element of a mini-batch, and FedAVG is similar to ensemble processing. When implementing FL, we used the widely known FedAVG aggregation method [5], but this does not guarantee the best choice. To solve this problem, many researchers have studied aggregation methods that can work well with abnormal distributions, robust aggregation, and efficient communication such as FedProx [16], FSVRG [17], CO-OP [18], LoAdaBoost FedAVG [19], and RFA [20]. Hyperparameter selection also requires further research.

In addition, many researchers have studied methods to reduce communication costs. First, it has been suggested to reduce the communication round through methods such as client selection, peer-to-peer, and local update [11-13]. Second, a method such as sparsification, subsampling, or quantization has been suggested to reduce the communication message size [12,13]. Third, the asynchronous update method in traditional parallel computation methods can be applied.

FL can be used to build medical artificial intelligence apps by protecting patient privacy. Although the data themselves are not exposed or gathered in the central repository in FL, these data can nevertheless be guessed during the aggregation process in the network [12]. Therefore, other privacy preservation methods such as differential privacy, secure multiparty computation, and homomorphic encryption [11,12,21,22] might be necessary to protect privacy from diverse up-to-date privacy attack methods.

In future studies, we plan to use the proposed FL methods in real clinical datasets rather than benchmark datasets. First, we will try to improve the FL framework based on the results from this study. We will then compare the performance of a breast cancer recurrence prediction model using data from two different medical centers in Korea.

## Conclusions

Our experiments demonstrated the potential of FL in terms of performance and data protection, which is important for dealing with sensitive medical data. Specifically, in FL, only weights are transferred, and the participants are unaware of each other's local datasets. This can prevent personal information leaks. In addition, the proposed approach can be used to supplement existing approaches and to avoid problems that may occur during the deidentification process. The future direction of research is to use FL for actual medical data through collaborations with multiple institutions. Tasks such as expanding the client-server version of FL and improving communication will be expected to be important for the application of FL in real-world medical data with multiple institutions.

## Acknowledgments

This work was supported by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (2018-0-00861, Intelligent SW Technology Development for Medical Data Analysis), and by the Fourth Stage of Brain Korea 21 Project (Department of Intelligent Precision Healthcare Convergence) in 2021.



## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Client-server communication logic.

[[PDF File \(Adobe PDF File\), 324 KB - jmir\\_v22i10e20891\\_app1.pdf](#)]

### Multimedia Appendix 2

Confusion matrices for the MNIST experiments.

[[PDF File \(Adobe PDF File\), 303 KB - jmir\\_v22i10e20891\\_app2.pdf](#)]

### Multimedia Appendix 3

Accuracy changes for each round of MNIST federated learning (FL) experiments.

[[PDF File \(Adobe PDF File\), 353 KB - jmir\\_v22i10e20891\\_app3.pdf](#)]

### Multimedia Appendix 4

Each digit class classification results of precision and recall in a centralized machine learning (CML) experiment using the MNIST dataset.

[[PDF File \(Adobe PDF File\), 230 KB - jmir\\_v22i10e20891\\_app4.pdf](#)]

### Multimedia Appendix 5

Each digit class classification result of precision and recall in the basic federated learning (FL) experiment using the MNIST dataset.

[[PDF File \(Adobe PDF File\), 230 KB - jmir\\_v22i10e20891\\_app5.pdf](#)]

### Multimedia Appendix 6

Each digit class classification result of precision and recall in the imbalanced federated learning (FL) experiment using the MNIST dataset.

[[PDF File \(Adobe PDF File\), 229 KB - jmir\\_v22i10e20891\\_app6.pdf](#)]

### Multimedia Appendix 7

Each digit class classification result of precision and recall in the skewed federated learning (FL) experiment using the MNIST dataset.

[[PDF File \(Adobe PDF File\), 229 KB - jmir\\_v22i10e20891\\_app7.pdf](#)]

### Multimedia Appendix 8

Each digit class classification result of precision and recall in an imbalanced and skewed federated learning (FL) experiment using the MNIST dataset.

[[PDF File \(Adobe PDF File\), 249 KB - jmir\\_v22i10e20891\\_app8.pdf](#)]

### Multimedia Appendix 9

Area under the receiver operating characteristic curve, which is the result of the in-hospital mortality prediction for each experiment.

[[PDF File \(Adobe PDF File\), 255 KB - jmir\\_v22i10e20891\\_app9.pdf](#)]

### Multimedia Appendix 10

Confusion matrix of federated learning (FL) to predict in-hospital mortality using MIMIC-III. (A) Basic FL. (B) Imbalanced FL.

[[PDF File \(Adobe PDF File\), 227 KB - jmir\\_v22i10e20891\\_app10.pdf](#)]

### Multimedia Appendix 11

Confusion matrices for the ECG experiments.

[[PDF File \(Adobe PDF File\), 251 KB - jmir\\_v22i10e20891\\_app11.pdf](#)]

### Multimedia Appendix 12

Each class classification result of precision and recall in a centralized machine learning (CML) experiment using the ECG dataset.

[[PDF File \(Adobe PDF File\), 229 KB - jmir\\_v22i10e20891\\_app12.pdf](#)]

## Multimedia Appendix 13

Each class classification result of precision and recall in the basic federated learning (FL) experiment using the ECG dataset.

[PDF File (Adobe PDF File), 229 KB - [jmir\\_v22i10e20891\\_app13.pdf](#)]

## Multimedia Appendix 14

Each class classification result of precision and recall in the imbalanced federated learning (FL) experiment using the ECG dataset.

[PDF File (Adobe PDF File), 229 KB - [jmir\\_v22i10e20891\\_app14.pdf](#)]

## References

1. A Study on Global Data Leaks in 2018. InfoWatch Analytics Center. URL: [https://infowatch.com/sites/default/files/report/analytics/Global\\_Data\\_Breaches\\_2018.pdf](https://infowatch.com/sites/default/files/report/analytics/Global_Data_Breaches_2018.pdf) [accessed 2020-10-13]
2. Varunkumar K, Prabakaran M, Kaurav A, Chakkaravarthy S, Thiagarajan S, Venkatesh P. Various Database Attacks and its Prevention Techniques. *Int J Eng Trends Technol* 2014 Mar 25;9(11):532-536. [doi: [10.14445/22315381/ijett-v9p302](#)]
3. Emam KE, Dankar FK, Vaillancourt R, Roffey T, Lysyk M. Evaluating the Risk of Re-identification of Patients from Hospital Prescription Records. *Can J Hosp Pharm* 2009 Jul;62(4):307-319 [FREE Full text] [doi: [10.4212/cjhp.v62i4.812](#)] [Medline: [22478909](#)]
4. Konečný J, McMahan H, Yu F, Richtárik P, Suresh A, Bacon D. Federated learning: strategies for improving communication efficiency. *arXiv preprint 2016:1610.05492* [FREE Full text]
5. McMahan H, Moore E, Ramage D, Hampson S. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint 2016:1602.05629* [FREE Full text]
6. Brendan H. Federated Learning: Collaborative Machine Learning without Centralized Training Data. Google AI Blog. 2017. URL: <https://ai.googleblog.com/2017/04/federated-learning-collaborative.html> [accessed 2020-10-13]
7. Hard A, Rao K, Mathews R, Ramaswamy S, Beaufays F, Augenstein S. Federated learning for mobile keyboard prediction. *arXiv preprint 2018:1811.03604* [FREE Full text]
8. Yang T, Andrew G, Eichner H, Sun H, Li W, Kong N. Applied federated learning: Improving google keyboard query suggestions. *arXiv preprint 2018:1812.02903* [FREE Full text]
9. Chen M, Mathews R, Ouyang T, Beaufays F. Federated learning of out-of-vocabulary words. *arXiv preprint 2019:1903.10635* [FREE Full text]
10. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digit Med* 2020;3:119. [doi: [10.1038/s41746-020-00323-1](#)] [Medline: [33015372](#)]
11. Kairouz K, McMahan H. Advances and Open Problems in Federated Learning. *arXiv preprint 2019:1912.04977* [FREE Full text]
12. Li T, Sahu AK, Talwalkar A, Smith V. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Process Mag* 2020 May;37(3):50-60. [doi: [10.1109/msp.2020.2975749](#)]
13. Jie X, Benjamin S, Glicksberg, Chang S, Peter W, Jiang B, et al. Federated learning for healthcare informatics. *arXiv preprint 2019:1911.06270* [FREE Full text]
14. Qinbin L, Zeyi W, Bingsheng H. A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection. *arXiv preprint 2019:1907.09693* [FREE Full text]
15. Zhao Y, Meng L, Liangzhen L, Naveen S, Damon C, Vikas C. Federated Learning with Non-IID Data. *arXiv preprint 2018:1806.00582* [FREE Full text]
16. Li T, Sahu A, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated Optimization in Heterogeneous Networks. *arXiv preprint 2018:1812.06127* [FREE Full text]
17. Konečný J, McMahan H. Federated Optimization: Distributed Machine Learning for On-Device Intelligence. *arXiv preprint 2016:1610.02527* [FREE Full text]
18. Wang Y. Cooperative Machine Learning from Mobile Devices Masters Thesis. University of Alberta. 2017. URL: <https://era.library.ualberta.ca/items/7d680f04-7987-45c5-b9cd-4fe43c87606f> [accessed 2020-10-15]
19. Huang L, Yin Y, Fu Z, Zhang S, Deng H, Liu D. LoAdaBoost: loss-based AdaBoost federated machine learning with reduced computational complexity on IID and non-IID intensive care data. *arXiv preprint 2018:1811.12629* [FREE Full text]
20. Pillutla K, Kakade S, Harchaoui Z. Robust Aggregation for Federated Learning. *arXiv preprint 2019:1912.13445* [FREE Full text]
21. Abadi M, Chu A, Goodfellow I, McMahan H, Mironov I, Talwar K, et al. Deep Learning with Differential Privacy. 2016 Presented at: CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security; 2016; Vienna. [doi: [10.1145/2976749.2978318](#)]
22. Pfohl Stephen R, Andrew M, Katherine H. Federated and Differentially Private Learning for Electronic Health Records. *arXiv preprint 2019:1911.05861* [FREE Full text]
23. Li W. Privacy-Preserving Federated Brain Tumour Segmentation. : Springer, Cham; 2019 Presented at: Machine Learning in Medical Imaging. MLMI 2019; 2019; Shenzhen. [doi: [10.1007/978-3-030-32692-0\\_16](#)]

24. Sheller MJ, Reina GA, Edwards B, Martin J, Bakas S. Multi-Institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation. 2019 Presented at: International MICCAI Brain Lesion Workshop; 2018; Granada p. 92-104 URL: <http://europepmc.org/abstract/MED/31231720> [doi: [10.1007/978-3-030-11723-8\\_9](https://doi.org/10.1007/978-3-030-11723-8_9)]
25. Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated Electronic Health Records. *Int J Med Inform* 2018 Apr;112:59-67 [FREE Full text] [doi: [10.1016/j.ijmedinf.2018.01.007](https://doi.org/10.1016/j.ijmedinf.2018.01.007)] [Medline: [29500022](https://pubmed.ncbi.nlm.nih.gov/29500022/)]
26. El Emam K, Dankar FK, Issa R, Jonker E, Amyot D, Cogo E, et al. A globally optimal k-anonymity method for the de-identification of health data. *J Am Med Inform Assoc* 2009;16(5):670-682 [FREE Full text] [doi: [10.1197/jamia.M3144](https://doi.org/10.1197/jamia.M3144)] [Medline: [19567795](https://pubmed.ncbi.nlm.nih.gov/19567795/)]
27. Taira RK, Bui AA, Kangarloo H. Identification of patient name references within medical documents using semantic selectional restrictions. *Proc AMIA Symp* 2002:757-761 [FREE Full text] [Medline: [12463926](https://pubmed.ncbi.nlm.nih.gov/12463926/)]
28. Thomas SM, Mamlin B, Schadow G, McDonald C. A successful technique for removing names in pathology reports using an augmented search and replace method. *Proc AMIA Symp* 2002:777-781 [FREE Full text] [Medline: [12463930](https://pubmed.ncbi.nlm.nih.gov/12463930/)]
29. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. HHS.gov. Washington, DC: U.S. Department of Health & Human Services; 1996. URL: <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html> [accessed 2020-10-13]
30. Nass SJ, Levit LA, Gostin LO. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington, DC: National Academies Press; 2009.
31. TensorFlow Federated: Machine Learning on Decentralized Data. TensorFlow. URL: <https://www.tensorflow.org/federated> [accessed 2020-10-13]
32. Ryffel T, Trask A, Dahl M, Wagner B, Mancuso J, Rueckert D. A generic framework for privacy preserving deep learning. *arXiv preprint* 2018:1811.04017 [FREE Full text]
33. OpenMined. URL: <https://www.openmined.org/> [accessed 2020-10-13]
34. An Industrial Grade Federated Learning Framework. FATE. URL: <https://fate.fedai.org/> [accessed 2020-10-13]
35. FederatedAI. GitHub. URL: <https://github.com/FederatedAI/FATE> [accessed 2020-10-13]
36. FL\_Server. GitHub. URL: [https://github.com/bmiskku/FL\\_Server](https://github.com/bmiskku/FL_Server) [accessed 2020-10-13]
37. FL\_Client. GitHub. URL: [https://github.com/bmiskku/FL\\_Client](https://github.com/bmiskku/FL_Client) [accessed 2020-10-13]
38. Harutyunyan H. MIMIC-III benchmarks. GitHub. 2018. URL: <https://github.com/YerevaNN/mimic3-benchmarks> [accessed 2020-10-13]
39. Johnson A, Pollard T, Mark R. MIMIC-III Clinical Database. Physionet. URL: <https://physionet.org/content/mimiciii/1.4/> [accessed 2020-10-13]
40. AF Classification from a Short Single Lead ECG Recording - The PhysioNet Computing in Cardiology Challenge 2017. PhysioNet. 2017. URL: <https://physionet.org/content/challenge-2017/1.0.0/> [accessed 2020-10-13]
41. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019 Jan;25(1):65-69 [FREE Full text] [doi: [10.1038/s41591-018-0268-3](https://doi.org/10.1038/s41591-018-0268-3)] [Medline: [30617320](https://pubmed.ncbi.nlm.nih.gov/30617320/)]
42. Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, Ivanov V. Towards Federated Learning at Scale: System Design. *arXiv preprint* 2019:1902.01046 [FREE Full text]
43. Ruder S. An overview of gradient descent optimization algorithms. *arXiv preprint* 2016:1609.04747 [FREE Full text]
44. Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. *Sci Data* 2019 Jun 17;6(1):96. [doi: [10.1038/s41597-019-0103-9](https://doi.org/10.1038/s41597-019-0103-9)] [Medline: [31209213](https://pubmed.ncbi.nlm.nih.gov/31209213/)]
45. Awni ECG. GitHub. URL: <https://github.com/awni/ecg> [accessed 2020-10-13]
46. Li M, Zhang T, Chen Y, Smola AJ. Efficient mini-batch training for stochastic optimization. 2014 Presented at: KDD '14: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2014; New York p. 661-670. [doi: [10.1145/2623330.2623612](https://doi.org/10.1145/2623330.2623612)]
47. Keskar N, Mudigere D, Nocedal J, Smelyanskiy M, Tang P. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. *arXiv preprint* 2016:1609.04836 [FREE Full text]
48. Masters D, Luschi C. Revisiting Small Batch Training for Deep Neural Networks. *arXiv preprint* 2018:1804.07612 [FREE Full text]
49. Dietterich T. Ensemble Methods in Machine Learning. : Springer; 2000 Presented at: International Workshop on Multiple Classifier Systems. MCS 2000; 2000; Berlin, Heidelberg. [doi: [10.1007/3-540-45014-9\\_1](https://doi.org/10.1007/3-540-45014-9_1)]

## Abbreviations

- ADAM:** adaptive moment estimation  
**API:** application programming interface  
**AUPRC:** area under the precision-recall curve  
**AUROC:** area under the receiver operating characteristic curve

**AWS:** Amazon Web Services

**CML:** centralized machine learning

**ECG:** electrocardiogram

**FedAVG:** federated averaging

**FL:** federated learning

**LSTM:** long short-term memory

**MIMIC-III:** Medical Information Mart for Intensive Care III

**MNIST:** Modified National Institute of Standards and Technology

**SOTA:** state of the art

*Edited by G Eysenbach; submitted 01.06.20; peer-reviewed by JH Yoon, M Abdalla; comments to author 22.06.20; revised version received 16.08.20; accepted 02.10.20; published 26.10.20.*

*Please cite as:*

*Lee GH, Shin SY*

*Federated Learning on Clinical Benchmark Data: Performance Assessment*

*J Med Internet Res 2020;22(10):e20891*

URL: <http://www.jmir.org/2020/10/e20891/>

doi: [10.2196/20891](https://doi.org/10.2196/20891)

PMID: [33104011](https://pubmed.ncbi.nlm.nih.gov/33104011/)

©Geun Hyeong Lee, Soo-Yong Shin. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Accuracy of Nutrient Calculations Using the Consumer-Focused Online App MyFitnessPal: Validation Study

Charlotte Evenepoel<sup>1</sup>, BSc, MSc; Egbert Clevers<sup>1</sup>, BSc, MSc, PhD; Lise Deroover<sup>1</sup>, BSc, MSc, PhD; Wendy Van Loo<sup>1</sup>, BA; Christophe Matthys<sup>2,3,4</sup>, BSc, MSc, PhD; Kristin Verbeke<sup>1,4</sup>, BSc, MSc, PhD

<sup>1</sup>Translational Research in Gastrointestinal Disorders, Department of Chronic Diseases, Metabolism and Aging, KU Leuven, Leuven, Belgium

<sup>2</sup>Nutrition & Obesity Unit, Clinical and Experimental Endocrinology, Department of Chronic Diseases, Metabolism and Aging, KU Leuven, Leuven, Belgium

<sup>3</sup>Department of Endocrinology, University Hospitals Leuven, KU Leuven, Leuven, Belgium

<sup>4</sup>Leuven Food Science and Nutrition Research Center, KU Leuven, Leuven, Belgium

**Corresponding Author:**

Kristin Verbeke, BSc, MSc, PhD

Translational Research in Gastrointestinal Disorders

Department of Chronic Diseases, Metabolism and Aging

KU Leuven

Herestraat 49

TARGID, O&N1, box 701

Leuven, 3000

Belgium

Phone: 32 16330150

Email: [kristin.verbeke@kuleuven.be](mailto:kristin.verbeke@kuleuven.be)

## Abstract

**Background:** Digital food registration via online platforms that are coupled to large food databases obviates the need for manual processing of dietary data. The reliability of such platforms depends on the quality of the associated food database.

**Objective:** In this study, we validate the database of MyFitnessPal versus the Belgian food composition database, Nubel.

**Methods:** After carefully given instructions, 50 participants used MyFitnessPal to each complete a 4-day dietary record 2 times (T1 and T2), with 1 month in between T1 and T2. Nutrient intake values were calculated either manually, using the food composition database Nubel, or automatically, using the database coupled to MyFitnessPal. First, nutrient values from T1 were used as a training set to develop an algorithm that defined upper limit values for energy intake, carbohydrates, fat, protein, fiber, sugar, cholesterol, and sodium. These limits were applied to the MyFitnessPal dataset extracted at T2 to remove extremely high and likely erroneous values. Original and cleaned T2 values were correlated with the Nubel calculated values. Bias was estimated using Bland-Altman plots. Finally, we simulated the impact of using MyFitnessPal for nutrient analysis instead of Nubel on the power of a study design that correlates nutrient intake to a chosen outcome variable.

**Results:** Per food portion, the following upper limits were defined: 1500 kilocalories for total energy intake, 95 grams (g) for carbohydrates, 92 g for fat, 52 g for protein, 22 g for fiber, 70 g for sugar, 600 mg for cholesterol, and 3600 mg for sodium. Cleaning the dataset extracted at T2 resulted in a 2.8% rejection. Cleaned MyFitnessPal values demonstrated strong correlations with Nubel for energy intake ( $r=0.96$ ), carbohydrates ( $r=0.90$ ), fat ( $r=0.90$ ), protein ( $r=0.90$ ), fiber ( $r=0.80$ ), and sugar ( $r=0.79$ ), but weak correlations for cholesterol ( $p=0.51$ ) and sodium ( $p=0.53$ ); all  $P$  values were  $\leq 0.001$ . No bias was found between both methods, except for a fixed bias for fiber and a proportional bias for cholesterol. A 5-10% power loss should be taken into account when correlating energy intake and macronutrients obtained with MyFitnessPal to an outcome variable, compared to Nubel.

**Conclusions:** Dietary analysis with MyFitnessPal is accurate and efficient for total energy intake, macronutrients, sugar, and fiber, but not for cholesterol and sodium.

(*J Med Internet Res* 2020;22(10):e18237) doi:[10.2196/18237](https://doi.org/10.2196/18237)

**KEYWORDS**

dietary assessment; MyFitnessPal; Nubel; nutrition; online application; diet



## Introduction

Analysis of dietary records is an important part of nutritional and epidemiological research investigating the effects of diet on human health. Food composition tables are used to convert recorded food intake into nutrient intake. Individual food items from the dietary record are matched with an entry in the food composition table so that nutrient information can be extracted. This matching process is typically done manually and is therefore time- and labor-intensive, especially in large-scale studies or studies that require repeated dietary assessment over time [1].

Over the last years, a large number of mobile apps have been developed to track lifestyle habits such as physical activity, sleep, time management, and dietary intake. Such dietary intake digital platforms contain an online food database that immediately converts food items into nutrient values. With these online platforms, users electronically record their food intake and track the amount of calories and nutrients consumed. To do so, users select consumed food items from the associated food database or, in case an item is not registered, they add that item and the related nutritional information to the platform's database. Therefore, these databases are partially user-based. Generally, these platforms have been developed to promote weight loss by increasing the awareness of habits and progress through self-monitoring; however, they also appeal to nutritional research applications as well. Yet, the accuracy and reliability of these platforms rely on the quality of the associated databases.

The aim of this study is to evaluate the accuracy of nutrient intake values based on the database of MyFitnessPal [2]. MyFitnessPal is the most popular commercial nutrition weight loss app, with more than 165 million users in 2016 [3]. In the United States, 83% of the dietitians who participated in a survey held at the Food and Nutrition Conference and Expo 2015 recommended the use of nutrition and health-related apps, with MyFitnessPal and Fitbit mentioned most [4,5]. Furthermore, in a survey across the United Kingdom, Australia, Canada, New Zealand, and the United States, it was the preferred app for sports dietitians [6]. MyFitnessPal is characterized by its extensive and country-specific database of over 6 million food items and brands. This consumer app is user-friendly, contains a barcode scanner to rapidly add packed foods, and has both smartphone and web-based versions. In addition, the basic version of MyFitnessPal is freely available. We extracted values for energy intake and nutrient composition from dietary records entered in MyFitnessPal and compared them to the values calculated using the Belgian food composition database, Nubel [7].

## Methods

### Study Sample

This study used data from a previous intervention study that aimed to evaluate the impact of modified wheat bran on microbial fermentation and intestinal health [8]. We used 100 dietary records from 50 participants. Each participant used MyFitnessPal to complete a 4-day dietary record 2 times, with 1 month between the 2 records. These 2 time points generated

2 dietary intake datasets, referred to as T1 and T2. All participants used MyFitnessPal to record their dietary intake.

### Dietary Intake Assessment

Each participant received an account on MyFitnessPal and detailed instructions with illustrations about the use of MyFitnessPal. The manual included information on how to (1) select food items in the MyFitnessPal database, (2) indicate portion size, (3) register home recipes, and (4) use the favorites lists. To register packed food items, participants were instructed to scan the attached barcode or to select the corresponding item (correct brand) from the food item list within MyFitnessPal. If the item was not included in the list, participants were asked to enter the missing food item, including the nutritional data mentioned on the package. For generic food items such as bread, rice, pasta, fruits, and vegetables, we preregistered items in the MyFitnessPal database with the tag "Targid" (ie, the name of our research unit). These items contained nutritional information from the Belgian food composition database, Nubel. This preregistration of Targid-tagged food items was done for standardization purposes, as generic items often have multiple entries in MyFitnessPal with highly variable nutritional information. Participants were asked to preferentially select these items, and in case the searched item was not available in the Targid list, they were instructed to choose a green-flagged item. The green flag indicates that administrators of MyFitnessPal have revised the nutritional information. With regard to portion size, participants were instructed to weigh the consumed food items as much as possible, and when they were unable to do so, they were asked to select from listed portion sizes.

### Dietary Nutrient Quantification

Food items registered in MyFitnessPal were converted to nutrient values (ie, total energy intake, carbohydrates, fat, protein, sugars, fiber, sodium, and cholesterol), either using the MyFitnessPal database or using Nubel. The Nubel database contains the composition of 1194 basic food items, each product expressed per 100 grams. The MyFitnessPal food records are directly linked to the food composition data available in the MyFitnessPal app. Therefore, downloading the food records from the MyFitnessPal account immediately yields nutrient intake values. Conversion of the MyFitnessPal-registered food items to nutrient values using the Nubel database was performed according to a standard in-house procedure. Registered food items in MyFitnessPal with a food match in the Nubel database (eg, an apple, lasagne) were translated to nutrients using the Nubel food composition information. Homemade food without a match in the Nubel database (eg, Belgian endive with ham and cheese sauce) was translated to nutrients using the recipes described in a basic and comprehensive Belgian cookbook named *Ons Kookboek* [9]. For packed food, the nutritional information on the food label was used. If portion size was available in grams, nutrient values per consumed portion were calculated. Portion sizes defined in measurements (eg, "a slice of," "a piece of") were converted to grams using guidelines on the standardized quantification of food products [10].

## Data Cleaning

As MyFitnessPal is partially user-based and thus inherently prone to errors, we developed an algorithm that removes extremely high nutrient values very likely to be erroneous using Monte Carlo simulations [11]. Using the T1 dataset as a training set, we determined an upper limit of intake per food portion (using an in-house R script) for each nutrient. These limit values were obtained for each nutrient by iteratively increasing the putative limit value and including only intake values below the limit in the database (Multimedia Appendix 1). For each iteration, the correlation between the included Nubel and MyFitnessPal values was calculated. The nutrient intake value for which this correlation was maximal was defined as the upper limit for that nutrient. These cut-offs constitute the best compromise between removing erroneous values (ie, true positives) and not removing correct values (ie, false positives). Subsequently, these limit values based on the T1 dataset were applied to the MyFitnessPal values of the T2 dataset to remove extremely high values, using an in-house R script (Multimedia Appendix 2).

## Data Analysis

For each 4-day diary, mean energy intake and nutrient values were calculated with Nubel and the MyFitnessPal database. To evaluate the impact of the data cleaning procedure, both the original and cleaned T2 data from MyFitnessPal were correlated to the Nubel values. The normality of the data was checked with the Shapiro-Wilk test and with visual inspection of the residual histogram. Depending on normality, Pearson or Spearman rank were applied for correlation analysis. Bland-Altman plots were used to assess the degree of agreement between both methods and to evaluate bias [12]. In addition, a paired *t* test or a Wilcoxon signed rank test was performed. Data analysis was performed using SAS software (version 9.4; SAS Institute). *P* values less than .05 were considered statistically significant.

To further assess the practical implication of using the MyFitnessPal database, we simulated the loss of statistical power that occurs when using the MyFitnessPal database compared to the Nubel database in a study design that aims to correlate nutrient intake to a random outcome variable. In addition, we calculated the corresponding increase in sample size required to compensate for the loss in power. Given a nutrient quantified using Nubel and MyFitnessPal, named *N* and *M*, respectively, and given a variable defined as the outcome (*O*), we used the correlation *NM* (known from this study) and a simulated correlation *NO* to retrieve the correlation of interest (ie, *MO*). Simulations were performed according to Monte Carlo simulations [11]. Briefly, for each nutrient, we simulated 100 times a value for *O* and this for 100 correlations, with  $0 \leq \rho \leq 0.5$  and a step length of 0.005. Subsequently, for each correlation between *O* and the Nubel-calculated nutrient *N*, and with a known sample size, we derived the power for *NO*. Corresponding correlations between the *O* variables and the MyFitnessPal-calculated nutrients *M* were computed, followed by the resulting power of *MO*. The power of *MO* indicates the power to detect a true effect for the correlation of the MyFitnessPal nutrient data (*M*) and the outcome variable (*O*) (eg, a health parameter). This simulation was performed for a

sample size of 50, 100, and 500 observations. All simulation analyses were performed with R (version 3.5.1; R Core Team). Power calculations were performed with R package “pwr” (version 1.3-0; Stéphane Champely).

## Results

Of the 50 participants, 78% (39/50) were women and 22% (11/50) were men. Mean BMI was 25.3 (SD 5.1) kg/m<sup>2</sup> and mean age was 28.2 (SD 11.3) years. All dietary records covered 4 consecutive days, except for 1 diary with only 3 days of food registration. The upper limit values per food portion were 1500 kilocalories (kcal) for total energy intake, 95 grams (g) for carbohydrates, 92 g for fat, 52 g for protein, 22 g for fiber, 70 g for sugar, 600 mg for cholesterol, and 3600 mg for sodium. The clean-up of dataset T2 removed certain nutrient values from 79 of the 2826 recorded food items (2.8%). For carbohydrates, 46 values were removed; for protein, 17 values were removed; for fat, 2 values were removed; for sugar, 8 values were removed; for cholesterol and sodium, 3 values were removed. No values for fiber and total energy intake were removed.

In the original T2 dataset, strong positive correlations between Nubel and MyFitnessPal were obtained for total energy intake, the 3 macronutrients, and for sugar and fiber (Table 1). Nevertheless, fiber intake was significantly underestimated when calculated with the MyFitnessPal database ( $P < .001$ ). Correlations for cholesterol and sodium, however, were weak, and their intake was also strongly underestimated by MyFitnessPal. After the cleaning of the T2 dataset, correlations with Nubel data were stronger for carbohydrates, fat, sugar, and sodium than for the original data (Table 1). Mean energy and fat intake did not differ significantly between cleaned MyFitnessPal and Nubel values, whereas carbohydrate and protein intake became slightly but significantly lower using MyFitnessPal. Correlations for original and cleaned T2 MyFitnessPal data with Nubel were identical for fiber and total energy intake, as data cleaning did not remove any value.

Bland-Altman plots of the cleaned T2 data displaying the agreement between both methods for all nutrients and energy intake are shown in Figure 1. No proportional nor fixed bias was observed for the macronutrients, sugar, and energy intake. In contrast, fiber intake showed an average fixed bias of about 4 g/day, which is about 20% of average fiber intake. Furthermore, sodium and cholesterol intake were clearly underreported in MyFitnessPal [the mean difference MyFitnessPal-Nubel amounts to -1345 (SD 241) mg/day for sodium and -187 (SD 124) mg/day for cholesterol]. In addition, the difference between MyFitnessPal and Nubel for cholesterol intake proportionally increased with increasing cholesterol intake.

The power simulation quantified the statistical power that is lost when correlating MyFitnessPal-derived data instead of Nubel data to an outcome parameter (Figure 2) and the additional sample size needed to compensate (Table 2). For example, for an intended power of 80% with Nubel, the power will be decreased when using MyFitnessPal with 4% for total energy intake, 8% for carbohydrates, 10% for protein and fat, 15% for sugar, and 19% for fiber. Depending on the intended

sample size, the required increase in sample size to compensate for this loss in power ranged from 10% to 65% (Table 2). Furthermore, the simulation showed a complete loss of power if MyFitnessPal would be used to assess cholesterol and sodium intake, resulting in extremely high sample sizes.

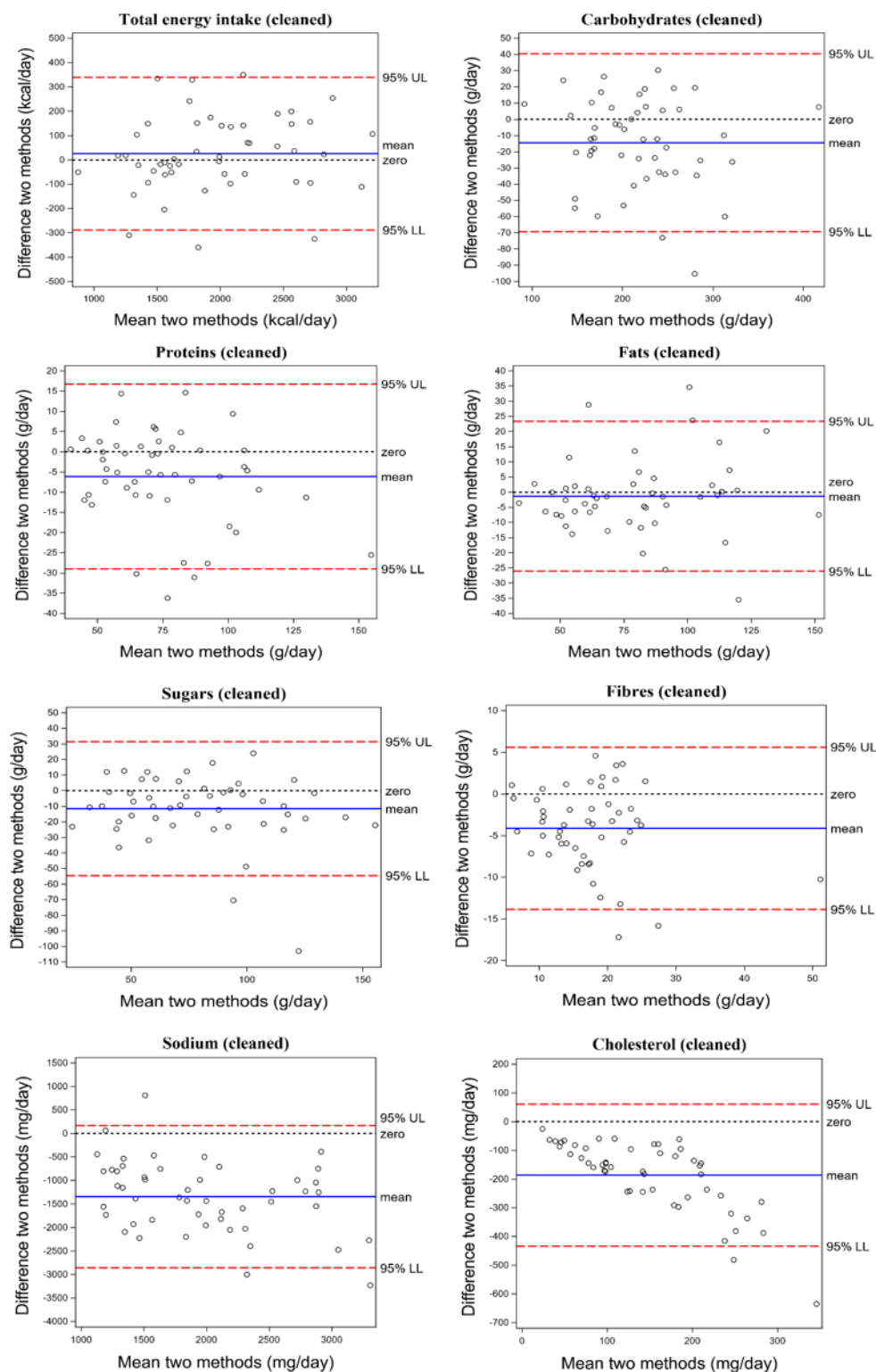
**Table 1.** Nutrient intake values of T2 data derived from Nubel and MyFitnessPal, prior to and after data cleaning (n=50).

Nutrient	Nubel		Original MyFitnessPal data				Cleaned MyFitnessPal data			
	Mean	SD	Mean	SD	<i>P</i> value <sup>a</sup>	Coefficient <sup>b</sup>	Mean	SD	<i>P</i> value <sup>a</sup>	Coefficient <sup>b</sup>
Energy intake (kcal/day)	1958	543	1984	567	.26	0.96	1984	567	.26	0.96
Carbohydrates (g/day)	225	61	241	80	.05	0.70	210	58	<.001	0.90
Fat (g/day)	81	27	82	31	.69	0.75	80	28	.44	0.90
Protein (g/day)	78	26	78	26	.85	0.94	72	23	<.001	0.90
Sugar (g/day)	85	35	80	40	.22	0.70	74	31	<.001	0.79
Fiber (g/day)	19	8	15	7	<.001	0.80	15	7	<.001	0.80
Sodium (mg/day)	2638	848	1551	1765	<.001	0.45	1293	607	<.001	0.53
Cholesterol (mg/day)	242	131	67	63	<.001	0.67	55	49	<.001	0.51

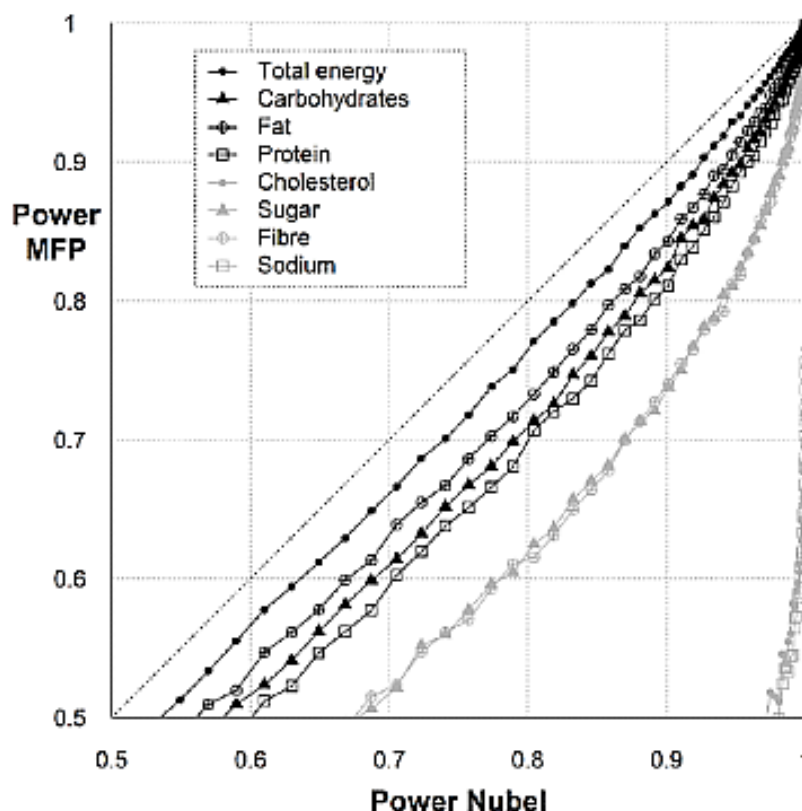
<sup>a</sup>Paired *t* test was used for energy intake, macronutrients, sugar, and fiber; the Wilcoxon signed rank test was used for for cholesterol and sodium.

<sup>b</sup>Pearson correlation coefficients were used for energy intake, macronutrients, sugar, and fiber; the Spearman rank correlation coefficients were used for cholesterol and sodium. Both Pearson and Spearman were significant at the level of .05, as they were all *P*<.001.

**Figure 1.** Bland-Altman plots for energy intake and nutrient values of the cleaned T2 dataset, with Nubel as the reference method and MyFitnessPal as the other method for nutrient intake analysis. The difference between the 2 methods is calculated as follows: MyFitnessPal – Nubel. The 95% upper limit (UL) and lower limit (LL) of agreement (SD 1.96) are depicted as long dashed lines. The full line and short-dashed line indicate the mean difference and zero, respectively.



**Figure 2.** Correlation analysis between the statistical power of Nubel and MyFitnessPal (MFP) to reject the null hypothesis that states there is no correlation between each of these methods and a simulated variable outcome. The total sample size is 100 power values. This correlation between the power of Nubel and MyFitnessPal was performed for all studied nutrients (macronutrients, sugar, fiber, cholesterol, and sodium) and for energy intake.



**Table 2.** Increase in sample size (%) required to maintain 80% power when using MyFitnessPal for nutrient analysis instead of Nubel to detect a true effect if correlated with an outcome variable.

Nubel sample size	Percentage (%) increase in MyFitnessPal sample size required to maintain a statistical power of 80%							
	Energy intake	Carbohydrate	Fat	Protein	Sugar	Fiber	Cholesterol	Sodium
50	11	28	21	33	64	65	347	364
100	10	27	20	32	60	65	277	307
500	10	25	19	28	40	36	68	72

## Discussion

### Principal Findings

This study evaluated the accuracy of the MyFitnessPal food composition database to assess energy and macro- and micronutrient intake. Strong positive correlations between MyFitnessPal and Nubel were observed for energy intake, macronutrients, sugar, and fiber, but not for cholesterol and sodium. Bland-Altman plots displayed a good agreement between MyFitnessPal and Nubel, except for cholesterol and sodium. Using MyFitnessPal over Nubel for quantifying total energy intake and macronutrients led to a power loss in studies that correlate nutrient values to an outcome variable of not more than 10%, which can be compensated by increasing the sample size.

Digital collection of food intake data facilitates the assessment of dietary intake, both at the level of dietary data registration and at the level of nutrient quantification [1,13]. Many studies

have compared digital methods with paper-and-pencil methods at the level of dietary registration using the same database for nutrient calculation. These studies, using digital registration tools such as My Meal Mate, YANA-C, Wellnavi, e-DIA or Easy Diet Diary, observed good agreement between the digital and paper method [14-18]. In this study, we used the same digital method for diet registration but applied 2 distinct databases to calculate the nutrient intake values. Our results indicate that the nutritional information extracted from MyFitnessPal is comparable to the information calculated with a standard food composition database (Nubel), except for cholesterol and sodium. Overall, dietary intake was slightly underestimated by MyFitnessPal, which is in agreement with previous studies [19,20]. The underestimation was even more pronounced after cleaning the dataset, a procedure that resulted in the rejection of extremely high and likely erroneous values. Compared to Nubel-derived values, MyFitnessPal underestimated protein intake by 7.8%, carbohydrate intake by 6.4%, and fat intake by 1.7%. In contrast, energy intake was



slightly overestimated (1.3%). In a study applying the Brazilian food composition data table as a reference, the MyFitnessPal database underestimated energy (0.7%), fat (16.8%), protein (11.9%), and carbohydrate intake (10.8%) [19]. In another study, nutrient intake estimates from thirty 24-hour dietary recalls collected using the Nutrition Data System for Research (NDSR) were compared with intake calculations from these recalls entered by the researcher into MyFitnessPal [20]. Compared to NDSR, a similar underestimation by MyFitnessPal was observed for energy (4.1%), fat (14.1%), and protein (8.0%), while carbohydrate intake was overestimated by 1.6% [20]. A reason for the underestimation by MyFitnessPal is most likely incomplete or missing information about nutrient composition for some food items in the database. Indeed, some entries in MyFitnessPal only have a value for total energy content without values for macronutrient composition or cholesterol and sodium content. Selecting such items for inclusion in the dietary record results in inaccurate information.

The use of digital platforms like MyFitnessPal has several advantages over other methods. Platforms available as a smartphone app allow participants to report food intake conveniently on their smartphones, regardless of their whereabouts or the occasion. In this way, the time delay between consuming and registering food data is reduced, which increases data quality [21]. Compared to food frequency questionnaires and 24-hour recalls, prospective food recording is less prone to memory bias [22,23]. Additionally, recording with a mobile app increases the satisfaction and adherence of the participants, further enhancing dietary assessment [24,25]. Moreover, the direct link of the consumed items to the nutritional facts facilitates and accelerates the work of the investigator. In contrast to paper data, MyFitnessPal food data do not require transfer to an electronic database, which reduces errors. In this way, the use of MyFitnessPal can reduce the high financial and human resources required for standard nutrient quantification [26]. A potential concern of digital platforms is the fact that the immediate feedback and nutritional information provided to participants may alter their eating behavior and induce misreporting. This is called the “reactivity effect” and should be taken into consideration when deciding on the use of a digital platform in a study design [27].

An important source of error in dietary assessment, independent of the registration method and database used, is the estimation of the portion size by the participant. A study that compared meals recorded by the participants with a personal digital assistant (PDA) against actual meals assessed by dietitians reported portion size to be the greatest source of error, accounting for 49% of the errors between recorded and actual meals [28]. Other major errors in food registration by participants included reporting incorrect food (25%) and omitting food (15%). In addition, in a study with the digital My Meal Mate diary, an incorrect portion size was the cause of most

errors, with users selecting the standard listed portion sizes rather than providing the true portions consumed [14].

An effective strategy to improve the accuracy of dietary assessment with digital platforms is to provide adequate training to the participants, in the form of a manual or a training session. A study of 78 adolescents showed a significant increase in proficiency and perception in the use of a mobile phone food record after additional training was provided [29]. Accordingly, the application of MyFitnessPal in a naturalistic setting (ie, without the provision of instructions to subjects unfamiliar with the app) resulted in poor correlations for 4-day mean energy and macronutrient intake (ranging from 0.21 to 0.42) compared to a 24-hour recall analyzed with the Australian Food, Supplement, and Nutrient Database (AUSNUT) 2011-2013 database [30]. Therefore, it is crucial that the training includes information on how to select items from the database and clear instructions on how to provide portion sizes. We assume that the high correlations between MyFitnessPal and Nubel found in this study are partly due to the extensive manual that was provided to the participants and the fact that we predefined a number of generic items to increase standardization. In daily life, this manual is not available, and as MyFitnessPal is a user-based platform, errors in nutritional information arise in the database. However, MyFitnessPal itself has built-in tools to enhance food registration, such as bar code scanning and the green flagging of items. If the user is motivated, MyFitnessPal may be a useful tool to track nutrient intake in daily life. Using MyFitnessPal as a self-guided approach was as effective in inducing weight loss as a combination of the app with dietary counseling in 100 obese subjects [31]. Another trial highlighted the importance of motivation when using MyFitnessPal for weight loss management in combination with dietary counseling [25]. In addition, it is important to realize that our study was performed in the context of a trial evaluating the health benefits of wheat bran. Therefore, participants were likely biased towards persons interested in human health. Good motivation and background knowledge on nutrition and food are known predictors for reliable and accurate reporting with digital dietary records [22]. Consequently, these predictors may have driven our study participants to select those food items with accurate nutritional information and may have directed our results.

## Conclusion

MyFitnessPal provides accurate estimates for energy, macronutrients, fiber, and sugar intake, but not for cholesterol and sodium, in a research setting. We advise the use of both a data cleaning procedure and a clear manual on dietary reporting for participants, as they contribute to better nutrient assessment. A manual on MyFitnessPal usage should include instructions on proper food item selection, portion size recording, and the use of features such as scanning bar codes of food items. To enhance standardization, it is useful to add entries with a tag for generic items, as the MyFitnessPal database contains many entries with considerable differences in nutritional information.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Flow chart representing the iterative process that was applied to define the upper limit for a nutrient intake value extracted from MyFitnessPal (MFP) using Monte Carlo simulations.

[PNG File, 17 KB - [jmir\\_v22i10e18237\\_app1.png](#)]

## Multimedia Appendix 2

In-house R script for nutrient data extraction from MyFitnessPal, including data cleaning.

[DOCX File, 20 KB - [jmir\\_v22i10e18237\\_app2.docx](#)]

## References

1. Ngo J, Engelen A, Molag M, Roesle J, García-Segovia P, Serra-Majem L. A review of the use of information and communication technologies for dietary assessment. *Br J Nutr* 2009 Jul;101 Suppl 2:S102-S112. [doi: [10.1017/S0007114509990638](#)] [Medline: [19594959](#)]
2. MyFitnessPal. URL: <https://www.myfitnesspal.com/> [WebCite Cache ID 75xFgNLZ2]
3. Jones D. How the MyFitnessPal app got 165 million users. URL: <https://www.linkedin.com/pulse/interview-how-myfitnesspal-app-got-165-million-users-david-jones> [accessed 2019-02-16]
4. Saucedo A, Frederico C, Pellechia K, Starin D. Results of the Academy of Nutrition and Dietetics' Consumer Health Informatics Work Group's 2015 Member App Technology Survey. *J Acad Nutr Diet* 2016 Aug;116(8):1336-1338. [doi: [10.1016/j.jand.2016.04.009](#)] [Medline: [27236643](#)]
5. FitBit. URL: <https://www.fitbit.com/> [WebCite Cache ID 6NLbJcpUC]
6. Jospe MR, Fairbairn KA, Green P, Perry TL. Diet app use by sports dietitians: a survey in five countries. *JMIR Mhealth Uhealth* 2015 Jan 22;3(1):e7 [FREE Full text] [doi: [10.2196/mhealth.3345](#)] [Medline: [25616274](#)]
7. Nubel. Belgian Food Composition Table. Brussels: The Ministry of Public Health; 2015.
8. ClinicalTrials.gov. Identifier: NCT02796989. The impact of modified wheat bran on carbohydrate fermentation in the colon in healthy and obese subjects. Bethesda (MD): U.S. National Library of Medicine; 2017 Dec. URL: <https://clinicaltrials.gov/ct2/show/NCT02796989> [accessed 2020-09-30]
9. KVLV. Ons Kookboek@. Wijgmaal: KVLV; 2013.
10. Hoge Gezondheidsraad. Maten en gewichten - Handleiding voor een gestandaardiseerde kwantificering van voedingsmiddelen in België. Brussels: Hoge Gezondheidsraad; 2005.
11. Harrison R. Introduction To Monte Carlo Simulation. In: Nuclear Physics Methods and Accelerators in Biology and Medicine: Fifth International Summer School on Nuclear Physics Methods and Accelerators in Biology and Medicine. 2010 Jan 05 Presented at: AIP Conference Proceedings; 2009; Bratislava p. 17-21 URL: <http://europemc.org/abstract/MED/20733932> [doi: [10.1063/1.3295638](#)]
12. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986 Feb 08;1(8476):307-310. [Medline: [2868172](#)]
13. Burke LE, Warziski M, Starrett T, Choo J, Music E, Sereika S, et al. Self-monitoring dietary intake: current and future practices. *J Ren Nutr* 2005 Jul;15(3):281-290. [doi: [10.1016/j.jrn.2005.04.002](#)] [Medline: [16007557](#)]
14. Carter MC, Burley VJ, Nykjaer C, Cade JE. 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. *Br J Nutr* 2013 Feb 14;109(3):539-546. [doi: [10.1017/S0007114512001353](#)] [Medline: [22717334](#)]
15. Vereecken CA, Covents M, Matthys C, Maes L. Young adolescents' nutrition assessment on computer (YANA-C). *Eur J Clin Nutr* 2005 May;59(5):658-667 [FREE Full text] [doi: [10.1038/sj.ejcn.1602124](#)] [Medline: [15741983](#)]
16. Wang D, Kogashiwa M, Kira S. Development of a new instrument for evaluating individuals' dietary intakes. *J Am Diet Assoc* 2006 Oct;106(10):1588-1593. [doi: [10.1016/j.jada.2006.07.004](#)] [Medline: [17000191](#)]
17. Rangan AM, O'Connor S, Giannelli V, Yap ML, Tang LM, Roy R, et al. Electronic Dietary Intake Assessment (e-DIA): Comparison of a Mobile Phone Digital Entry App for Dietary Data Collection With 24-Hour Dietary Recalls. *JMIR Mhealth Uhealth* 2015 Oct 27;3(4):e98 [FREE Full text] [doi: [10.2196/mhealth.4613](#)] [Medline: [26508282](#)]
18. Ambrosini GL, Hurworth M, Giglia R, Trapp G, Strauss P. Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. *Nutr J* 2018 Jan 09;17(1):5 [FREE Full text] [doi: [10.1186/s12937-018-0315-4](#)] [Medline: [29316930](#)]
19. Teixeira V, Voci SM, Mendes-Netto RS, da Silva DG. The relative validity of a food record using the smartphone application MyFitnessPal. *Nutr Diet* 2018 Apr;75(2):219-225. [doi: [10.1111/1747-0080.12401](#)] [Medline: [29280547](#)]
20. Griffiths C, Harnack L, Pereira MA. Assessment of the accuracy of nutrient calculations of five popular nutrition tracking applications. *Public Health Nutr* 2018 Jun;21(8):1495-1502. [doi: [10.1017/S1368980018000393](#)] [Medline: [29534771](#)]
21. Illner A, Freisling H, Boeing H, Huybrechts I, Crispim S, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. *Int J Epidemiol* 2012 Aug;41(4):1187-1203. [doi: [10.1093/ije/dys105](#)] [Medline: [22933652](#)]
22. Willett W. Nutritional epidemiology. New York: Oxford University Press; 1998.

23. Gibson R. Principles of Nutritional Assessment. New York: Oxford University Press; 2005.
24. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a Smartphone Application for Weight Loss Compared to Website and Paper Diary: Pilot Randomized Controlled Trial. *J Med Internet Res* 2013 Apr 15;15(4):e32. [doi: [10.2196/jmir.2283](https://doi.org/10.2196/jmir.2283)] [Medline: [23587561](https://pubmed.ncbi.nlm.nih.gov/23587561/)]
25. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. *Ann Intern Med* 2014 Nov 18;161(10 Suppl):S5-12. [doi: [10.7326/M13-3005](https://doi.org/10.7326/M13-3005)] [Medline: [25402403](https://pubmed.ncbi.nlm.nih.gov/25402403/)]
26. Touvier M, Kesse-Guyot E, Méjean C, Pollet C, Malon A, Castetbon K, et al. Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. *Br J Nutr* 2011 Apr;105(7):1055-1064. [doi: [10.1017/S0007114510004617](https://doi.org/10.1017/S0007114510004617)] [Medline: [21080983](https://pubmed.ncbi.nlm.nih.gov/21080983/)]
27. Rebore SM, Patterson RE, Kristal AR, Cheney CL. The effect of keeping food records on eating patterns. *J Am Diet Assoc* 1998 Oct;98(10):1163-1165. [doi: [10.1016/S0002-8223\(98\)00269-7](https://doi.org/10.1016/S0002-8223(98)00269-7)] [Medline: [9787724](https://pubmed.ncbi.nlm.nih.gov/9787724/)]
28. Beasley J, Riley WT, Jean-Mary J. Accuracy of a PDA-based dietary assessment program. *Nutrition* 2005 Jun;21(6):672-677. [doi: [10.1016/j.nut.2004.11.006](https://doi.org/10.1016/j.nut.2004.11.006)] [Medline: [15925290](https://pubmed.ncbi.nlm.nih.gov/15925290/)]
29. Six BL, Schap TE, Zhu FM, Mariappan A, Bosch M, Delp EJ, et al. Evidence-based development of a mobile telephone food record. *J Am Diet Assoc* 2010 Jan;110(1):74-79 [FREE Full text] [doi: [10.1016/j.jada.2009.10.010](https://doi.org/10.1016/j.jada.2009.10.010)] [Medline: [20102830](https://pubmed.ncbi.nlm.nih.gov/20102830/)]
30. Chen J, Berkman W, Bardouh M, Ng CYK, Allman-Farinelli M. The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. *Nutrition* 2019 Jan;57:208-216. [doi: [10.1016/j.nut.2018.05.003](https://doi.org/10.1016/j.nut.2018.05.003)] [Medline: [30184514](https://pubmed.ncbi.nlm.nih.gov/30184514/)]
31. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing Self-Monitoring Strategies for Weight Loss in a Smartphone App: Randomized Controlled Trial. *JMIR Mhealth Uhealth* 2019 Feb 28;7(2):e12209 [FREE Full text] [doi: [10.2196/12209](https://doi.org/10.2196/12209)] [Medline: [30816851](https://pubmed.ncbi.nlm.nih.gov/30816851/)]

## Abbreviations

**AUSNUT:** Australian Food, Supplement, and Nutrient Database

**g:** grams

**kcal:** kilocalories

**NDSR:** Nutrition Data System for Research

**PDA:** personal digital assistant

*Edited by G Eysenbach; submitted 13.02.20; peer-reviewed by D Segovia Lizano, PH Lin, A Wong; comments to author 26.06.20; revised version received 12.08.20; accepted 02.09.20; published 21.10.20.*

*Please cite as:*

Evenepoel C, Clevers E, Deroover L, Van Loo W, Matthys C, Verbeke K

Accuracy of Nutrient Calculations Using the Consumer-Focused Online App MyFitnessPal: Validation Study

*J Med Internet Res* 2020;22(10):e18237

URL: <http://www.jmir.org/2020/10/e18237/>

doi: [10.2196/18237](https://doi.org/10.2196/18237)

PMID: [33084583](https://pubmed.ncbi.nlm.nih.gov/33084583/)

©Charlotte Evenepoel, Egbert Clevers, Lise Deroover, Wendy Van Loo, Christophe Matthys, Kristin Verbeke. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 21.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Reducing Alert Fatigue by Sharing Low-Level Alerts With Patients and Enhancing Collaborative Decision Making Using Blockchain Technology: Scoping Review and Proposed Framework (MedAlert)

Paul Kengfai Wan<sup>1</sup>, MSc; Abylay Satybaldy<sup>2</sup>, MSc; Lizhen Huang<sup>1</sup>, PhD; Halvor Holtskog<sup>3</sup>, PhD; Mariusz Nowostawski<sup>2</sup>, PhD

<sup>1</sup>Department of Manufacturing and Civil Engineering, Norwegian University of Science and Technology, Gjøvik, Norway

<sup>2</sup>Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway

<sup>3</sup>Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Gjøvik, Norway

**Corresponding Author:**

Paul Kengfai Wan, MSc

Department of Manufacturing and Civil Engineering

Norwegian University of Science and Technology

Teknologiveien 22

Gjøvik, 2815

Norway

Phone: 47 93984604

Email: [paul.k.wan@ntnu.no](mailto:paul.k.wan@ntnu.no)

## Abstract

**Background:** Clinical decision support (CDS) is a tool that helps clinicians in decision making by generating clinical alerts to supplement their previous knowledge and experience. However, CDS generates a high volume of irrelevant alerts, resulting in alert fatigue among clinicians. Alert fatigue is the mental state of alerts consuming too much time and mental energy, which often results in relevant alerts being overridden unjustifiably, along with clinically irrelevant ones. Consequently, clinicians become less responsive to important alerts, which opens the door to medication errors.

**Objective:** This study aims to explore how a blockchain-based solution can reduce alert fatigue through collaborative alert sharing in the health sector, thus improving overall health care quality for both patients and clinicians.

**Methods:** We have designed a 4-step approach to answer this research question. First, we identified five potential challenges based on the published literature through a scoping review. Second, a framework is designed to reduce alert fatigue by addressing the identified challenges with different digital components. Third, an evaluation is made by comparing MedAlert with other proposed solutions. Finally, the limitations and future work are also discussed.

**Results:** Of the 341 academic papers collected, 8 were selected and analyzed. MedAlert securely distributes low-level (nonlife-threatening) clinical alerts to patients, enabling a collaborative clinical decision. Among the solutions in our framework, Hyperledger (private permissioned blockchain) and BankID (federated digital identity management) have been selected to overcome challenges such as data integrity, user identity, and privacy issues.

**Conclusions:** MedAlert can reduce alert fatigue by attracting the attention of patients and clinicians, instead of solely reducing the total number of alerts. MedAlert offers other advantages, such as ensuring a higher degree of patient privacy and faster transaction times compared with other frameworks. This framework may not be suitable for elderly patients who are not technology savvy or in-patients. Future work in validating this framework based on real health care scenarios is needed to provide the performance evaluations of MedAlert and thus gain support for the better development of this idea.

(*J Med Internet Res* 2020;22(10):e22013) doi:[10.2196/22013](https://doi.org/10.2196/22013)

**KEYWORDS**

blockchain; health care; alert fatigue; clinical decision support; smart contracts; information sharing



## Introduction

### Background

Clinical decision support (CDS) is a tool to facilitate medical decision making by generating clinical alerts [1], ranging from simple medication-specific alerts based on stored clinical rules and information to more complex patient-specific alerts by integrating CDS with electronic health records (EHRs) [2]. For example, CDS warns clinicians by generating an alert if a new prescription poses a threat to patients [3]. This real-time alert disrupts the workflow and draws clinicians' attention so they can evaluate and make appropriate decisions in a quick and efficient manner [4]. CDS has replaced previous situations in which clinicians make decisions solely on the basis of their knowledge and past experience [5]. CDS is now considered an essential health information technology that improves the overall quality of health care [6]. However, current CDS tools generate a high volume of irrelevant alerts, resulting in alert fatigue [7].

Alert fatigue or alert burden is defined as the mental state that results when alerts or reminders consume too much time and mental energy, which can cause clinicians to override or ignore both clinically irrelevant and relevant alerts unjustifiably [8]. Clinicians are now drowning with alerts and gradually becoming less responsive to and less respectful of them [9]. This is mainly because generated alerts are mostly irrelevant or low priority, and fortunately, they are not life threatening. In the long term, these *cry-wolf* alerts have desensitized clinicians, resulting in high overriding rates ranging between 77% and 90% [10-12], which opens the door to preventable medication errors.

Alert fatigue started becoming increasingly common in the health care sector decades ago and is now widely recognized as a national concern, often due to the lack of a corresponding action plan [13]. CDS failures and errors caused by individuals have resulted in direct costs of more than US \$20 billion in the United States [14,15]. Alert fatigue is perceived as a major problem because it extends beyond the health care industry. Other sectors, such as off-shore oil drilling [16] and heating, ventilation, and air-conditioning systems in buildings [17], are also experiencing alert fatigue. For example, fault detection systems generate high volumes of alerts, leading to operator alert fatigue and resulting in energy wastage in buildings. Currently, there is a persistent upward trend and increasing requests for new alerts [13], which does not help alert fatigue. This only exacerbates the alert fatigue and makes it more widespread.

Overriding alerts is clinically appropriate if the alert generated is incorrect [7]. However, due to the low specificity and high volume of alerts generated by CDS, relevant alerts may also be dismissed, resulting in preventable prescription errors and adverse drug events. Deactivation [18] or running low-priority alerts in silence [19] are among the suggestions for reducing alert fatigue. However, these approaches in managing alerts

effectively are difficult because of strict regulatory bodies and other external pressures. Many are in fact pushing for more rather than fewer alerts to reduce or avoid preventable medication errors [13].

In Norway, approximately 12% of patient harm is caused by the incorrect use of drugs [20]. One in three elderly people have been given the wrong medication, and an estimated one thousand deaths per year are thought to be due to medication errors, despite the use of e-prescriptions [21,22]. During a meeting at the Norwegian University of Science and Technology (NTNU), a health care representative from Innlandet Hospital presented in his presentation that approximately 8% of total health care spending went on correcting medication errors within the Innlandet region. [23]. We, therefore, agree with Wright et al [24] that the health care sector can only benefit from the potential value of CDS-generated alerts when they are well designed and properly implemented. Thus, there is a need to seek an alternative, innovative approach to improve the management of clinical alerts and reduce alert fatigue among clinicians.

### Objectives

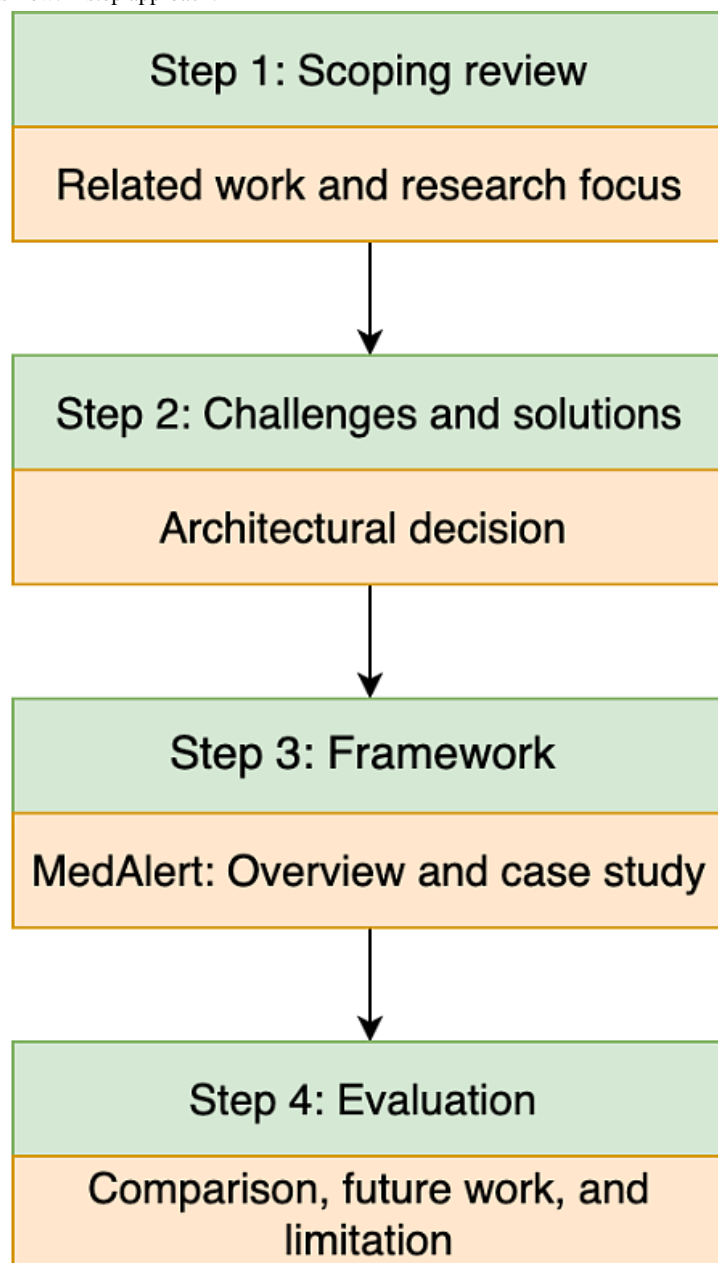
Blockchain technology has gained attention as a potential solution in the health care sector, mainly due to its potential in moving toward collaborative treatments and decision making [25-27]. A large range of literature has been published anticipating this technology with a view to improving the health sector with respect to the overall well-being of clinicians and the quality of patients' health care by sharing medical records and history [26]. However, studies focusing on clinical alerts using blockchain remain limited. This has led us to our main research question in this paper, which is to explore and understand how a blockchain-based solution can help to reduce alert fatigue in the health sector by sharing alerts and thus enhancing collaborative decision making. To answer this question, we designed a 4-step approach, which is explained in the *Methods* section.

## Methods

### Design Approach

The 4-step approach, shown in Figure 1, is designed to answer our research question and explain how the paper is organized. The first step is to conduct a scoping review to explore the current state of the art in this area. The literature we finally selected and the existing solutions we have chosen are then analyzed in step 1. Step 2 is designed to identify potential challenges and technical solutions for reducing alert fatigue. Architectural decisions are explained in this step. The framework is designed in step 3. An overview of MedAlert, together with a case study, is elaborated in this step. Finally, the framework is evaluated by comparing it with other proposed solutions. The comparison, future work, limitations, and benefits are also discussed in step 4.

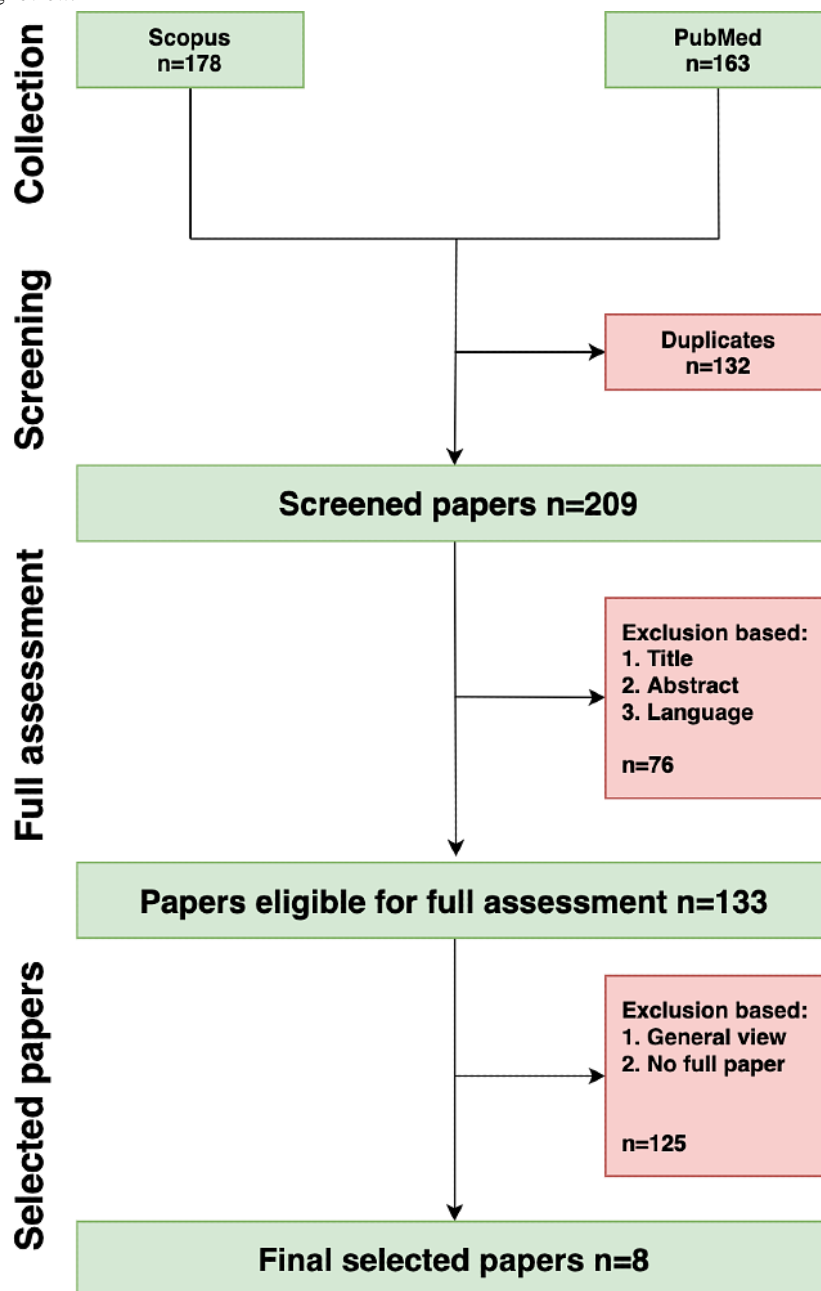


**Figure 1.** Research design process flow: 4-step approach.

### Scoping Review: Search Strategy

A scoping review was conducted with the aim of exploring the current state of the art in academic research with the widest possible coverage of all the published literature. The reporting of this scoping review was guided by PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-analysis extension for Scoping Reviews) [28]. We performed searches on 2 bibliographic databases, Scopus and PubMed. To be as comprehensive as possible, generic keyword strings such as *blockchain*, *clinical decision support*, *alert burden*, and *alert fatigue* were used as search criteria. [Multimedia Appendix 1](#) details the structures of the keyword strings.

We acknowledge that industries are also working on blockchain-based solutions within the health care sector, but often, the details of the frameworks are not disclosed. Therefore, in our research, we focus primarily on the academic sphere because the architecture frameworks and solutions are described in published work. Peer-reviewed articles, conferences, reviews and proceedings, and dissertations are included to provide a broad overview of different aspects of alert fatigue resulting from CDS. Only English papers were included, with no restrictions on the year or country of publication. We excluded general views, no full paper, and conference abstracts. The selection process for the scoping review is summarized in [Figure 2](#).

**Figure 2.** Process of scoping review.

## Results

### Related Work and Research Focus

Alert fatigue is a major problem faced by clinicians and is now a rising concern in the health care sector. The published literature on alert fatigue in the academic sphere started as early as 2007. We collected a total of 341 published items and finally selected a total of 8 [13,15,29-34] that fit our research criteria based on the scoping review in Figure 2. We then entered these items in Microsoft Word and Excel for deeper analysis. We summarize and sort these literatures according to their different key foci, methods, and benefits in Multimedia Appendix 2 [13,15,29-34].

Carli et al [35], Powers et al [36], and Hussain et al [37] pointed out that the high degree of alerts with low clinical relevance is one of the root causes of alert fatigue in their systematic

literature reviews. This is because hospitals and other private health care institutions use or purchase commercial CDS tools to improve the overall quality of their health care systems. It is common for vendors and designers of commercial CDS tools to sharply restrict the ability to modify the setup for alert systems, resulting in a high volume of low-relevance alerts [2]. The strict, low-specificity settings imposed by vendors are due to their fear of being exposed to potential litigation if the removal of alerts fails to prevent a potential medication error.

One common attempt to address alert fatigue is to reduce the number of alerts of low clinical relevance by clustering alerts with similar clinical management options [32] or better specifications to generate useful alerts [31]. The machine learning algorithm-based CDS is another suggested method to generate more context-driven alerts [15] and patient-centric alerts [34]. Soundararajan et al [30] designed a blockchain architecture framework to leverage blockchain and smart

contracts in support of clinical support tools that generate more patient context-appropriate alerts and thus generate fewer inappropriate alerts, which could reduce physician burnout. However, the actual benefits to patients and the extent of the positive impact on alert fatigue remain unclear.

All these efforts have managed to reduce the total number of alerts generated, but the fundamental issue of alert fatigue has still not been tackled. Bryant et al [38] pointed out that despite intensive efforts to reduce irrelevant alerts of commercial systems, overriding rates remain as high as reported over a decade ago. Medical experts suggested that improving alert fatigue should go beyond just reducing the total number of alerts [39].

Getting someone to attend the alerts is one way to reduce alert fatigue. Smithburger et al [5] suggested a potential strategy for directing alerts to medical professionals other than clinicians, for example, nurses. A study conducted in three academic medical centers in the Netherlands evaluated shifting time-dependent drug interaction alerts to medical staff such as nurses or pharmacists [40]. These results demonstrated the ability to improve the efficiency and effectiveness of such alerts and showed that incorrect administration times were reduced by 29% when they were directed at nurses. This can enable more collaborative treatment and decision care, whereas blockchain technology can be leveraged to enable alert sharing [25].

In our work, we have explored how blockchain can be leveraged to reduce alert fatigue by directing low-level alerts to patients in achieving high-quality collaborative clinical decisions. There has been a recent shift toward a more patient-centric data sharing for better collaborative decision making within the health care sector [41]. However, the relevant work remains limited. Thus, we contribute by designing an exploratory blockchain-based framework that enables low-level alert sharing with patients to enable more collaborative decision making while maintaining a high level of privacy and security. To design a sound framework, we need to understand and consider the challenges involved in facilitating the sharing of clinical alerts.

Data integrity and user privacy are two of the main concerns of the health care industry worldwide [42]. One of the reasons for this is that most of the current health care systems have weak and vulnerable centralized data storage procedures for preserving and managing sensitive medical data [43]. In 2019, the database of the Health Sciences Authority in Singapore was hacked for the third time in less than a year because of security loopholes, and more than 800,000 personal details were exposed [44]. Identity theft is another issue of concern in the health sector. According to Pandey et al [43], 10% of data breaches in the health industry in the last 10 years were categorized as identity theft.

There is a range of literature on blockchain-based frameworks that serves as an alternative to current vulnerable centralized database systems. EMRshare [45], Medchain [46], FHIRchain [25], and MedBlock [47] are examples of blockchain-based solutions that ensure high levels of data integrity and privacy for sharing medical records. In addition, smart contracts can enable a new service for health care to facilitate information sharing without a third party. For example, Medchain enables medical record access between multiple roles, such as patients, requesters, and health care providers, and helps them to achieve higher levels of efficiency and to satisfy security requirements [46]. This can improve collaborative decision making between different stakeholders, for example, clinicians and patients, in the health care sector.

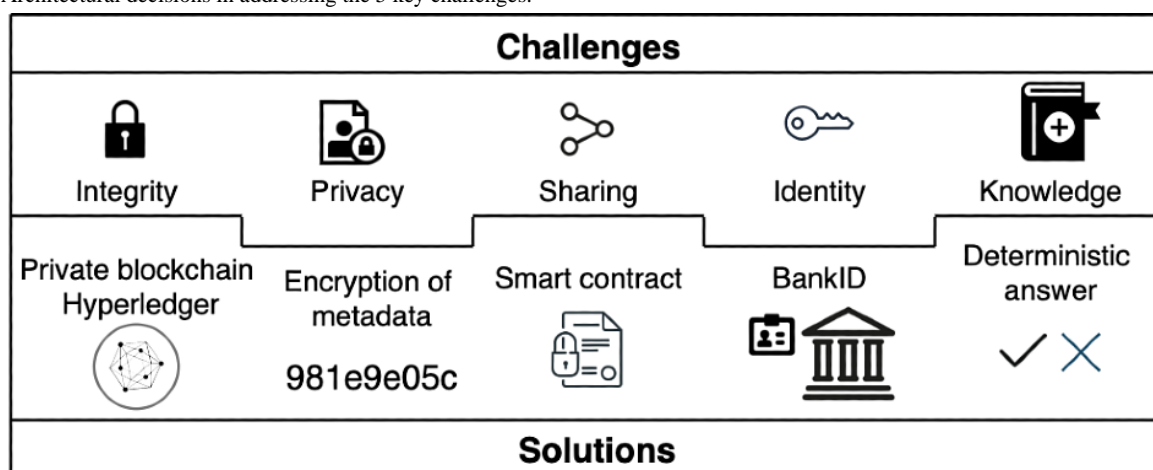
Five key challenges must be addressed to develop a secure and effective blockchain framework and thus reduce the alert burden within health care. These 5 key challenges are as follows:

1. Data integrity
2. Privacy issues
3. Verifying and authenticating participants' identities
4. Lack of secure information sharing
5. The extent of patients' knowledge in the medical field

### Architectural Decisions to Address These Five Key Challenges

In this section, we address the challenges suggested in the previous section. The architectural decisions are summarized in Figure 3.

**Figure 3.** Architectural decisions in addressing the 5 key challenges.



## Challenge 1: Data Integrity

### Context

Health information is sensitive and must be highly secured, without the possibility of any data manipulation. Any alteration in a patient's medical history could result in severe medication errors and even death. Medical data are often stored in and managed by centralized trusted third-party databases. However, such a centralized database can be vulnerable to single-point failures, resulting in loss or corrupted medical data and blocking of access deliberately during disputes by service providers [48]. Some modern EHR systems can be configured to have a backup and data redundancy mechanism that improves data storage resilience, but it requires additional configuration and maintenance that might be error prone due to human factors. According to some security experts, current systems in protecting our health data do not achieve the desired modern security standards [49].

### Solution

A solution that offers resiliency out of the box and a tamper-sensitive storing environment to prevent any *silent* manipulation by making alterations obvious to members in the network.

### Technical Requirement

Blockchain technology is a distributed ledger that contains replicated and synchronized digital data. It provides a platform for real-time data sharing between a large number of members in a network with a higher level of data trust [50,51]. *Data trust* denotes the reliability of the information and data provided [52]. A high level of data trust is important for decision making.

The data storage structure is a salient feature of blockchain, which ensures that information and data are stored in a tamper-evident environment [53]. All valid transactions are recorded in a block format, and each block is linked with a time stamp and hash references forming a chain of blocks [54]. Any attempt to alter information, for example, in the off-chain database, regardless of the intention, breaks the hash reference and thus makes it obvious to the other members of the network. This way, a hash reference creates a tamper-evident environment that maintains and ensures data integrity. The transactions recorded on the blockchain remain immutable and tamper-proofed owing to the structure and writing rights of the blockchain itself, which can guarantee a high level of data integrity.

## Challenge 2: Privacy Issues

### Context

Medical information, including medical records, prescription histories, patients' personal information, and surgical records that are stored in digital formats, are classified as digital assets. This information requires high levels of privacy protection because it relates to the patient's current physical or mental health and can reveal information about his or her health status [55]. Ensuring that current or new health services are in compliance with standards, such as General Data Protection Regulation (GDPR) is crucial to avoid unlawful behavior. For example, encryption, pseudonymization, or anonymization of

personal data, whenever possible, to prevent unlawful data processing [55].

### Solution

A private permissioned blockchain is a better option when it comes to ensuring on-chain data privacy and compliance with privacy regulations because transactions are visible only to members. Certain members of the network are granted permissions to read and write on the blockchain. By storing only metadata instead of actual health data, we can avoid exposing actual sensitive personal data, such as full name, diagnoses, and prescribed drugs, which could violate a patient's privacy.

### Technical Requirement

To increase the level of privacy protection, private blockchains such as Hyperledger are preferred over public blockchains, primarily because of the lower degree of visibility and level of *openness*. Information on private blockchains is only accessible to authorized members of the network and not just anyone with internet access. Only an authorized member, in our case clinician, has permission to write and store on the blockchain. This allows the framework to be more compliant with data protection regulations such as GDPR or HIPAA (Health Insurance Portability and Accountability Act) without compromising the privacy of patients [49].

Encrypting metadata in blockchain provides a higher level of security and protection for patients [25] because metadata are treated as sensitive data in health care. This prevents any unauthorized hacker from obtaining actual health information improperly. Encrypted metadata can act as a reference pointer to the patient's prescription profile in the health system. The reference pointer links transactional data from the blockchain to the actual data stored on an off-chain database. This acts as a form of protection because it isolates the patients' actual medical information from the reference itself. The pointer breaks and becomes invalidated when any alteration to the patient's data occurs in the off-chain database. Another benefit in storing encrypted metadata is the lightweight reference pointer, which is more suitable and efficient to store on blockchain, which currently has limited storage capability. This can be a scalable alternative [25].

## Challenge 3: Verifying and Authenticating Participants' Identities

### Context

It is important to ensure that the right patient receives the designated clinical alert from the clinicians. Clinicians working in hospitals can verify and authenticate themselves with the credentials offered by health care institutions through logging into the health care system. However, health care systems today lack a standard platform [56], particularly for patients, to verify and authenticate their digital identities.

### Solution

Use a trusted digital identity management system to verify patients' digital identities. Digital identity denotes the digital representation of entity attributes such as birth or other registered name, national ID number, and registered mobile number to

access systems and applications using an identity mediation process [57-59]. This allows patients to authenticate their identities accurately and thus either authorize or revoke access to certain requestors. This is a way of protecting patients' sensitive data, including managing their medical records, and it guarantees that security and privacy are compliant with local legislation and laws [60,61].

### Technical Requirement

Federated digital identity management, registered once and trusted by many concepts, is widely used in consumer spaces such as Facebook and Google and is trusted by many applications [62]. Unlike traditional centralized identity management, users do not need to set up and register their digital identities with every service provider. In this system, mutual trust is established by receiving components of proof distributed by two or more centralized owners or by mutually recognizing each other's trust and proofing standards [62]. Consortiums of leading banks and mobile operators have created private federated identification procedures, such as BankID in Norway [63] and Smart-ID in Estonia [64], to facilitate the distribution of verified and authenticated identities, thus enabling their citizens or users to access various portals, services, and platforms directly.

### Challenge 4: Lack of Secure Information Sharing

#### Context

Each medical institution has its own way of governing medical records and data. Often, moreover, they are not interaccessible, thus making information sharing difficult. Along with strict legal regulations and the lack of trust in medical institutions outside the organization, information exchange becomes more challenging [45].

#### Solution

Use a common layer to enable information sharing securely without altering the current health care IT infrastructure and to enhance collaborative decision making.

#### Technical Requirement

Smart contracts can govern and facilitate information exchange between two different actors accurately and verifiably without the intervention of an intermediate third party. It also enables autonomous self-execution, once a set of predefined rules is met [65]. For example, when an alert is generated from CDS, it triggers a smart contract to direct the alert to the identified patient. The integration of smart contracts can increase the efficiency of members' real-time decision making and overall information exchange. All events are recorded in the blockchain with a time stamp, and the blockchain structure can act as a common layer of information storage without changing the existing IT infrastructure. Smart contracts can track real-time

performance and also query past events for the purposes of analysis.

### Challenge 5: The Extent of Patients' Knowledge in the Medical Field

#### Context

When directing alerts to patients for a collaborative decision, the main problem is that they may not have sufficient knowledge to make the correct decision. Making a wrong decision can be fatal to patients.

#### Solution

Only low-level and nonlife-threatening alerts are directed to patients governed by smart contracts. Patients will receive clinical alerts and then provide information back to the clinician. The aim of directing alerts to patients is bring the alert to their attention, instead risking its rapid dismissal by clinicians due to the high volumes of alerts. This could reduce alert fatigue and the total number of alerts because clinicians can place the emphasis on higher-level alerts.

#### Technical Requirement

Smart contracts execute actions by sending notifications to patients when the CDS generates an alert. The alert is then directed to the patient in the form of a question with a deterministic answer, either *Yes* or *No*. Given a real-time response, the clinician is able to modify the prescription accordingly and eliminate medication prescription errors based on the responses provided by patients.

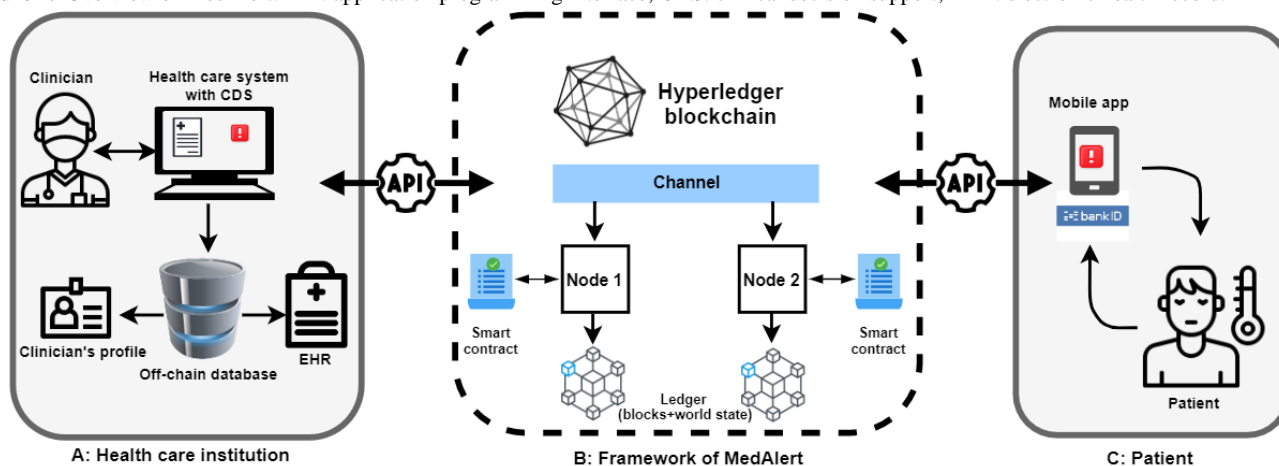
### Principal Finding: MedAlert

#### Overview

This section provides an overview of MedAlert as a potential solution for reducing alert fatigue and enabling a more collaborative process of clinical decision making. This case study is developed as a two-step scenario: (1) *how a patient logs in with BankID to verify and authenticate his or her identity before revealing the alert* and (2) *how a patient is involved in the decision-making process*.

Figure 4 shows how MedAlert (B) enables the interaction between a clinician in a health care institution (A) and a patient (C). The MedAlert is hosted in a private blockchain framework such as Hyperledger. The clinician authorizes through logging into his or her profile with their credentials issued by the health care institution, whereas the patient can log in with BankID to verify and authenticate himself or herself. The blockchain nodes can be administered by a collection of health care organizations such as hospitals but not on a patient's mobile device due the high requirement of computational resources and a consistent network connectivity. These nodes host ledgers and smart contracts that can be queried and updated by peer-connected applications.



**Figure 4.** Overview of MedAlert. API: application programming interface; CDS: clinical decision support; EHR: electronic health record.

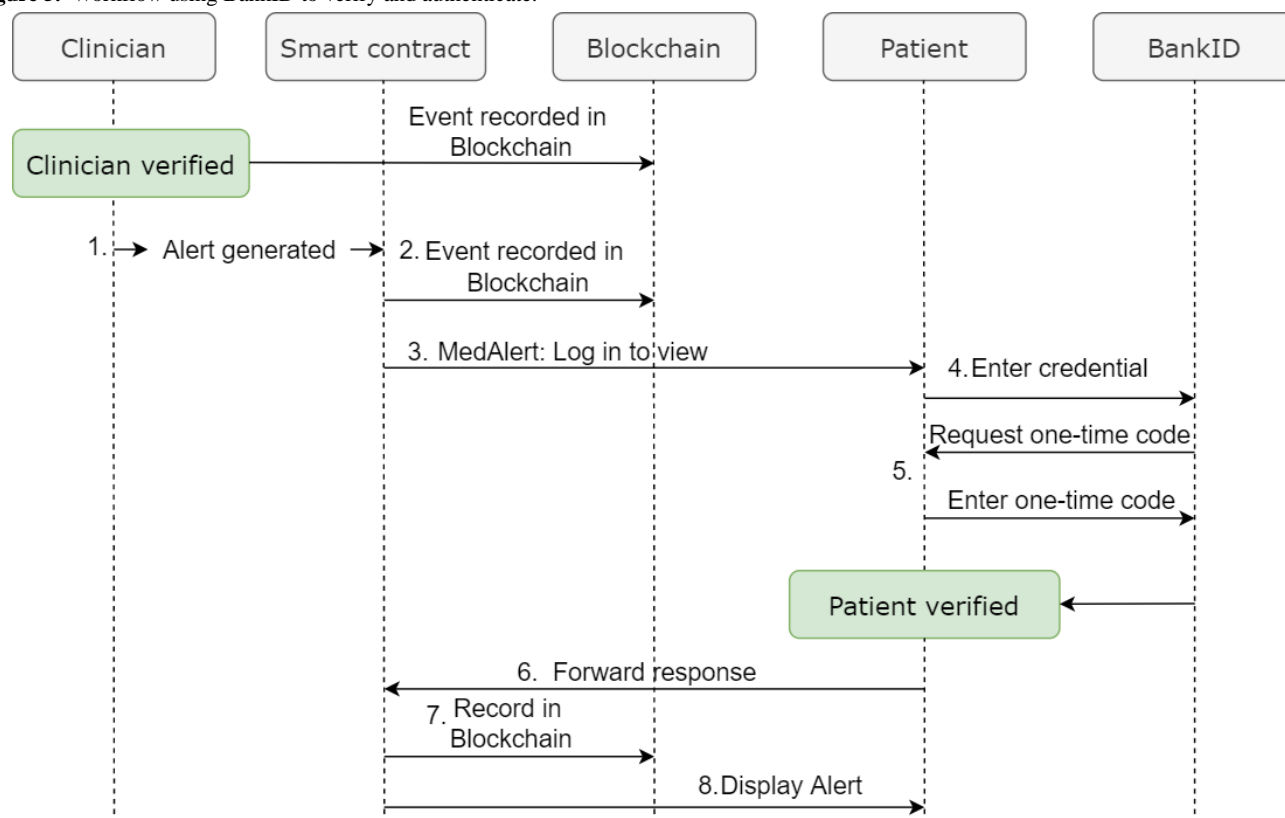
Application programming interfaces (APIs) can enable alert sharing with multiple health care systems. Representational state transfer (REST) APIs can establish communication between mobile client apps and the blockchain network. A client app sends a transaction proposal using organization-specific REST APIs that enable apps to connect to nodes; invoke smart contracts that generate transactions; submit transactions to the network that will be ordered, validated, and committed to the distributed ledger; and receive events when this process is complete.

The consensus protocol in the private blockchain enables transaction data integrity. For every transaction, each node will verify that the transaction has been endorsed by the required organizations according to the endorsement policy of the smart contract that generated the transaction. For example, some

transactions may only need to be endorsed by a single organization, whereas others may require multiple endorsements before they are considered valid. This process of validation verifies that all relevant organizations have generated the same outcome or result.

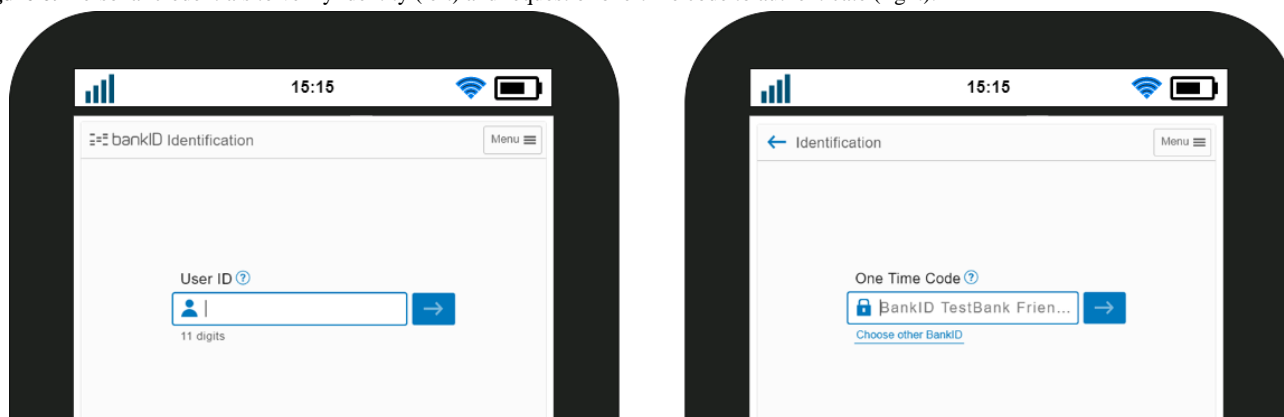
#### *How a Patient Logs In With BankID to Verify and Authenticate His or Her Identity Before Revealing the Alert*

This section describes the step-by-step workflow, as shown in Figure 5. Before clinicians can access patients' EHRs or prescribe new drugs, they need to authenticate their identities by logging in their credentials into the health care system. This event is recorded in the blockchain. When the clinician prescribes a drug to a patient and assumes that it could pose a threat to the patient:

**Figure 5.** Workflow using BankID to verify and authenticate.

1. A clinical alert (red exclamation mark) is generated from the CDS system, as shown in Figure 4. This triggers the smart contract.
2. This event is then recorded in the blockchain.
3. The smart contract also sends the alert to the patient's registered mobile number.
4. The patient receives a message with a link to verify and authenticate his or her identity. Then he or she must log in to verify and authentic himself or herself by providing his or her registered user ID (eg, the 11 digits of a social security number) as sketched in Figure 6 (left).
5. The patient is then required to enter his or her one-time code for final authentication, as shown in Figure 6 (right).
6. When the authentication and verification is successful, the response is forwarded to the smart contract.
7. This event is also recorded in the blockchain.
8. The patient is then able to view the alert.

**Figure 6.** Personal credentials to verify identity (left) and request of one-time code to authenticate (right).

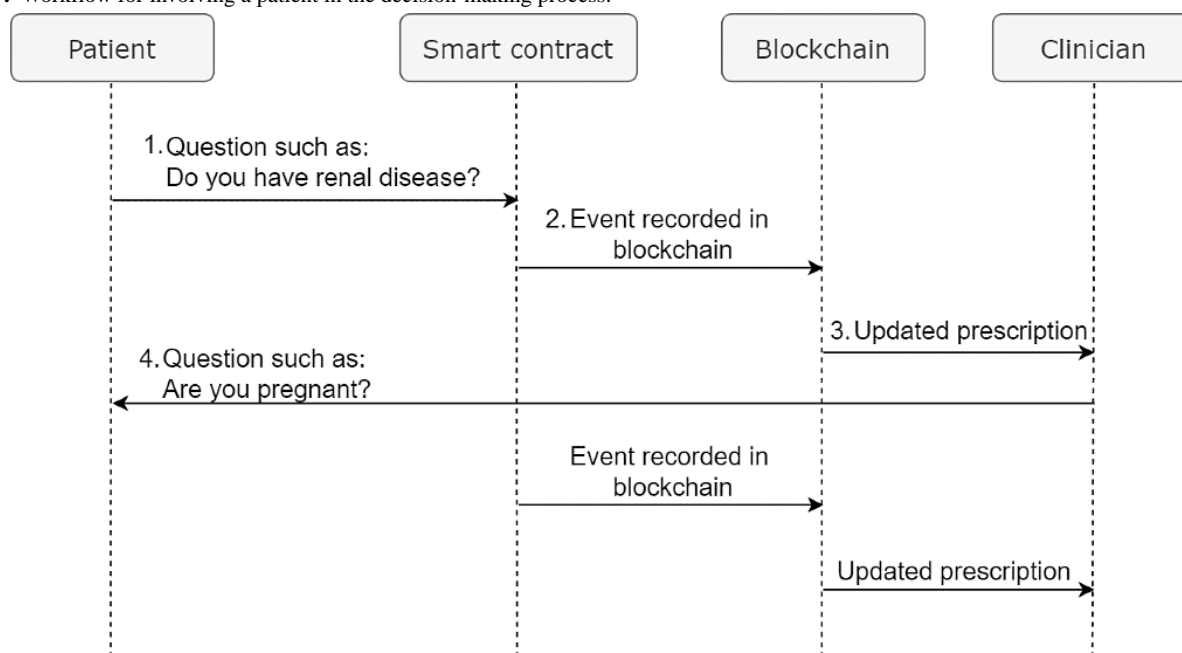


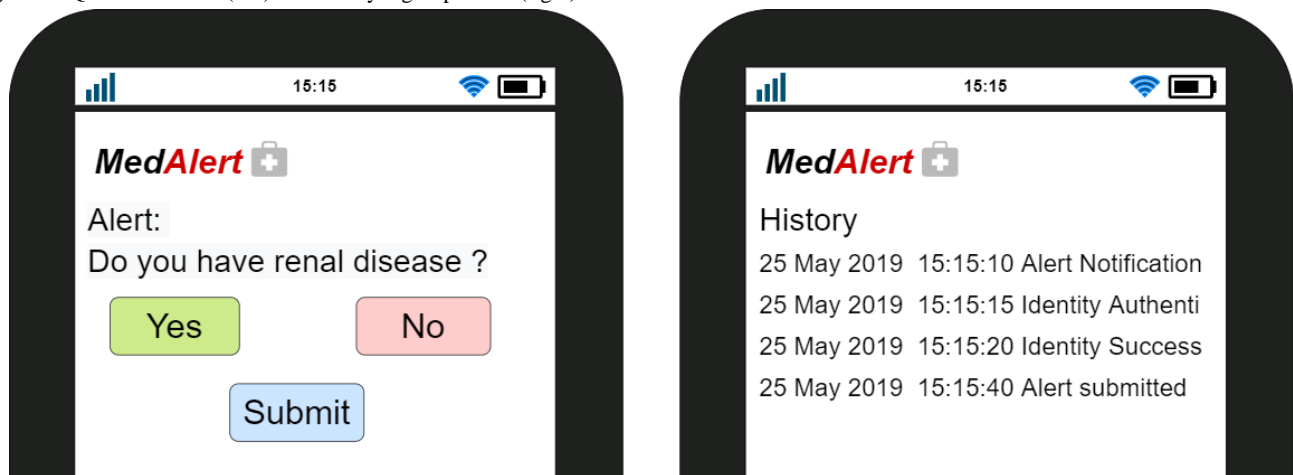
### How a Patient Is Involved in the Decision-Making Process

After the patient has verified and authenticated his or her identity, the patient can access and read the information in the alert. The workflow is shown in Figure 7.

1. The first alert asks: "Do you have renal disease?" The answer to the question is either *Yes* or *No*, as shown in Figure 8 (left).
2. When the patient responds, the smart contract is then triggered, and the patient sends the response back to the clinician. The transaction is recorded in the blockchain.
3. The clinician updates the prescription according to the answer provided.
4. If another low-level alert pops up, the patient has to respond in real time before the prescription is finalized. The patient can view his or her history, as shown in Figure 8 (right).

**Figure 7.** Workflow for involving a patient in the decision-making process.



**Figure 8.** Question in alert (left) and history log of patients (right).

## Discussion

### Comparison With Prior Work

Three different frameworks (MedAlert, MedRec, and MedAware) are compared in Table 1. These solutions can reduce the total number of alerts generated but with fundamentally different technologies. Both MedRec and MedAware focus on reducing alert fatigue by filtering irrelevant alerts. MedRec

utilizes a smart contract embedded in a blockchain platform, from which CDS retrieves medical records via MedRec to retrieve relevant patient information and generate alerts that are more context based. MedAware uses a machine learning algorithm to flag more relevant and accurate alerts in real time after analyzing patients' historical medical records. However, improving alert fatigue should go beyond just reducing the total number of alerts [39]. These two solutions only capture the clinician's attention.

**Table 1.** Comparison of different framework solutions.

Solution	MedAlert	MedRec	MedAware
Alert reduction	Yes	Yes	Yes
<b>Alert capturing</b>			
Clinician	Yes	Yes	Yes
Patient	Yes	No	No
<b>Privacy</b>			
Ownership	Clinician and patient	Clinician and patient	N/A <sup>a</sup>
Encryption	Yes	No	N/A
<b>Blockchain</b>			
Type	Private: Hyperledger	Public: Ethereum	No, machine learning
Smart contract	Yes	Yes	N/A
Miners	No	Yes, medical background	N/A

<sup>a</sup>N/A: not applicable.

Unlike MedRec and MedAware, MedAlert reduces alert fatigue by capturing the attention of both clinicians and patients. We believe that the way to reduce alert fatigue is to get the clinician's attention, but there is no *perfect* solution in which clinicians are able to pay attention to all alerts [15], not even after the removal of irrelevant alerts. Therefore, MedAlert directs low-level alerts to patients and induces them to pay attention to provide real-time responses. This is a novel initiative moving toward a clinician-patient collaborative decision-making process to avoid potential medication errors resulting from action being overridden. This can improve the quality of the health care domain with respect to better patient outcomes and reducing physician burnout.

MedAlert runs on the private Hyperledger blockchain, which ensures a higher privacy compared with MedRec, which runs on the public Ethereum blockchain. This is because private blockchain is better suited to a highly regulated industry such as health care due to the stricter requirements regarding patient privacy and data protection. To avoid information leakage, both MedAlert and MedRec record only metadata or reference pointers rather than patient's medical data on blockchain. To enhance patients' data privacy, all metadata is encrypted and stored on MedAlert blockchain, where only authenticated patients can view the transactions and authorized clinicians can read and write transactions. This makes MedAlert better

compliant to standards such as GDPR (Art. 32. Security of processing) compared with MedRec.

Apart from ensuring a higher-level privacy environment, MedAlert, deployed in private Hyperledger, has a better performance than MedRec, deployed in Ethereum. The assessments from Pongnumkul et al [66] show that Hyperledger outperforms Ethereum in 3 evaluation metrics: execution time, latency, and throughput. For example, the average latency of Ethereum is about 2 times at a low number of transactions and can increase up to 14 times that of Hyperledger at a high number of transactions. This is important when fast information sharing is needed between a clinician and a patient during collaborative decision making.

MedAlert can improve the flow of communication between clinicians and patients. Clinicians may need to ask for and validate information with patients because without this step, there is a significant risk of error in ordering or prescribing medication [67]. This risk can increase when alerts generated by CDS are simply overridden. MedAlert can reduce this and prevent it from happening by sharing clinical alerts with patients. Patients can receive the alert and be asked to provide information. If they are uncertain, they can enter into direct communication with the clinician and deal with the alert that way.

### Future Work

Validation work such as threat analysis is needed in future work to elucidate the effectiveness and the potential vulnerabilities of using MedAlert before deploying it in the eHealth sector [68]. This would provide a documented performance evaluation of MedAlert to persuade health care leaders of the benefits of this new digital tool and gain sufficient support from them for its deployment. Despite numerous published literature on how blockchain can record immutable transactions and enhance interoperability and thus improve health care, many leaders remain unsure about what blockchain has to do with health care. Proof of validation is an important step in scaling up this framework and making it applicable to the real world [69].

Second, sorting and tiering alerts based on severity, for example, sorting into 3 tiers: low, mid, and high, are needed as a part of future work to validate MedAlert. This is to determine which low-level alerts are suitable for patients because clinicians tend to accept high-severity alerts slightly more often than mid- or low-severity interaction alerts [11]. However, the process of tiering alerts is highly subjective when it comes to deciding which alerts are considered low level and time consuming for all medical experts before reaching a common consensus. Thus, this initial step in selecting which alerts are to be shared with patients can be challenging.

Decentralized identity management is an alternative way of verifying and authenticating users. It eliminates the limitations of centralized identity systems, helps achieve compliance with the most comprehensive national data protection laws, and returns ownership and control of identity data back to the individual. Various decentralized identity management systems exist that provide solutions using a distributed ledger technology. Evernym [70], uPort [71], and Sovrin [70] are some examples

of identity projects that are working on decentralized identity platforms. However, these sophisticated solutions are still at a provisional stage, where more validation, discussion, and investigation are needed [60].

### Limitations

MedAlert is suitable for a specific group of users. Collaborative decision making may be challenging for patients who are less technology savvy, particularly for elderly patients, who may not be able to use MedAlert effectively. For example, the steps where patients need to verify and authenticate themselves and thus gain access to alerts could be confusing for the elderly and may induce unnecessary stress on them. MedAlert is not suitable for in-patients either where they require constant monitoring. This is because they may not be able to provide a response when they are unwell in the hospital.

Directing low-level alerts to patients may create ethical issues where the responsibility is indirectly shifted on to them in cases when they provide incorrect responses. In a study conducted in medical centers in the Netherlands where alerts were directed to nurses, despite improvements in efficiency and effectiveness, the study concluded that such alerts should not be directed to nurses [40]. It is difficult to find the right balance of responsibilities between clinicians, nurses, and patients in a collaborative decision-making process.

Privacy concerns are covered by the GDPR. Storing digital assets, such as medical records on blockchain, could violate personal privacy. Although MedAlert only stores patients' metadata on blockchain, it is not entirely anonymous. Malicious acts include attempting to learn about and identify actual personal patients based on the pseudo-anonymous information on blockchain. In addition, the permanent storage of information, both data and metadata, belonging to a person could violate GDPR (Art. 17 Right to erasure or to be forgotten) in cases when users want to have their data completely erased or deleted.

### Conclusions

CDS supports the decision-making process in preventing medication errors by generating alerts. Clinicians can now rely on these alerts along with their knowledge and past experience to avoid medication errors. Due to the low specificity and highly restricted modifications of the CDS setting, a high volume of irrelevant alerts has caused clinicians to experience alert fatigue. This results in a high overriding rate, which can cause medication errors.

From our scoping review, we found different methods of reducing the number of alerts, such as machine learning algorithms and blockchain technology, by filtering out irrelevant alerts. We developed a different solution that is similar to what medical experts pointed out, where improving alert fatigue should go beyond just reducing the total number of alerts.

In line with this idea, we designed MedAlert, a blockchain-based solution, by sharing low-level alerts with patients where clinicians typically have a greater tendency to override low-level alerts. The goal is to ensure that alerts catch the attention of both patients and clinicians, thus preventing medication errors, instead of being habitually overridden. In our own work, we

introduced a second layer by engaging patients in providing a response and making them, at least, partially responsible for alert verification. This second layer reduces alert fatigue of clinicians and, at the same time, engages patients in the collaborative process, making it harder for medication errors to occur.

Other potential advantages of MedAlert over other frameworks include ensuring a greater degree of patient privacy and the ability to establish a new communication layer between patients and clinicians. Smart contracts and the use of BankID (federated identity management) are useful in authenticating patients and ensuring that the right person receives the alert.

Directing alerts to patients faces challenges such as finding a balance between patients and clinicians without raising ethical issues. This solution may not be suitable for elderly patients or in-patients where they require constant monitoring. Sorting and tiering the alerts based on levels of severity is also challenging because it is subjective and may vary between different panels of medical experts.

For the health care sector to benefit from the potential value of this innovative idea, future work, for example, on the validation of MedAlert based on real-world scenarios, such as the degree of compliance with GDPR, is needed. Providing documented evaluations of the performance of MedAlert is crucial to gain the support of health care leaders in nurturing this idea as a potential solution to reducing alert fatigue.

---

## Acknowledgments

The authors would like to thank Egil Utheim from Innlandet Hospital, Norway, for presenting insights and current issues in health care in Norway, which inspired the authors and led to the development of this paper. This study is funded by the NTNU digital transformation project: Trust and transparency in digital society through blockchain technology.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Search strings.

[DOC File, 59 KB - [jmir\\_v22i10e22013\\_app1.doc](#)]

---

## Multimedia Appendix 2

Summary of selected literature.

[DOCX File, 25 KB - [jmir\\_v22i10e22013\\_app2.docx](#)]

---

## References

1. Berner E, La LT. Overview of Clinical Decision Support Systems. In: Berner ES, editor. *Clinical Decision Support Systems: Theory and Practice*. New York, NY: Springer New York; 2007:3-22.
2. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation. *Health Aff (Millwood)* 2011 Dec;30(12):2310-2317. [doi: [10.1377/hlthaff.2010.1111](#)] [Medline: [22147858](#)]
3. Morris ME, Aguilera A. Mobile, Social, and Wearable Computing and the Evolution of Psychological Practice. *Prof Psychol Res Pr* 2012 Dec;43(6):622-626. [doi: [10.1037/a0029041](#)] [Medline: [25587207](#)]
4. Kamel Boulos MN, Wilson JT, Clauson KA. Geospatial blockchain: promises, challenges, and scenarios in health and healthcare. *Int J Health Geogr* 2018 Dec 05;17(1):25 [FREE Full text] [doi: [10.1186/s12942-018-0144-x](#)] [Medline: [29973196](#)]
5. Smithburger PL, Buckley MS, Bejian S, Burenheide K, Kane-Gill SL. A critical evaluation of clinical decision support for the detection of drug-drug interactions. *Expert Opin Drug Saf* 2011 Nov;10(6):871-882. [doi: [10.1517/14740338.2011.583916](#)] [Medline: [21542665](#)]
6. Kassakian SZ, Yackel TR, Gorman PN, Dorr DA. Clinical decisions support malfunctions in a commercial electronic health record. *Appl Clin Inform* 2017 Sep 06;8(3):910-923. [doi: [10.4338/ACI-2017-01-RA-0006](#)] [Medline: [28880046](#)]
7. Van DSH. Errors related to alert fatigue. *Safety of Health IT: Clinical Case Studies*: Springer International Publishing; 2016:41-54.
8. Van DSH. Drug safety alerting in computerized physician order entry: Unraveling and counteracting alert fatigue. Amsterdam: Erasmus University Rotterdam; 2009:E.
9. Rayo MF, Kowalczyk N, Liston BW, Sanders EB, White S, Patterson ES. Comparing the Effectiveness of Alerts and Dynamically Annotated Visualizations (DAVs) in Improving Clinical Decision Making. *Hum Factors* 2015 Sep;57(6):1002-1014. [doi: [10.1177/0018720815585666](#)] [Medline: [25957043](#)]



10. McGreevey JD, Mallozzi CP, Perkins RM, Shelov E, Schreiber R. Reducing Alert Burden in Electronic Health Records: State of the Art Recommendations from Four Health Systems. *Appl Clin Inform* 2020 Jan;11(1):1-12. [doi: [10.1055/s-0039-3402715](https://doi.org/10.1055/s-0039-3402715)] [Medline: [31893559](https://pubmed.ncbi.nlm.nih.gov/31893559/)]
11. Isaac T, Weissman JS, Davis RB, Massagli M, Cyrulik A, Sands DZ, et al. Overrides of medication alerts in ambulatory care. *Arch Intern Med* 2009 Mar 09;169(3):305-311. [doi: [10.1001/archinternmed.2008.551](https://doi.org/10.1001/archinternmed.2008.551)] [Medline: [19204222](https://pubmed.ncbi.nlm.nih.gov/19204222/)]
12. van DSH, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc* 2006;13(2):138-147. [doi: [10.1197/jamia.M1809](https://doi.org/10.1197/jamia.M1809)] [Medline: [16357358](https://pubmed.ncbi.nlm.nih.gov/16357358/)]
13. Chaparro JD, Hussain C, Lee JA, Hehmeyer J, Nguyen M, Hoffman J. Reducing Interruptive Alert Burden Using Quality Improvement Methodology. *Appl Clin Inform* 2020 Jan;11(1):46-58. [doi: [10.1055/s-0039-3402757](https://doi.org/10.1055/s-0039-3402757)] [Medline: [31940671](https://pubmed.ncbi.nlm.nih.gov/31940671/)]
14. Mello MM, Chandra A, Gawande AA, Studdert DM. National costs of the medical liability system. *Health Aff (Millwood)* 2010 Sep;29(9):1569-1577 [FREE Full text] [doi: [10.1377/hlthaff.2009.0807](https://doi.org/10.1377/hlthaff.2009.0807)] [Medline: [20820010](https://pubmed.ncbi.nlm.nih.gov/20820010/)]
15. Segal G, Segev A, Brom A, Lifshitz Y, Wasserstrum Y, Zimlichman E. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting. *J Am Med Inform Assoc* 2019 Dec 01;26(12):1560-1565. [doi: [10.1093/jamia/ocz135](https://doi.org/10.1093/jamia/ocz135)] [Medline: [31390471](https://pubmed.ncbi.nlm.nih.gov/31390471/)]
16. Walker GH, Waterfield S, Thompson P. All at sea: An ergonomic analysis of oil production platform control rooms. *International Journal of Industrial Ergonomics* 2014 Sep;44(5):723-731. [doi: [10.1016/j.ergon.2014.08.001](https://doi.org/10.1016/j.ergon.2014.08.001)]
17. Narayanaswamy B, Balaji B, Gupta R, Agarwal Y. editors. Data driven investigation of faults in HVAC systems with Model, Cluster and Compare (MCC): Association for Computing Machinery, Inc 2014:A. [doi: [10.1145/2674061.2674067](https://doi.org/10.1145/2674061.2674067)]
18. Simpao AF, Ahumada LM, Desai BR, Bonafide CP, Gálvez JA, Rehman MA, et al. Optimization of drug-drug interaction alert rules in a pediatric hospital's electronic health record system using a visual analytics dashboard. *J Am Med Inform Assoc* 2015 Mar;22(2):361-369. [doi: [10.1136/amiajnl-2013-002538](https://doi.org/10.1136/amiajnl-2013-002538)] [Medline: [25318641](https://pubmed.ncbi.nlm.nih.gov/25318641/)]
19. Saiyed SM, Greco PJ, Fernandes G, Kaelber DC. Optimizing drug-dose alerts using commercial software throughout an integrated health care system. *J Am Med Inform Assoc* 2017 Nov 01;24(6):1149-1154. [doi: [10.1093/jamia/ocx031](https://doi.org/10.1093/jamia/ocx031)] [Medline: [28444383](https://pubmed.ncbi.nlm.nih.gov/28444383/)]
20. Lundberg L. Electronic prescribing does not prevent medication errors. *sciencenorwaycom Internet*. Norway: sciencenorway; 2019 Oct 19. URL: <https://partner.sciencenorway.no/e-health-research-health-services-medication/electronic-prescribing-does-not-prevent-medication-errors/1579788> [accessed 2020-03-01]
21. Morris ME, Aguilera A. Mobile, Social, and Wearable Computing and the Evolution of Psychological Practice. *Prof Psychol Res Pr* 2012 Dec;43(6):622-626 [FREE Full text] [doi: [10.1037/a0029041](https://doi.org/10.1037/a0029041)] [Medline: [25587207](https://pubmed.ncbi.nlm.nih.gov/25587207/)]
22. The NDOE. The Norwegian Directorate of EHealth. 2019. URL: <https://ehelse.no> [accessed 2020-02-12]
23. Hospital I. Blockchain meeting: Distributive Value chain In: Technology NUoS, editor. In: Blockchain meeting: Distributive Value chain. 2019 Nov 15 Presented at: Blockchain meeting: Distributive Value chain; 15 November 2019; Gjøvik, Norway.
24. Wright A, Sittig DF, Ash JS, Feblowitz J, Meltzer S, McMullen C, et al. Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems. *J Am Med Inform Assoc* 2011 May 01;18(3):232-242 [FREE Full text] [doi: [10.1136/amiajnl-2011-000113](https://doi.org/10.1136/amiajnl-2011-000113)] [Medline: [21415065](https://pubmed.ncbi.nlm.nih.gov/21415065/)]
25. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data. *Comput Struct Biotechnol J* 2018;16:267-278. [doi: [10.1016/j.csbj.2018.07.004](https://doi.org/10.1016/j.csbj.2018.07.004)] [Medline: [30108685](https://pubmed.ncbi.nlm.nih.gov/30108685/)]
26. Morris ME, Aguilera A. Mobile, Social, and Wearable Computing and the Evolution of Psychological Practice. *Prof Psychol Res Pr* 2012 Dec;43(6):622-626. [doi: [10.1037/a0029041](https://doi.org/10.1037/a0029041)] [Medline: [25587207](https://pubmed.ncbi.nlm.nih.gov/25587207/)]
27. Tandon A, Dhir A, Islam AN, Mäntymäki M. Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda. *Computers in Industry* 2020 Nov;122:103290. [doi: [10.1016/j.compind.2020.103290](https://doi.org/10.1016/j.compind.2020.103290)]
28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 2018 Oct 02;169(7):467-473. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
29. Chen J, Chokshi S, Hegde R, Gonzalez J, Iturrate E, Aphinyanaphongs Y, et al. Development, Implementation, and Evaluation of a Personalized Machine Learning Algorithm for Clinical Decision Support: Case Study With Shingles Vaccination. *J Med Internet Res* 2020 Apr 29;22(4):e16848 [FREE Full text] [doi: [10.2196/16848](https://doi.org/10.2196/16848)] [Medline: [32347813](https://pubmed.ncbi.nlm.nih.gov/32347813/)]
30. Soundararajan V, McDaniel B, Shin J, Sneha S, Soundararajan V. Leveraging blockchain to improve clinical decision support systems. 2019 May 24 Presented at: CEUR Workshop Proceedings. ;2408; 2019; Lisbon, Portugal.
31. Heringa M, van der Heide A, Floor-Schreuderling A, De Smet PAGM, Bouvy ML. Better specification of triggers to reduce the number of drug interaction alerts in primary care. *Int J Med Inform* 2018 Jan;109:96-102. [doi: [10.1016/j.jimedinf.2017.11.005](https://doi.org/10.1016/j.jimedinf.2017.11.005)] [Medline: [29195711](https://pubmed.ncbi.nlm.nih.gov/29195711/)]
32. Heringa M, Siderius H, Floor-Schreuderling A, de Smet PAGM, Bouvy ML. Lower alert rates by clustering of related drug interaction alerts. *J Am Med Inform Assoc* 2017 Jan;24(1):54-59. [doi: [10.1093/jamia/ocw049](https://doi.org/10.1093/jamia/ocw049)] [Medline: [27107437](https://pubmed.ncbi.nlm.nih.gov/27107437/)]
33. Khalifa M, Zabani I. Improving Utilization of Clinical Decision Support Systems by Reducing Alert Fatigue: Strategies and Recommendations. *Stud Health Technol Inform* 2016;226:51-54. [Medline: [27350464](https://pubmed.ncbi.nlm.nih.gov/27350464/)]
34. Jafarpour B, Abidi SR, Ahmad AM, Abidi SSR. INITIATE: An Intelligent Adaptive Alert Environment. *Stud Health Technol Inform* 2015;216:285-289. [Medline: [26262056](https://pubmed.ncbi.nlm.nih.gov/26262056/)]

35. Carli D, Fahrni G, Bonnabry P, Lovis C. Quality of Decision Support in Computerized Provider Order Entry: Systematic Literature Review. *JMIR Med Inform* 2018 Jan 24;6(1):e3 [[FREE Full text](#)] [doi: [10.2196/medinform.7170](#)] [Medline: [29367187](#)]
36. Powers EM, Shiffman RN, Melnick ER, Hickner A, Sharifi M. Efficacy and unintended consequences of hard-stop alerts in electronic health record systems: a systematic review. *J Am Med Inform Assoc* 2018 Nov 01;25(11):1556-1566 [[FREE Full text](#)] [doi: [10.1093/jamia/ocy112](#)] [Medline: [30239810](#)]
37. Hussain MI, Reynolds TL, Zheng K. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review. *J Am Med Inform Assoc* 2019 Oct 01;26(10):1141-1149 [[FREE Full text](#)] [doi: [10.1093/jamia/ocz095](#)] [Medline: [31206159](#)]
38. Bryant AD, Fletcher GS, Payne TH. Drug interaction alert override rates in the Meaningful Use era: no evidence of progress. *Appl Clin Inform* 2014;5(3):802-813 [[FREE Full text](#)] [doi: [10.4338/ACI-2013-12-RA-0103](#)] [Medline: [25298818](#)]
39. Sheffer J, Cvach M, Edworthy JR, Patterson ES. A Roundtable Discussion: Improving the 'Alarm Problem' Will Require Much More Than Just Reducing the Number of Alarms. *Biomed Instrum Technol* 2018;52(6):454-461. [doi: [10.2345/0899-8205-52.6.454](#)] [Medline: [30479154](#)]
40. van der Sijs H, Lammers L, van den Tweel A, Aarts J, Berg M, Vulto A, et al. Time-dependent drug-drug interaction alerts in care provider order entry: software may inhibit medication error reductions. *J Am Med Inform Assoc* 2009;16(6):864-868 [[FREE Full text](#)] [doi: [10.1197/jamia.M2810](#)] [Medline: [19717806](#)]
41. Gordon WJ, Catalini C. Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability. *Comput Struct Biotechnol J* 2018;16:224-230 [[FREE Full text](#)] [doi: [10.1016/j.csbj.2018.06.003](#)] [Medline: [30069284](#)]
42. Hathaliya JJ, Tanwar S. An exhaustive survey on security and privacy issues in Healthcare 4.0. *Computer Communications* 2020 Mar;153:311-335. [doi: [10.1016/j.comcom.2020.02.018](#)]
43. Pandey AK, Khan AI, Abushark YB, Alam MM, Agrawal A, Kumar R, et al. Key Issues in Healthcare Data Integrity: Analysis and Recommendations. *IEEE Access* 2020;8:40612-40628. [doi: [10.1109/access.2020.2976687](#)]
44. Siew A. HSA blood donor data leak: When 'sorry' may not be enough. *Today Online*. Singapore: Today Online; 2019 Mar 21. URL: <https://www.todayonline.com/commentary/hsa-blood-donor-data-leak-when-sorry-may-not-be-enough> [accessed 2020-03-03]
45. Xiao Z, Li Z, Liu Y, Feng L, Zhang W, Lertwuthikarn T. EMRShare: A Cross-Organizational Medical Data Sharing and Management Framework Using Permissioned Blockchain C3 - Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS. 2019 Feb 21 Presented at: 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS); 11-13 December 2018; Singapore. [doi: [10.1109/PADSW.2018.8645049](#)]
46. Shen B, Guo J, Yang Y. MedChain: Efficient Healthcare Data Sharing via Blockchain. *Applied Sciences* 2019 Mar 22;9(6):1207. [doi: [10.3390/app9061207](#)]
47. Fan K, Wang S, Ren Y, Li H, Yang Y. MedBlock: Efficient and Secure Medical Data Sharing Via Blockchain. *J Med Syst* 2018 Jun 21;42(8):136. [doi: [10.1007/s10916-018-0993-7](#)] [Medline: [29931655](#)]
48. Das M, Tao X, Cheng J, editors. A Secure and Distributed Construction Document Management System Using Blockchain 2021;Cham:850-862. [doi: [10.1007/978-3-030-51295-8\\_59](#)]
49. Alam BM, Wang G, Zaman A, Tao H, Wang T, Hassan M. Blockchain and big data to transform the healthcare. 2018 Presented at: ICDPA2018: Proceedings of the International Conference on Data Processing and Applications; May 12-14, 2018; Guangdong, China p. 62-68. [doi: [10.1145/3224207.3224220](#)]
50. Glaser F. Pervasive Decentralisation of Digital Infrastructures: A Framework for Blockchain enabled System and Use Case Analysis. 2017 Jan Presented at: Proceedings of the 50th Hawaii International Conference on System Sciences; 2017; Hawaii p. 1543-1552. [doi: [10.24251/hicss.2017.186](#)]
51. Nærland K, Müller-Bloch C, Beck R, Palmund S. Blockchain to Rule the Waves - Nascent Design Principles for Reducing Risk and Uncertainty in Decentralized Environments. Presented at: International Conference on Information Systems (ICIS); 2017; Seoul URL: <https://core.ac.uk/download/pdf/145226366.pdf> [doi: [10.17705/4icis](#)]
52. Wang Y, Han JH, Beynon-Davies P. Understanding blockchain technology for future supply chains: a systematic literature review and research agenda. *Supp Chain Mngmnt* 2019 Jan 14;24(1):62-84. [doi: [10.1108/scm-03-2018-0148](#)]
53. Drescher D. Blockchain basics : a non-technical introduction in 25 steps. New York: Apress; Mar 16, 2017:273.
54. Antonopoulos A. Mastering Bitcoin: Unlocking Digital Cryptocurrencies. Newton, MA,USA: O'Reilly Medi; 2015.
55. GDPR. GDPR checklist for data controllers. URL: <https://gdpr.eu/checklist/> [accessed 2020-06-29]
56. DeSalvo K. RE: Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap. 2015. URL: [https://www.healthit.gov/sites/default/files/comments\\_upload/tascet\\_comments\\_to\\_onc\\_interoperability\\_roadmap\\_2015\\_04\\_01.pdf](https://www.healthit.gov/sites/default/files/comments_upload/tascet_comments_to_onc_interoperability_roadmap_2015_04_01.pdf) [accessed 2020-05-16]
57. Torres J, Nogueira M, Pujolle G. A Survey on Identity Management for the Future Network. *IEEE Commun. Surv. Tutorials* 2013;15(2):787-802. [doi: [10.1109/surv.2012.072412.00129](#)]
58. Ning H, Liu X, Ye X, He J, Zhang W, Daneshmand M. Edge Computing-Based ID and nID Combined Identification and Resolution Scheme in IoT. *IEEE Internet Things J* 2019 Aug;6(4):6811-6821. [doi: [10.1109/jiot.2019.2911564](#)]

59. Camp L. Digital identity. *IEEE Technol. Soc. Mag* 2004;23(3):34-41. [doi: [10.1109/mtas.2004.1337889](https://doi.org/10.1109/mtas.2004.1337889)]
60. Bouras MA, Lu Q, Zhang F, Wan Y, Zhang T, Ning H. Distributed Ledger Technology for eHealth Identity Privacy: State of The Art and Future Perspective. *Sensors (Basel)* 2020 Jan 15;20(2):- [FREE Full text] [doi: [10.3390/s20020483](https://doi.org/10.3390/s20020483)] [Medline: [31952172](https://pubmed.ncbi.nlm.nih.gov/31952172/)]
61. Viswanathan H, Chen B, Pompili D. Research challenges in computation, communication, and context awareness for ubiquitous healthcare. *IEEE Commun. Mag* 2012 May;50(5):92-99. [doi: [10.1109/mcom.2012.6194388](https://doi.org/10.1109/mcom.2012.6194388)]
62. Jensen HH, Hewett N. Inclusive Deployment of Blockchain for Supply Chains Part 2: Trustworthy verification of digital identities. *World Economic Forum Internet.: World Economic Forum*; 2019 Apr 24. URL: [www3.weforum.org/docs/WEF\\_Trustworthy\\_Verification\\_of\\_Digital\\_Identities\\_2019.pdf](http://www3.weforum.org/docs/WEF_Trustworthy_Verification_of_Digital_Identities_2019.pdf) [accessed 2020-03-12]
63. BankID. About Us. URL: <https://www.bankid.no/en/private/about-us/>
64. Smart-ID. Smart-ID is a smart way to identify yourself. URL: <https://www.smart-id.com/>
65. Kim H, Laskowski M. A Perspective on Blockchain Smart Contracts: Reducing Uncertainty and Complexity in Value Exchange. : IEEE; 2017 Sep 17 Presented at: 2017 26th International Conference on Computer Communication and Networks (ICCCN); 31 July-3 Aug. 2017; Vancouver, BC, Canada. [doi: [10.1109/icccn.2017.8038512](https://doi.org/10.1109/icccn.2017.8038512)]
66. Pongnumkul S, Siripanpornchana C, Thajchayapong S. Performance Analysis of Private Blockchain Platforms in Varying Workloads. : IEEE; 2018 Sep 18 Presented at: 2017 26th International Conference on Computer Communication and Networks (ICCCN); 31 July-3 Aug. 2017; Vancouver, BC, Canada. [doi: [10.1109/icccn.2017.8038517](https://doi.org/10.1109/icccn.2017.8038517)]
67. Porter SC, Manzi SF, Volpe D, Stack AM. Getting the data right: information accuracy in pediatric emergency medicine. *Qual Saf Health Care* 2006 Aug;15(4):296-301 [FREE Full text] [doi: [10.1136/qshc.2005.017442](https://doi.org/10.1136/qshc.2005.017442)] [Medline: [16885256](https://pubmed.ncbi.nlm.nih.gov/16885256/)]
68. Wan PK, Huang L, Holtskog H. Blockchain-Enabled Information Sharing Within a Supply Chain: A Systematic Literature Review. *IEEE Access* 2020;8:49645-49656. [doi: [10.1109/access.2020.2980142](https://doi.org/10.1109/access.2020.2980142)]
69. Saberi S, Kouhizadeh M, Sarkis J, Shen L. Blockchain technology and its relationships to sustainable supply chain management. *International Journal of Production Research* 2018 Oct 17;57(7):2117-2135. [doi: [10.1080/00207543.2018.1533261](https://doi.org/10.1080/00207543.2018.1533261)]
70. Windley P, Reed D. Sovrin: A Protocol and Token for Self-Sovereign Identity and Decentralized Trust. Utah: Sovrin Foundation; 2018 Jan 16. URL: <https://sovrin.org/wp-content/uploads/Sovrin-Protocol-and-Token-White-Paper.pdf>
71. Dunphy P, Petitcolas FA. A First Look at Identity Management Schemes on the Blockchain. *IEEE Secur. Privacy* 2018 Jul;16(4):20-29. [doi: [10.1109/msp.2018.3111247](https://doi.org/10.1109/msp.2018.3111247)]

## Abbreviations

**API:** application programming interface  
**CDS:** clinical decision support  
**EHR:** electronic health record  
**GDPR:** General Data Protection Regulation  
**IT:** information technology  
**NTNU:** Norwegian University of Science and Technology  
**REST:** representational state transfer

*Edited by G Eysenbach; submitted 07.07.20; peer-reviewed by C Esposito, R Sutton; comments to author 02.08.20; revised version received 08.09.20; accepted 12.09.20; published 28.10.20.*

### *Please cite as:*

Wan PK, Satybaldy A, Huang L, Holtskog H, Nowostawski M  
 Reducing Alert Fatigue by Sharing Low-Level Alerts With Patients and Enhancing Collaborative Decision Making Using Blockchain Technology: Scoping Review and Proposed Framework (MedAlert)  
*J Med Internet Res* 2020;22(10):e22013  
 URL: <http://www.jmir.org/2020/10/e22013/>  
 doi: [10.2196/22013](https://doi.org/10.2196/22013)  
 PMID: [33112253](https://pubmed.ncbi.nlm.nih.gov/33112253/)

©Paul Kengfai Wan, Abylay Satybaldy, Lizhen Huang, Halvor Holtskog, Mariusz Nowostawski. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org/>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Raising the Digital Profile of Facial Palsy: National Surveys of Patients' and Clinicians' Experiences of Changing UK Treatment Pathways and Views on the Future Role of Digital Technology

Ala Szczepura<sup>1</sup>, BA, DPhil; Nikki Holliday<sup>2</sup>, BSc, MSc; Catriona Neville<sup>3</sup>, BSc; Karen Johnson<sup>4</sup>; Amir Jahan Khan Khan<sup>5</sup>, BSc, MSc, PhD; Samuel W Oxford<sup>6</sup>, MSc, PhD; Charles Nduka<sup>3</sup>, MA, MD, FRCS

<sup>1</sup>Faculty Health & Life Sciences, Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom

<sup>2</sup>Health & Life Sciences, Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom

<sup>3</sup>Queen Victoria Hospital NHS Foundation Trust, East Grinstead, West Sussex, United Kingdom

<sup>4</sup>Facial Palsy UK (Charity), Peterborough, United Kingdom

<sup>5</sup>Department of Economics, Institute of Business Administration (IBA), Karachi, Pakistan

<sup>6</sup>Exercise & Life Sciences, Faculty Health & Life Sciences, Centre for Sport, Coventry University, Coventry, United Kingdom

**Corresponding Author:**

Ala Szczepura, BA, DPhil  
Faculty Health & Life Sciences  
Centre for Intelligent Healthcare  
Coventry University  
Richard Crossman Building  
Priory Street  
Coventry, CV1 5FB, UK  
United Kingdom  
Phone: 44 7557425463  
Email: [ab5794@coventry.ac.uk](mailto:ab5794@coventry.ac.uk)

## Abstract

**Background:** Facial nerve palsy leaves people unable to move muscles on the affected side of their face. Challenges exist in patients accessing facial neuromuscular retraining (NMR), a therapy used to strengthen muscle and improve nerve function. Access to therapy could potentially be improved through the use of digital technology. However, there is limited research available on patients' and clinicians' views about the potential benefits of such telerehabilitation based on their lived experiences of treatment pathways.

**Objective:** This study aims to gather information about facial palsy treatment pathways in the United Kingdom, barriers to accessing NMR, factors influencing patient adherence, measures used to monitor recovery, and the potential value of emerging wearable digital technology.

**Methods:** Separate surveys of patients with facial palsy and facial therapy specialists were conducted. Questionnaires explored treatment pathways and views on telerehabilitation, were co-designed with users, and followed a similar format to enable cross-referencing of responses. A follow-up survey of national specialists investigated methods used to monitor recovery in greater detail. Analysis of quantitative data was conducted allowing for data distribution. Open-text responses were analyzed using thematic content analysis.

**Results:** A total of 216 patients with facial palsy and 25 specialist therapists completed the national surveys. Significant variations were observed in individual treatment pathways. Patients reported an average of 3.27 (SD 1.60) different treatments provided by various specialists, but multidisciplinary team reviews were rare. For patients diagnosed most recently, there was evidence of more rapid initial prescribing of corticosteroids (prednisolone) and earlier referral for NMR therapy. Barriers to NMR referral included difficulties accessing funding, shortage of specialist therapists, and limited awareness of NMR among general practitioners. Patients traveled long distances to reach an NMR specialist center; 9% (8/93) of adults reported traveling  $\geq 15$  miles. The thematic content analysis demonstrates positive attitudes to the introduction of digital technology, with similar incentives and barriers identified by both patients and clinicians. The follow-up survey of 28 specialists uncovered variations in the measures currently used to monitor recovery and no agreed definitions of a clinically significant change for any of these. The main barriers to NMR adherence identified by patients and therapists could all be addressed by using suitable real-time digital technology.



**Conclusions:** The study findings provide valuable information on facial palsy treatment pathways and views on the future introduction of digital technology. Possible ways in which emerging sensor-based digital technology can improve rehabilitation and provide more rigorous evidence on effectiveness are described. It is suggested that one legacy of the COVID-19 pandemic will be lower organizational barriers to this introduction of digital technology to assist NMR delivery, especially if cost-effectiveness can be demonstrated.

(*J Med Internet Res* 2020;22(10):e20406) doi:[10.2196/20406](https://doi.org/10.2196/20406)

## KEYWORDS

Bell palsy; facial nerve paralysis; patient experience; treatment pathway; facial exercise therapy; neuromuscular retraining; treatment adherence; digital technology; outcome measures; telerehabilitation; biosensors; COVID-19

## Introduction

### Facial Palsy

Bell palsy is an acute unilateral paralysis of the facial nerve, resulting in a patient partially, or completely, losing the ability to voluntarily move facial muscles on the affected side of the face [1]. It is the most common acute disorder affecting a single nerve, and its cause is unknown [2]. Each year, the condition affects 11 to 40 people per 100,000 in the population, most commonly in the age group of 30 to 45 years [3]. The annual incidence of Bell palsy in the United Kingdom is currently 37.7 per 100,000 population [4]. Bell palsy represents only approximately 60% of all facial nerve paralysis (FNP) cases [5]. The total number of FNP cases occurring annually in the United Kingdom is estimated to be at least 22,500, and 1 in 60 individuals will be affected over the course of their lifetime [6,7].

Epidemiological studies indicate that this neurological condition occurs more commonly in those with diabetes, obesity, hypertension, and upper respiratory conditions and people who are immunocompromised or pregnant [2,6,8] or following infection by a virus such as herpes simplex [9]. Data from the United States show a recent rise in incidence, possibly linked to increasing rates of herpes infections [10]. Without intervention, some patients will show an element of recovery within 2 to 3 weeks and complete recovery within 3 to 4 months [2,3]. However, although normal facial function is completely restored in approximately 70% of cases, 30% will have a poor recovery [11,12] with facial disfigurement and sometimes facial pain [3,13], and up to 16% of those affected will have residual involuntary movements known as synkinesis [3]. Research shows that people with these residual deficits experience a long-term reduction in quality of life, psychological distress, depression, and social alienation, often relinquishing a previous public-facing role [3,13-16]. As a result, patients with FNP continue to have relatively low public visibility, unless a high-profile international star reveals their own diagnosis [17].

### Available Treatments

Although various treatments are available, uncertainty exists regarding the effectiveness of many of these. Cochrane systematic reviews have confirmed the effectiveness and cost-effectiveness of corticosteroids (prednisolone) administered within 72 hours of onset of symptoms [18-20]. Beyond this initial treatment, for those with incomplete recovery, there are a number of medical options available. Various surgical

procedures, together with botulinum toxin injections can attempt to normalize facial appearance [21-25]. However, a Cochrane review of surgical interventions has reported that there is insufficient evidence to decide whether such procedures are beneficial and has also concluded that further trials are unlikely [26].

Physical rehabilitation therapy can be used as an adjunct to medical treatments. The use of facial neuromuscular retraining (NMR) to strengthen muscle and improve nerve function has been evaluated more than other physical therapies [27-32]. A 2011 Cochrane review has concluded that there is some evidence that NMR can improve facial function (for moderate nerve paralysis and chronic cases) and reduce sequelae in acute cases, although it was recommended that both need to be confirmed in randomized controlled trials [33]. An update of this systematic review is currently underway [34]. A recent review of physical therapy combined with standard drug treatment (SDT) has reported evidence of positive effects on grade and time to recovery compared with SDT alone [35].

### Practice Guidelines and Patient Involvement

In the context of an incomplete evidence base, current international guidelines highlight the need to consider patients' experiences and preferences [2,36]. Recent clinical practice guidelines from the United States, which conclude that physical therapy can provide potential functional and psychological benefit, add that there is a "large role for shared decision making" [2]. Although Canadian guidelines make no recommendation regarding the use of facial NMR in the acute phase, owing to a lack of good quality trials and risk of bias, its use is suggested for patients who do not have complete facial recovery [36]. In the United Kingdom, clinical guidelines produced by the National Institute for Health and Care Excellence (NICE) recommend rapid initial medication (prednisolone) and referral to a range of hospital-based medical specialists [37]. Guidelines for commissioning neurology services in the United Kingdom also include a recommendation to consider new *transformational technologies* [38].

### Objectives

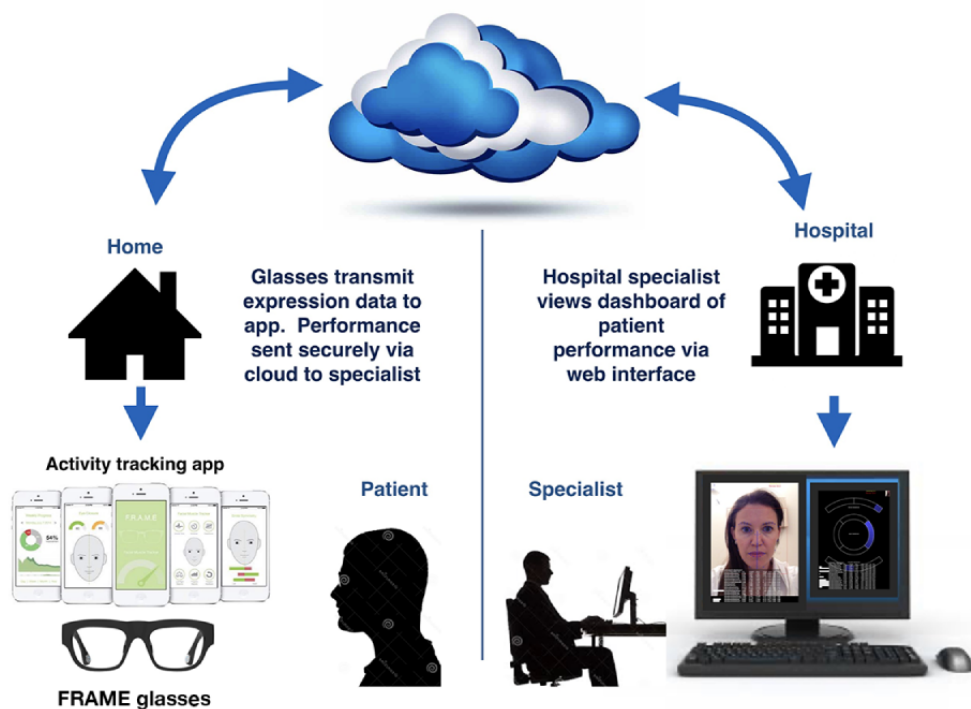
The potential for the "digital patient" to transform the delivery of care has been demonstrated in many clinical areas [39], but there is a lack of evidence for FNP. The aim of this study, conducted in collaboration with patients and specialist clinicians, is to gather evidence about FNP treatment pathways in the United Kingdom, the referral process and timing of NMR, current outcomes used to monitor recovery, and the potential



role of digital technology to assist in rehabilitation. The study was conducted as part of a research program (Facial Remote Activity Monitoring Eyewear [FRAME]) funded by the UK National Institute for Health Research (NIHR). The FRAME program aims to develop inconspicuous miniaturized sensor

devices in spectacle frames to measure facial movement, providing biofeedback to patients and access for clinicians to outcome data, and to conduct early health technology assessment of the wearable FRAME digital technology shown in Figure 1 [40].

**Figure 1.** Facial Remote Activity Monitoring Eyewear (FRAME) system overview.



## Methods

### National Surveys

To provide context for technology introduction, national surveys were conducted to explore patients' and specialist therapists' experiences of FNP care pathways in the United Kingdom and the potential role of digital technology. Separate questionnaires were co-designed for patients and therapists, piloted, and refined following feedback. Each questionnaire included a mix of closed and open-ended questions (Multimedia Appendix 1) and followed a similar format to enable cross-referencing of some responses [41]. Questionnaires collected demographic details, information on treatment pathways, and ratings of the importance of treatments personally experienced. Respondents were encouraged to add textual comments to expand on responses to closed questions. Further open-ended questions explored the potential value of the emerging FRAME technology. Both questionnaires were uploaded on a web-based platform.

### Recruitment to National Surveys

For patients, an open recruitment strategy was adopted to achieve geographical spread. People with experience of FNP were recruited in collaboration with the national charity, Facial Palsy UK [42]; the patient survey was advertised widely, including via social media. Recruitment of specialist therapists was coordinated through a nation-wide professional group,

Facial Therapy Specialists UK [43]; all members were emailed a personal invitation containing a link to the questionnaire, and the initial email was followed by 2 reminders over a period of 5 months.

### Follow-Up Survey

A preliminary analysis of responses uncovered a range of methods used to report treatment outcomes. Therefore, a further survey was conducted to examine these in greater detail, especially any use of validated scales, and whether there is consensus on the definition of a clinically significant change for these measures. A convenience sample of 50 clinicians attending the Facial Therapy Specialists Annual Meeting held in London (October 1, 2018) were invited to complete this questionnaire.

### Analysis of Responses

The closed questions were analyzed using SPSS (version 25, IBM Corporation). Response patterns were summarized using mean and SD or median and IQR, depending on data distribution. Certain variables were grouped to explore changes over time (eg, time since diagnosis). Open-text comments expanding on responses to closed questions and text replying to open questions were analyzed using thematic content analysis [44]. Inductive coding was used, following a flexible analysis approach that helped account for any further themes emerging during the coding process [45]. Data were coded and analyzed

for thematic patterns and meanings within the data until saturation was reached.

### Ethics

Ethical approval was granted by the Health and Life Sciences Research Ethics Committee, University of Coventry (ref: P48908).

## Results

### Respondent Rate and Participants

#### *National Surveys*

The response rate (RR) for patients was calculated based on the number of people accessing the invitation link (after viewing an advertisement) versus the number of patients completing the questionnaire (216/216, 100% RR). Patients were resident in England (all 9 English regions), Scotland, Ireland, and Wales.

The RR for clinicians was based on the number of UK therapists contacted who completed the questionnaire (25/49, 51% RR); 5 responses from therapists not currently practicing in the United Kingdom were excluded.

Analysis of patient responses identified that 77.3% (167/216) of patients had acquired FNP as adults, 12.0% (26/216) acquired it at birth or during childhood, 7.9% (17/216) were carers of a child with the condition, and 2.8% (6/216) had another personal or professional connection. Patients with adult-acquired FNP had a mean of 6.96 (SD 7.00) years of experience since first being diagnosed. [Table 1](#) shows that the most common cause of their condition reported by patients was Bell palsy. Specialist therapists were mainly physiotherapists by training (22/25, 88%), with the remainder being speech and language therapists. Clinicians had a mean of 9.72 (SD 7.68) years of experience in treating FNP cases.

**Table 1.** Patients in the United Kingdom with adult-acquired facial nerve paralysis: reported diagnosis and treatment pathways (n=167).

Patient question	Response	Values, n (%)
<b>Cause of condition (164 responses)</b>		
	Bell Palsy	89 (54.3)
	Acoustic neuroma or vestibular schwannoma	27 (16.5)
	Ramsay Hunt syndrome	23 (14.0)
	Salivary gland or parotid tumor	4 (2.4)
	Facial nerve neuroma	3 (1.7)
	Birth trauma	1 (0.6)
	Lyme disease	1 (0.6)
	Stroke	1 (0.6)
	Other <sup>a</sup>	8 (4.9)
	Do not know	7 (4.3)
<b>Treatments provided to date (166 responses)</b>		
	Advice on eye care	111 (66.9)
	Prednisolone or other corticosteroids	100 (60.2)
	Antivirals	43 (25.9)
	Antibiotics	26 (15.7)
	Botox injections	71 (42.8)
	Facial neuromuscular retraining	101 (60.8)
	Electrical stimulation therapy	35 (21.1)
	Plastic surgery <sup>b</sup>	33 (19.9)
	Psychological therapy (eg, cognitive behavioral therapy)	16 (9.6)
	Other <sup>c</sup>	12 (7.2)
	No treatment <sup>d</sup>	6 (3.6)
<b>Stage at which first treatment started (164 responses)</b>		
	Within 72 hours following symptoms	109 (66.4)
	Within 1 month of onset	18 (11.0)
	1-6 months postonset	9 (5.5)
	6-9 months postonset	6 (3.7)
	>9 months postonset	11 (6.7)
	Do not know or other	11 (6.7)

<sup>a</sup>Other causes include the virus of the brain stem, postoperative complications, otitis media, skull fracture, side effect of radiotherapy, and accidental injury.

<sup>b</sup>Plastic surgery includes face lift, brow lift, eyelid surgery, and facial sling.

<sup>c</sup>Other treatments provided include acupuncture, self-funded chiropractic, and massage.

<sup>d</sup>No treatment group includes 2 patients diagnosed with acoustic neuroma or vestibular schwannoma and 4 with Bell palsy (1 assigned to the trial control group).

### Follow-Up Survey

The follow-up questionnaire was completed by 28 of 50 clinicians (RR 56%); 75% (21/28) were facial therapists, 11% (3/28) were hospital doctors, 8% (2/28) were neurological physiotherapists, and 4% (1/28) were clinical psychologists.

### Treatment Pathways

An analysis of treatment pathways was conducted for patients with adult-acquired FNP. Respondents reported receiving an average of 3.27 (SD 1.60) different treatments following initial diagnosis, most commonly corticosteroids, advice on eye care, and facial NMR, as shown in [Table 1](#).

### Diagnosis and Initial Treatment

Overall, 66.4% (109/164) of adult-acquired cases who reported the timing of their first treatment said this was within 72 hours of symptom onset (Table 1); for those most recently diagnosed ( $\leq 1$  year ago), this figure was 91% (31/34) versus 47% (32/68) for patients diagnosed 5 to 18 years ago. The average time to first review of their case was 64 (SD 26.8) days; for those diagnosed  $\leq 1$  year ago, this figure was an average of 6 (SD 28.8) days. Fewer than 9.6% (16/166) of patients had been referred for psychological therapy.

### Referral for NMR

Of 167 respondents with adult-acquired FNP, 98 (58.6%) had been referred for facial NMR, and these patients were treated

in 35 different centers. Table 2 shows that nearly half (44/98, 45%) were referred for NMR by a hospital consultant following other treatments; only 28% (27/98) were referred by their general practitioner (GP), and a further 14% (14/98) indicated that they initiated the referral themselves (usually via their GP) following information provided by friends, family, or patient support groups or based on their own research. Therapists reported that their specialist center received a mean of 73.2 (SD 75.5) new facial NMR referrals in an average year, with a large variation between centers (median 30, IQR 123). Table 3 indicates that the mean percentage of referrals from a GP is 37% (SD 32%), with hospital consultants accounting for between 11% (11/98) and 18% (18/98), depending on the specialty.

**Table 2.** Patients with adult-acquired facial nerve paralysis referred for facial neuromuscular retraining (n=98).

Patient question	Response	Values, n (%)
<b>Referral route to therapy (98 responses)</b>		
	General practitioner	27 (27.6)
	Self-initiated (usually via a general practitioner)	14 (14.3)
	Plastic surgeon	18 (18.4)
	Ear, nose, and throat specialist	15 (15.3)
	Neurologist	11 (11.2)
	Other <sup>a</sup>	11 (11.2)
	Do not know	2 (2.0)
<b>Any problems with referral (97 responses)</b>		
	Yes	22 (22.7)
	No	69 (71.1)
	Do not know	6 (6.2)
<b>Feedback provided during therapy<sup>b</sup> (96 responses)</b>		
	Yes	70 (72.9)
	No	11 (11.5)
	Do not know	15 (15.6)

<sup>a</sup>Referral routes—Other includes solicitor, speech and language therapist, and Botox consultant.

<sup>b</sup>Feedback tended to be given verbally, with the addition of photographic evidence, sharing of electromyography results, scores from the Sunnybrook Scale, or via percent recovered score.

**Table 3.** Current referral and treatment pathways reported by facial therapy specialists in the United Kingdom.

Referrals	Values
<b>Source of referral (24 responses) , mean percentage (SD)</b>	
General practitioner	37 (32)
Plastic surgeon	18 (25)
Ear, nose, and throat specialist	14 (15)
Neurologist	10 (11)
Eye specialist	7 (18)
Other <sup>a</sup>	14 (28)
<b>Treatments pathway, n (%)</b>	
<b>Wait for first appointment following referral (weeks; 25 responses)</b>	
<1	4 (16)
1-2	3 (12)
3-4	8 (32)
5-6	4 (16)
8-12	5 (20)
<b>Treatments patients receive before referral (25 responses)</b>	
Advice on eye care	14 (56)
Facial neuromuscular retraining	6 (24)
Prednisolone or other corticosteroids	20 (80)
Botox injections	12 (48)
Plastic surgery	11 (44)
Psychological therapy (eg, cognitive behavioral therapy)	4 (16)
Other <sup>b</sup>	6 (24)
<b>Follow-on referral to other specialists (25 responses)</b>	
Ophthalmology	18 (72)
Botox injections	19 (76)
Psychological therapy	16 (64)
Surgery (dynamic facial reanimation)	13 (52)
Other <sup>c</sup>	7 (28)
<b>Feedback provided to referring clinician (25 responses)</b>	
Feedback on progress and final outcome	13 (52)
Feedback on final outcome only	10 (40)
No feedback provided	2 (8)

<sup>a</sup>Other sources of referral: community pediatricians, neurosurgical or maxillofacial consultants, physiotherapists, speech and language therapists.

<sup>b</sup>Other prior treatments: blood tests, magnetic resonance imaging, electromyography, referral to peer support group, education, soft tissue mobilization, facial massage, and taping.

<sup>c</sup>Other follow-up referrals: radiology; nerve conduction studies; maxillofacial; ear, nose, and throat; speech and language therapy; restorative dentistry; vestibular physiotherapy; and audiology.

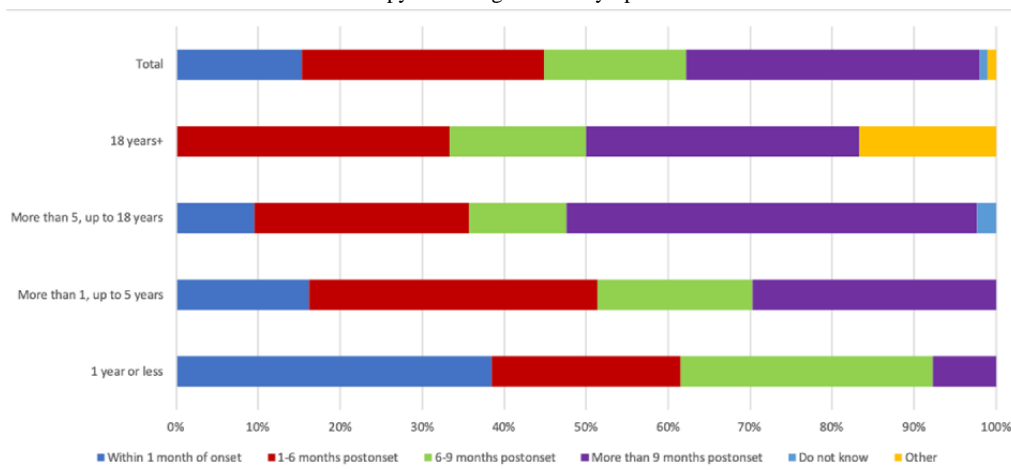
Figure 2 shows that the referral for NMR occurs at different points following the onset of symptoms. Overall, 15% (15/98) of adult cases were referred within 1 month; 38% (8/21) of people diagnosed most recently ( $\leq 1$  year ago), compared with 0% for those diagnosed  $>18$  years ago. Once referred, patients reported that they had to wait an average of 7.14 weeks (range 0.5-23 weeks) for a first appointment. Table 3 shows that

therapy centers currently record a mean wait time of 4.67 (SD 3.35) weeks. One in 4 adult patients (22/97, 23%) described experiencing difficulties with their referral, most commonly owing to problems accessing National Health Service (NHS) funding, a shortage of specialist therapists, or difficulties in persuading their GP to refer. Overall, 80% (20/25) of centers were aware that some patients experienced difficulties; the 2



main causes identified were limited awareness of facial NMR among GPs and a shortage of NHS funding.

**Figure 2.** Timing of referral for tailored facial exercise therapy following onset of symptoms.



### Other Treatments

Therapists reported that before referral for NMR, their patients usually received a number of other treatments, as detailed in Table 3, most commonly corticosteroids, advice on eye care, Botox injections, and plastic surgery. In addition, 24% (6/25) of centers reported receiving referrals following failed NMR, presumably provided by a nonspecialist. On completion of facial NMR, 64% (16/25) of centers regularly refer some patients for psychological therapy, 72% (18/25) to ophthalmology, and 52% (13/25) to surgery (dynamic facial reanimation). Although several specialties are involved in the treatment pathway, only 52% (13/25) of therapists had participated in a multidisciplinary team (MDT) review. Similarly, only 23.0% (38/165) of adult patients were aware of such a review of their case, and a further 24.8% (41/165) were uncertain.

### Views on Treatment

When asked to rate the importance of various treatments, based on their own experience, both patients and therapists rated the same 6 most highly, although median scores indicated little discrimination among them. Patients' scores produced the following ranking: 1st—advice on eye care, 2nd—facial NMR, 3rd—psychological therapy, 4th—corticosteroids, 5th—Botox injections, and 6th—plastic surgery. Therapists identified a similar order, although they ranked psychological therapy lower as 4th and surgery higher as 5th. All other treatments were rated as important by fewer than 4% of respondents.

### Adherence to NMR

Adult patients reported varying levels of adherence to their prescribed NMR program, with only 33% (32/97) recording very high levels, 41% (40/97) reporting medium-to-high adherence, 21% (20/97) recording poor-to-medium levels, and 5% (5/97) being uncertain about adherence. An analysis of patients' open-text comments identified 3 main barriers, regardless of the level achieved: "difficulties in fitting facial exercises into daily life," "using a mirror while completing exercises," and "insufficient regular follow-up." Patients who reported a very high level of adherence described 3 further facilitating factors: "observing improved outcomes in self,"

"belief that the treatment will work," and "observing positive outcomes in others." These patients also highlighted a common fear that acted as a spur: "scared of not recovering." An analysis of comments made by the medium-to-high adherence patient group uncovered 2 additional demotivating factors: an "inability to see any improvement" and "lack of funds to travel for check-ups." In terms of travel, distances can be considerable. Although most adult patients (75/97, 81%) were referred to a center within their own region, 9% (8/93) reported traveling  $\geq 115$  miles to reach a specialist center. Among patients with poor-to-medium adherence, 2 further barriers were identifiable: "pain associated with facial exercises" and "exercises can be tiring", possibly indicative of poor execution of a prescribed facial exercise regime.

When asked to describe their patients' level of compliance, 36% (9/25) reported very high levels, 56% (14/25) reported medium-to-high adherence, 4% (1/25) recorded poor-to-medium levels, and 4% (1/25) were uncertain about adherence. Interestingly, therapists appear to have underestimated the level of poor-to-medium adherence, perhaps suggesting overoptimistic feedback by patients. The thematic content analysis shows that therapists identify the same 3 barriers limiting adherence (ie, "fitting exercises into daily life," "lack of evidence of improvement," and "use of a mirror"). Unlike patients, therapists did not identify patients' travel costs. However, they did additionally think that the timing of other treatments influences adherence, for example, "poorer adherence post-surgery because this may be viewed as the primary treatment for managing their condition" and "Botox given too early may reduce compliance if patients rely on this to relieve feelings of tightness and synkinesis."

### Feedback on Recovery

Table 2 shows that 73% (70/96) of patients received regular feedback on their recovery during the course of treatment. This is provided in a number of ways with no consistent pattern nationally. Table 3 indicates that the regularity of feedback to referring physicians varies; 52% (12/25) of therapists provide feedback throughout treatment, but 40% (10/25) only report a final outcome. Final discharge summaries also vary, some

include validated measures such as the Sunnybrook Scale and others provide photographs showing the patient's progress.

In response to the follow-up survey, clinicians reported the experience of using various methods for recording treatment outcomes ranging from photographic evidence to recognized disease-specific scales. When asked to name scales they had used, 17 different instruments were identified (Table 4). The

most frequently cited was the Sunnybrook Scale (26/28, 93% of respondents) [46], followed by the Facial Disability Index (FDI) identified by 54% (15/28) of respondents [29], and the House-Brackmann (HB) Scale identified by 50% (14/28) of clinicians [47]. Relatively few respondents were able to define a clinically significant change for these measures: 27% (7/26) for Sunnybrook, 7% (1/15) for FDI, and 21% (3/14) for HB. Where answers were provided, there was no consensus.

**Table 4.** Follow-up survey: outcome measures and minimum clinically significant change (n=28).

Outcome measure	Experience of use, n (%)	Minimum clinically significant change
Sunnybrook Scale	26 (93)	<ul style="list-style-type: none"> <li>Change by 10 points (4 respondents)</li> <li>Change by <math>\geq 5</math> points (1 respondent)</li> <li>5:1 change (2 respondents)</li> <li>No comment (19 respondents)</li> </ul>
Facial Disability Index	15 (54)	<ul style="list-style-type: none"> <li>Change by <math>\geq 5</math> points (1 respondent)</li> <li>No comment (14 respondents)</li> </ul>
House-Brackmann Scale	14 (50)	<ul style="list-style-type: none"> <li>Change by 1 grade (2 respondents)</li> <li>Grade 3 eye closure or above (1 respondent)</li> <li>Not sensitive enough for therapy outcomes (2 respondents)</li> <li>No comment (9 respondents)</li> </ul>
FaCE (Facial Clinimetric Evaluation) Scale	10 (36)	<ul style="list-style-type: none"> <li>Change by 5 points (1 respondent)</li> <li>Change by 10 points (1 respondent)</li> <li>Change by 15 points (1 respondent)</li> <li>No comment (7 respondents)</li> </ul>
Synkinesis Assessment Questionnaire	7 (25)	<ul style="list-style-type: none"> <li>Change by 1 point (1 respondent)</li> <li>Change by 10 points (1 respondent)</li> <li>Change by 15 points (1 respondent)</li> <li>Other<sup>a</sup> (1 respondent)</li> <li>No comment (3 respondents)</li> </ul>
SF-36 (36-item Short Form Health Survey)	4 (14)	<ul style="list-style-type: none"> <li>Change by 10 points (1 respondent)</li> <li>No comment (3 respondents)</li> </ul>
EuroQol-5D	4 (14)	<ul style="list-style-type: none"> <li>No comment (4 respondents)</li> </ul>
Face-Q <sup>b</sup> ; MEEI facegram <sup>c</sup> ; CORE-10 <sup>d</sup> ; MBLF <sup>e</sup> ; Lazarini <sup>f</sup>	2 (7)	<ul style="list-style-type: none"> <li>No comment (2 respondents)</li> </ul>
eFACE (electronic Facial Paralysis Assessment)	2 (7)	<ul style="list-style-type: none"> <li>Change by 10 points (1 respondent)</li> <li>No comment (1 respondent)</li> </ul>
Smile Index; CCE angle <sup>g</sup> ; Satisfaction with Appearance; Hospital Anxiety and Depression Scale	1 (4)	<ul style="list-style-type: none"> <li>No comment (1 respondent)</li> </ul>

<sup>a</sup>After the treatment score was compared with the opinion of a therapist for training or education purposes.

<sup>b</sup>Developed for facial aesthetic patients, enables users to tailor a version to suit their needs based on over 40 scales measuring a range of concepts important to patients.

<sup>c</sup>FACE-Gram software (MEEI, Boston, Mass).

<sup>d</sup>Comprises 10 items drawn from CORE-OM which is used in evaluation of counselling and psychological therapies in the UK.

<sup>e</sup>French oro-facial myofunctional assessment to quantify impairment and specify motor and functional deficit.

<sup>f</sup>Graphic-visual adaptation of House-Brackmann facial nerve grading for peripheral facial palsy

<sup>g</sup>Angle between the cheilion, contralateral cheilion, and ipsilateral endocanthion.

## Digital Technology

A thematic analysis of open-text comments on the potential value of digital technology identified 4 superordinate themes, as shown in Textbox 1. The first theme, *System*, was

predominantly voiced by specialist therapists who highlighted the potential for this technology to help reduce pressure on their time, improve their ability to monitor a patient's progress, increase coverage, and reduce travel for outpatient reviews. An extra subtheme, expressed by patients within the *System* theme,

was that new technology could help raise medical awareness, especially in primary care. A second theme, *Self-management*, was identified by both groups. This focused on improving treatment adherence and emphasized factors such as enabling people to fit facial exercises into their daily routines and providing motivational feedback through regular review of daily performance.

The third theme, *Identity*, emerged mainly from patients' responses. This highlighted psychological benefits such as improved confidence in people who are frightened to draw attention to themselves and hope in individuals that their

condition might improve. The final theme, *Innovation*, encompassed views, such as, that innovation is natural and that it can bring positive societal benefits, together with an awareness that funding constraints limit implementation of innovations. Patients demonstrated a great willingness to support future introduction of digital technology, but they also emphasized the need to raise awareness among GPs alongside implementation. Therapists expressed a similarly positive view of telerehabilitation, with a clear consensus that digital technology could not replace the initial face-to-face consultation because patients need to be carefully trained to be able to complete their exercises properly.

**Textbox 1.** Themes and subthemes: representative quotes from questionnaires.

## System:

- Therapist time
  - “It is a good adjunct to one-to-one therapy for most patients.” (Therapist)
  - “I think that for our patient group and geography, Telerehab would work well... we have no dedicated time for facial therapy. It’s included as an acute treatment” (Therapist)
- Monitoring
  - “To be able to monitor a patient's progress remotely will be of great value. It also has the potential to generate and collate a lot of objective data on facial function which will be invaluable in efficient data collection for future research.” (Therapist)
  - “Any technology in the form of an app and use of EMG biofeedback would be massively useful. This would offer real time feedback about how facial muscles are performing and how successful they are in performing their exercises.” (Therapist)
- Access
  - “Our community colleagues I think would benefit greatly and more the patients as many travel from long distances to see the team here as rare specialism.” (Therapist)
  - “I am always interested in new advances in medicine. I think if it can provide cover for patients in areas where there is no specialism [in facial neuromuscular], then it will be invaluable.” (Therapist)
  - “These are essential for the future of intervention with this patient group especially as there are few specialists and patients are currently having to travel long distances.” (Therapist)
- Medical awareness
  - “It [technology] needs to be embraced, further developed and used to educate the medical profession about the condition, its treatments, causes and correct treatment regime.” (Patient)
  - “New technology can be useful, but primary diagnosis is important - this [facial palsy] needs more awareness.” (Patient)

## Self-management:

- Exercise performance
  - “It would benefit some patients and encourage those patients who find it hard to fit exercises into their daily routine.” (Therapist)
  - “Using the 'Fitbit' technology whereby users are given rewards for exercising, can check on their daily performance and review weekly summaries of how well they have done. Pop up reminders to do exercises may also [be] very useful.” (Therapist)
- Motivation
  - “May be very useful to provide motivational feedback.” (Therapist)
  - “Technology could give users rewards for exercising, check on their daily performance and review weekly summaries of how well they have done.” (Therapist)

## Identity:

- Confidence
  - “Anything that can help the physical issues as well as facial symmetry/making someone feel better about themselves/more confident is so important.” (Patient)
  - “If facial palsy patients can be helped by the use of new technologies; they should be. They are hardly likely to fight for treatment as we are traumatised, embarrassed and scared about bringing attention to themselves.” (Patient)
- Hope
  - “Personally, I would try whatever technology was available!” (Patient)
  - “I am open to trying absolutely anything that could help my symptoms.” (Patient)
  - “I would try any new technology in the hope my palsy is improved. Living with it can make you have low self-esteem.” (Patient)

## Innovation:

- Natural
  - “I think it’s great, it’s good to know there is always research into new ways to aid recovery.” (Patient)

- “I think novel ways of assessing and treating patients using technology must be embraced.” (Therapist)
- “I am all for new technology if it proves helpful to the patient.” (Patient)
- Societal benefit
  - “This is exactly what we need! We need to move forward with the times. I feel many new treatment may be expensive initially but if it can change a person life then it is worth it.” (Patient)
  - “In the long term it may work out cheaper ... [the patient] may have their confidence back and be able to work and contribute back to society.” (Patient)
- Funding
  - “This is a very interesting area however costing and funding within NHS would be a concern for clinical use.” (Therapist)
  - “I would like to see it made available on the NHS.” (Patient)
  - “Financial pressures on health services, the relative rarity of facial therapists, the increasing familiarity / dependence of many people with / on evolving technologies all make it an important part of facial palsy interventions.” (Therapist)

## Discussion

### Principal Findings

This exploration of the experiences of patients with facial palsy and their clinicians provides new information on UK treatment pathways, access to facial neuromuscular retraining, the methods in use to monitor recovery, and patients' and clinicians' views on the introduction of an emerging digital technology to support facial NMR. The findings highlight the need to understand patients' and clinicians' experiences before introducing such a technology. As funding barriers currently limit access to routine NMR, evidence of cost-effectiveness may be needed before implementation.

### Treatment Pathways

Although variations were observed in the treatment pathways experienced by patients, there are signs that more recent cases (diagnosed  $\leq 1$  year ago) received treatment sooner than cases diagnosed  $>5$  years ago. For the only medical treatment supported by Cochrane systematic reviews (rapid initial prednisolone treatment [19]), we observed high adherence by GPs, especially for patients diagnosed most recently (91%, 31/34 treated). For facial NMR therapy (partially supported by Cochrane review evidence [33]), this study found a move toward earlier NMR referrals in more recent cases. Even so, 1 in 4 patients experienced difficulties with their referral, commonly citing poor GP awareness. The fact that guidelines currently contain no firm guidance on referrals for NMR may offer some explanation [37].

In the United Kingdom, NICE advises that uncomplicated FNP cases can be managed by primary care to include referral to various hospital medical specialists and therapists [37]. This is challenging because an average GP will only see one patient every 2 years [48]. An earlier UK study identified an overall reduction in GP referral rates of patients with FNP to hospital medical specialists over the period 2001 to 2012 [4]. In this study, twice as many referrals to specialist facial therapists were initiated by hospital consultants compared with GP referrals, with patients often prompting the GP referrals. The NICE guidance for GPs does not specifically mention NMR referral,

instead indicating a need for confirmatory trial evidence as suggested by the Cochrane review [33]. An update to this review may help inform future guidance on referrals [34]. Finally, although patients in our study ranked psychological therapy third highest in terms of its importance, fewer than 10% had been referred; this is presumably linked to the fact that NICE guidelines only contain a weak recommendation for GPs to consider referral for counseling [37].

Once referred for NMR, access to an appropriate therapist can be problematic. Several patients reported traveling very long distances to reach a specialist center; the cost of travel for regular checkups was also identified as a barrier to adherence by patients. Specialist centers reported receiving patients previously referred to a nonspecialist therapist, reinforcing the need for increased access to appropriately trained professionals [49]. Patient adherence to prescribed facial exercises will influence treatment effectiveness. Both patients and therapists identified the same barriers: fitting exercises into daily life, the use of a mirror, and the need for regular feedback. The fact that clinicians perceive higher levels of adherence than those reported by patients indicates a need for improved monitoring of adherence. Other research has shown that a collaboration between specialists can reduce the burden of long-term disability for acute onset FNP [50]. In this study, although patients experienced care pathways that involved referral to several medical specialties, such collaboration was limited; just over half (52%) of the therapists had any experience of participating in an MDT review, and only 23% (38/165) of adult patients were aware of such a review of their case. Interestingly, reports are now emerging of efforts to integrate physical therapy with treatment by ophthalmologists, oculoplastic surgeons, and ENT and other specialists [51]. It is also considered that MDTs are likely to play an important role in standardizing outcome measures and implementing relevant data collection [52].

Evidence of effectiveness of FNP treatments currently remains reliant on subjective measures, including reduction in *crocodile tears*, *incomplete recovery of motor function*, and *cosmetically disabling sequelae* [19,33]. This study identified inconsistency in the methods used in the United Kingdom to report treatment outcomes. In addition to photographs, various validated scales



are used. Among these, the Sunnybrook Scale, mentioned by 96% of specialists, is considered to grade patients in a more objective and continuous manner than the HB Scale, which was mentioned by half [53]. However, FDI, mentioned by 54% of clinicians, better represents impairment, disability, and psychosocial status than the 36-item Short Form Survey (SF-36) health status measure, mentioned by 14% [54]. In addition to variation in the scales used to record treatment outcomes, there was no consensus on what represents a clinically significant change in any scale.

Interestingly, very few clinicians (14%) reported the experience of using EuroQol-5D, the health-related quality of life (HRQoL) measure used for NHS reimbursement decisions [55]. Converted into incremental quality-adjusted life years (QALYs), this is used to quantify long-term treatment outcomes [56]. QALYs are particularly relevant as 30% of patients with FNP will continue to live with reduced HRQoL over the rest of their lives [3,13-15]. In addition, because such individuals may give up their original employment [57], this can lead to a significant long-term societal cost burden, especially if the condition was acquired in early life [58]. An international collaboration has recently been established, focusing on pediatric patients with FNP (using a patient-centered approach, similar to this study) with the aim of comparing FNP treatment pathways, standardizing outcome measurement, and developing value-based reimbursement strategies [59].

### Digital Technology

To our knowledge, this is the first UK study to explore the potential use of wearable digital technology to support facial NMR therapy. The findings show that patients and therapists both demonstrate a positive attitude toward the introduction of such technology, with patients recognizing benefits centered on better self-management and improved confidence, therapists identifying better monitoring of patients' progress and reduced work pressures, and both highlighting the potential for improved adherence to facial exercise programs. The main barriers to adherence could all be addressed by an appropriate real time (synchronous) digital solution that addresses patients' and clinicians' feedback. A review of telerehabilitation articles has recently concluded that patients' feedback will help improve future areas of applications, although no FNP telerehabilitation studies were identified [60]. A review of real time, web-based consultation has highlighted a number of general barriers and facilitators [61]. The key barriers and facilitators mirror those found in this study. Asynchronous methods for monitoring FNP treatment outcomes have also recently been evaluated by 2 research teams. Tan et al [62] reported that the assessment of videos using the Sunnybrook and HB scales is as good as a face-to-face assessment, although the lack of real time

interaction was judged to limit the value of this approach. Mothes et al [63] found that automatic Sunnybrook grading of photographs using machine learning can deliver fair agreement compared with the subjective rating of the same photographs. Neither study addressed real time monitoring and biofeedback.

### Limitations

A number of limitations should be borne in mind when considering this research. First, patient respondents may not be representative of the wider population because participants were recruited via a specialist support group. Second, there may be recall inaccuracy when participants are asked to provide information sometime after the event. Third, the therapist RR (51% national survey and 56% follow-up survey) means that data may not fully reflect the national picture. Finally, although the greatest care was taken in the questionnaire design, as with all surveys that record individuals' views, the validity and reliability of the data could not be tested independently.

### Conclusions and Implications for Technology

#### Introduction

To date, little research has explored the potential value of digital technology in assisting facial NMR therapy. This study provides a baseline overview of FNP treatment pathways in the United Kingdom, the factors limiting access to NMR and influencing therapy adherence, the main methods used to record treatment outcome, and the potential role of digital technology. The study indicates that harmonization of outcome measures is required to both strengthen the evidence on treatment effectiveness and to better support MDT management. The main factors limiting NMR adherence could all be addressed through the use of real time digital technology. However, for the type of wearable technology being considered, product design will be an additional factor likely to influence adherence [64], especially in the younger 30- to 45-year age group affected by this condition [3]. However, although the study clearly demonstrates positive attitudes toward the introduction of digital technology, economic barriers may prove to be a challenge. Previous research has identified funding as a barrier for access to surgical treatments for FNP in England [65] and more recently in Wales, Scotland, and Northern Ireland [66]. This study has similarly identified funding barriers to NMR referral. Finally, organizational and cultural factors are acknowledged to act as important barriers to implementation for all digital health innovations [67], often reinforced by policy priorities [68]. One legacy of the COVID-19 pandemic is that health systems worldwide are rapidly adopting digital options in many clinical areas [69]. Thus, barriers to the introduction of digital technology to assist facial NMR therapy may now be lower, especially if cost-effectiveness can be demonstrated.

### Acknowledgments

The authors would like to thank all the patients and clinicians who took the time to complete detailed questionnaires describing their experiences.

This study is funded as part of the FRAME program, an inconspicuous, noninvasive, mobile sensor device for real time control of assistive technologies through facial expression, which is supported by the NIHR (reference number II-LA-0814-20008). The

research funder had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the paper for publication.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

National Survey Questionnaires.

[DOCX File, 219 KB - [jmir\\_v22i10e20406\\_app1.docx](#)]

## References

- Hernández RA, Sullivan F, Donnan P, Swan I, Vale L, BELLS Trial Group. Economic evaluation of early administration of prednisolone and/or aciclovir for the treatment of Bell's palsy. *Fam Pract* 2009 Apr;26(2):137-144. [doi: [10.1093/fampra/cmn107](#)] [Medline: [19244470](#)]
- Baugh RF, Basura GJ, Ishii L, Schwartz SR. Clinical practice guideline: Bell's Palsy. *Otolaryngol Head Neck Surg* 2013 Aug 23;149(2\_suppl):P13-P14. [doi: [10.1177/0194599813493390a16](#)] [Medline: [24189771](#)]
- Peitersen E. Bell's palsy: the spontaneous course of 2,500 peripheral facial nerve palsies of different etiologies. *Acta Otolaryngol Suppl* 2002(549):4-30. [Medline: [12482166](#)]
- Morales DR, Donnan PT, Daly F, Staa TV, Sullivan FM. Impact of clinical trial findings on Bell's palsy management in general practice in the UK 2001-2012: interrupted time series regression analysis. *BMJ Open* 2013;3(7):e003121 [FREE Full text] [doi: [10.1136/bmjopen-2013-003121](#)] [Medline: [23864211](#)]
- Murthy JM, Saxena A. Bell's palsy: treatment guidelines. *Ann Indian Acad Neurol* 2011 Jul;14(Suppl 1):S70-S72 [FREE Full text] [doi: [10.4103/0972-2327.83092](#)] [Medline: [21847333](#)]
- Holland NJ, Weiner GM. Recent developments in Bell's palsy. *Br Med J* 2004 Sep 4;329(7465):553-557 [FREE Full text] [doi: [10.1136/bmj.329.7465.553](#)] [Medline: [15345630](#)]
- Hauser WA, Karnes WE, Annis J, Kurland LT. Incidence and prognosis of Bell's palsy in the population of Rochester, Minnesota. *Mayo Clin Proc* 1971 Apr;46(4):258-264. [Medline: [5573820](#)]
- Phan NT, Panizza B, Wallwork B. A general practice approach to Bell's palsy. *Aust Fam Physician* 2016 Nov;45(11):794-797 [FREE Full text] [Medline: [27806447](#)]
- Adour KK, Bell DN, Hilsinger RL. Herpes simplex virus in idiopathic facial paralysis (Bell palsy). *J Am Med Assoc* 1975 Aug 11;233(6):527-530. [Medline: [167209](#)]
- Alshami S, Hodge D, Bradley E. The epidemiology and surgical outcomes of facial nerve palsy in a population-based cohort. *Investig Ophthalmol Vis Sci* 2017;58(8):1 [FREE Full text] [doi: [10.1007/978-3-540-69000-9\\_100651](#)]
- Engström M, Berg T, Stjernquist-Desatnik A, Axelsson S, Pitkäranta A, Hultcrantz M, et al. Prednisolone and valaciclovir in Bell's palsy: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet Neurol* 2008 Nov;7(11):993-1000. [doi: [10.1016/S1474-4422\(08\)70221-7](#)] [Medline: [18849193](#)]
- Peitersen E. Natural history of Bell's palsy. *Acta Otolaryngol Suppl* 1992;492:122-124. [doi: [10.3109/00016489209136829](#)] [Medline: [1632236](#)]
- Neely JG, Neufeld PS. Defining functional limitation, disability, and societal limitations in patients with facial paresis: initial pilot questionnaire. *Am J Otol* 1996 Mar;17(2):340-342. [Medline: [8723973](#)]
- Weir A, Pentland B, Crosswaite A, Murray J, Mountain R. Bell's palsy: the effect on self-image, mood state and social activity. *Clin Rehabil* 2016 Jul;9(2):121-125 [FREE Full text] [doi: [10.1177/026921559500900206](#)]
- Fu L, Bundy C, Sadiq SA. Psychological distress in people with disfigurement from facial palsy. *Eye (Lond)* 2011 Oct;25(10):1322-1326 [FREE Full text] [doi: [10.1038/eye.2011.158](#)] [Medline: [21720412](#)]
- Valente SM. Visual disfigurement and depression. *Plast Surg Nurs* 2004;24(4):140-6; quiz 147. [doi: [10.1097/00006527-200410000-00003](#)] [Medline: [15632721](#)]
- Miller SG. Angelina Jolie's Facial Paralysis: What Is Bell's Palsy? Live Science. 2017. URL: <https://www.livescience.com/59949-angelina-jolie-bells-palsy.html> [accessed 2020-03-20]
- Gagyor I, Madhok VB, Daly F, Sullivan F. Antiviral treatment for Bell's palsy (idiopathic facial paralysis). *Cochrane Database Syst Rev* 2019 Sep 5;9:CD001869. [doi: [10.1002/14651858.CD001869.pub9](#)] [Medline: [31486071](#)]
- Madhok VB, Gagyor I, Daly F, Somasundara D, Sullivan M, Gammie F, et al. Corticosteroids for Bell's palsy (idiopathic facial paralysis). *Cochrane Database Syst Rev* 2016 Jul 18;7:CD001942 [FREE Full text] [doi: [10.1002/14651858.CD001942.pub5](#)] [Medline: [27428352](#)]
- Linder TE, Abdelkafy W, Caverio-Vanek S. The management of peripheral facial nerve palsy: 'paresis' versus 'paralysis' and sources of ambiguity in study designs. *Otol Neurotol* 2010 Feb;31(2):319-327. [doi: [10.1097/MAO.0b013e3181cabd90](#)] [Medline: [20009779](#)]
- Miwa H, Kondo T, Mizuno Y. Bell's palsy-induced blepharospasm. *J Neurol* 2002 Apr;249(4):452-454. [doi: [10.1007/s004150200038](#)] [Medline: [11967652](#)]

22. Sullivan FM, Swan IR, Donnan PT, Morrison JM, Smith BH, McKinstry B, et al. Early treatment with prednisolone or acyclovir in Bell's palsy. *N Engl J Med* 2007 Oct 18;357(16):1598-1607. [doi: [10.1056/NEJMoa072006](https://doi.org/10.1056/NEJMoa072006)] [Medline: [17942873](https://pubmed.ncbi.nlm.nih.gov/17942873/)]
23. Lyons C, McNab A. Symptomatic nocturnal lagophthalmos. *Aust N Z J Ophthalmol* 1990 Nov;18(4):393-396. [doi: [10.1111/j.1442-9071.1990.tb01823.x](https://doi.org/10.1111/j.1442-9071.1990.tb01823.x)] [Medline: [2076287](https://pubmed.ncbi.nlm.nih.gov/2076287/)]
24. Hayashi A, Maruyama Y, Okada E, Ogino A. Use of a suture anchor for correction of ectropion in facial paralysis. *Plast Reconstr Surg* 2005 Jan;115(1):234-239. [Medline: [15622256](https://pubmed.ncbi.nlm.nih.gov/15622256/)]
25. Ramakrishnan Y, Alam S, Kotecha A, Gillett D, D'Souza A. Reanimation following facial palsy: present and future directions. *J Laryngol Otol* 2010 Nov;124(11):1146-1152. [doi: [10.1017/S0022215110001507](https://doi.org/10.1017/S0022215110001507)] [Medline: [20546647](https://pubmed.ncbi.nlm.nih.gov/20546647/)]
26. McAllister K, Walker D, Donnan P, Swan I. Surgical interventions for the early management of Bell's palsy. *Cochrane Database Syst Rev* 2013 Oct 16(10):CD007468. [doi: [10.1002/14651858.CD007468.pub3](https://doi.org/10.1002/14651858.CD007468.pub3)] [Medline: [24132718](https://pubmed.ncbi.nlm.nih.gov/24132718/)]
27. Zhang X, Feng L, Du L, Zhang A, Tang T. Literature study on clinical treatment of facial paralysis in the last 20 years using web of science: comparison between rehabilitation, physiotherapy and acupuncture. *Neural Regen Res* 2012 Jan 15;7(2):152-159 [FREE Full text] [doi: [10.3969/j.issn.1673-5374.2012.02.013](https://doi.org/10.3969/j.issn.1673-5374.2012.02.013)] [Medline: [25767492](https://pubmed.ncbi.nlm.nih.gov/25767492/)]
28. Beurskens CH, Heymans PG. Mime therapy improves facial symmetry in people with long-term facial nerve paresis: a randomised controlled trial. *Aust J Physiother* 2006;52(3):177-183 [FREE Full text] [doi: [10.1016/s0004-9514\(06\)70026-5](https://doi.org/10.1016/s0004-9514(06)70026-5)] [Medline: [16942452](https://pubmed.ncbi.nlm.nih.gov/16942452/)]
29. Brach J, VanSwearingen J. Physical therapy for facial paralysis: a tailored treatment approach. *Phys Ther* 1999;79(4):- [doi: [10.1093/ptj/79.4.397](https://doi.org/10.1093/ptj/79.4.397)] [Medline: [10201545](https://pubmed.ncbi.nlm.nih.gov/10201545/)]
30. Segal B, Zompa I, Danys I, Black M, Shapiro M, Melmed C, et al. Symmetry and synkinesis during rehabilitation of unilateral facial paralysis. *J Otolaryngol* 1995 Jun;24(3):143-148. [Medline: [7674438](https://pubmed.ncbi.nlm.nih.gov/7674438/)]
31. Ross B, Nedzelski JM, McLean JA. Efficacy of feedback training in long-standing facial nerve paresis. *Laryngoscope* 1991 Jul;101(7 Pt 1):744-750. [doi: [10.1288/00005537-199107000-00009](https://doi.org/10.1288/00005537-199107000-00009)] [Medline: [2062155](https://pubmed.ncbi.nlm.nih.gov/2062155/)]
32. Lindsay R, Robinson M, Hadlock T. Comprehensive facial rehabilitation improves function in people with facial paralysis: a 5-year experience at the Massachusetts eye and ear infirmary. *Phys Ther* 2010 Mar;90(3):391-397. [doi: [10.2522/ptj.20090176](https://doi.org/10.2522/ptj.20090176)] [Medline: [20093325](https://pubmed.ncbi.nlm.nih.gov/20093325/)]
33. Teixeira L, Valbuza J, Prado G. Physical therapy for Bell's palsy (idiopathic facial paralysis). *Cochrane Database Syst Rev* 2011 Dec 7(12):CD006283. [doi: [10.1002/14651858.CD006283.pub3](https://doi.org/10.1002/14651858.CD006283.pub3)] [Medline: [22161401](https://pubmed.ncbi.nlm.nih.gov/22161401/)]
34. Khan A, Szczepura A, Nduka C. The Effectiveness of Facial Exercises in the Treatment of Bell's Palsy (Sudden Weakness of Facial Muscles): a Systematic Review. *CRD Database - University of York*. 2017. URL: [http://www.crd.york.ac.uk/PROSPERO/display\\_record.php?ID=CRD42017073067](http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017073067) [accessed 2020-05-20]
35. Ferreira M, Marques EE, Duarte JA, Santos PC. Physical therapy with drug treatment in Bell palsy: a focused review. *Am J Phys Med Rehabil* 2015 Apr;94(4):331-340. [doi: [10.1097/PHM.0000000000000255](https://doi.org/10.1097/PHM.0000000000000255)] [Medline: [25785922](https://pubmed.ncbi.nlm.nih.gov/25785922/)]
36. de Almeida JR, Guyatt GH, Sud S, Dorion J, Hill MD, Kolber MR, Bell Palsy Working Group, Canadian Society of Otolaryngology - Head and Neck Surgery and Canadian Neurological Sciences Federation. Management of Bell palsy: clinical practice guideline. *Can Med Assoc J* 2014 Sep 2;186(12):917-922 [FREE Full text] [doi: [10.1503/cmaj.131801](https://doi.org/10.1503/cmaj.131801)] [Medline: [24934895](https://pubmed.ncbi.nlm.nih.gov/24934895/)]
37. Suspected Neurological Conditions: Recognition and Referral. National Institute for Health and Care Excellence (NICE). 2019. URL: <https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf> [accessed 2020-02-19]
38. Cader Z, Kings J, Kipps C. Transforming Community Neurology. Neurological Alliance. 2020. URL: <http://tinyurl.com/CNPReport> [accessed 2020-02-19]
39. Castle-Clarke S, Imison C. The Digital Patient: Transforming Primary Care?. London, UK: New Cavendish Street London Nuffield Trust; 2016.
40. Szczepura A, Khan A, Holliday N, Nduka C, Neville C, Johnson K, et al. PP26 facial palsy therapy: can novel 'smart spectacles' help people smile? *Int J Technol Assess Health Care* 2019 Jan 3;34(S1):76 [FREE Full text] [doi: [10.1017/s0266462318001964](https://doi.org/10.1017/s0266462318001964)]
41. Online Surveys. 2018. URL: <https://www.onlinesurveys.ac.uk/> [accessed 2020-08-14]
42. Inform, Support, Research. Facial Palsy UK. URL: <https://www.facialpalsy.org.uk> [accessed 2020-05-26]
43. Facial Therapy Specialists UK. URL: <https://fts-uk.org> [accessed 2020-05-26]
44. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med* 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: [10.1097/ACM.0000000000000388](https://doi.org/10.1097/ACM.0000000000000388)] [Medline: [24979285](https://pubmed.ncbi.nlm.nih.gov/24979285/)]
45. Saldaña J. The Coding Manual for Qualitative Researchers. Second Edition. London, UK: Sage Publications; 2013.
46. Hu W, Ross B, Nedzelski J. Reliability of the Sunnybrook facial grading system by novice users. *J Otolaryngol* 2001 Aug;30(4):208-211. [doi: [10.2310/7070.2001.20148](https://doi.org/10.2310/7070.2001.20148)] [Medline: [11771031](https://pubmed.ncbi.nlm.nih.gov/11771031/)]
47. House JW, Brackmann DE. Facial nerve grading system. *Otolaryngol Head Neck Surg* 1985 Apr;93(2):146-147. [doi: [10.1177/019459988509300202](https://doi.org/10.1177/019459988509300202)] [Medline: [3921901](https://pubmed.ncbi.nlm.nih.gov/3921901/)]
48. Glass GE, Tzafetta K. Optimising treatment of Bell's Palsy in primary care: the need for early appropriate referral. *Br J Gen Pract* 2014 Dec;64(629):e807-e809 [FREE Full text] [doi: [10.3399/bjgp14X683041](https://doi.org/10.3399/bjgp14X683041)] [Medline: [25452547](https://pubmed.ncbi.nlm.nih.gov/25452547/)]

49. Cronin G, Steenerson RL. The effectiveness of neuromuscular facial retraining combined with electromyography in facial paralysis rehabilitation. *Otolaryngol Head Neck Surg* 2003 Apr;128(4):534-538. [doi: [10.1016/s0194-5998\(03\)00005-6](https://doi.org/10.1016/s0194-5998(03)00005-6)] [Medline: [12707657](#)]
50. Eviston TJ, Croxson GR, Kennedy PG, Hadlock T, Krishnan AV. Bell's palsy: aetiology, clinical features and multidisciplinary care. *J Neurol Neurosurg Psychiatry* 2015 Dec;86(12):1356-1361. [doi: [10.1136/jnnp-2014-309563](https://doi.org/10.1136/jnnp-2014-309563)] [Medline: [25857657](#)]
51. van Landingham SW, Diels J, Lucarelli M. Physical therapy for facial nerve palsy: applications for the physician. *Curr Opin Ophthalmol* 2018 Sep;29(5):469-475. [doi: [10.1097/ICU.0000000000000503](https://doi.org/10.1097/ICU.0000000000000503)] [Medline: [29994853](#)]
52. Butler DP, Grobbelaar A. Facial palsy: what can the multidisciplinary team do? *J Multidiscip Healthc* 2017;10:377-381 [FREE Full text] [doi: [10.2147/JMDH.S125574](https://doi.org/10.2147/JMDH.S125574)] [Medline: [29026314](#)]
53. Berg T, Jonsson L, Engström M. Agreement between the Sunnybrook, House-Brackmann, and Yanagihara facial nerve grading systems in Bell's palsy. *Otol Neurotol* 2004 Nov;25(6):1020-1026. [doi: [10.1097/00129492-200411000-00027](https://doi.org/10.1097/00129492-200411000-00027)] [Medline: [15547437](#)]
54. VanSwearingen J, Brach J. The facial disability index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. *Phys Ther* 1996 Dec;76(12):1288-98; discussion 1298. [doi: [10.1093/ptj/76.12.1288](https://doi.org/10.1093/ptj/76.12.1288)] [Medline: [8959998](#)]
55. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. Third Edition. Oxford, UK: Oxford University Press; 2005.
56. Ogden J. QALYs and their role in the NICE decision-making process. *Prescriber* 2017 May 2;28(4):41-43 [FREE Full text] [doi: [10.1002/psb.1562](https://doi.org/10.1002/psb.1562)]
57. Nishikura M. Difficulties With Employment: Understanding the Experiences of People With Visible Differences. READ Discussion Paper. 2009. URL: <http://www.rease.e.u-tokyo.ac.jp/read/en/archive/dp/f09/f0906.pdf> [accessed 2020-09-14]
58. Weatherly H, Drummond M, Claxton K, Cookson R, Ferguson B, Godfrey C, et al. Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. *Health Policy* 2009 Dec;93(2-3):85-92. [doi: [10.1016/j.healthpol.2009.07.012](https://doi.org/10.1016/j.healthpol.2009.07.012)] [Medline: [19709773](#)]
59. Butler DP, de la Torre A, Borschel GH, Hadlock TA, Beurskens C, Bogart K, et al. An international collaborative standardizing patient-centered outcome measures in pediatric facial palsy. *JAMA Facial Plast Surg* 2019 Sep 1;21(5):351-358 [FREE Full text] [doi: [10.1001/jamafacial.2019.0224](https://doi.org/10.1001/jamafacial.2019.0224)] [Medline: [31070677](#)]
60. Peretti A, Amenta F, Tayebati SK, Nittari G, Mahdi SS. Telerehabilitation: review of the state-of-the-art and areas of application. *JMIR Rehabil Assist Technol* 2017 Jul 21;4(2):e7 [FREE Full text] [doi: [10.2196/rehab.7511](https://doi.org/10.2196/rehab.7511)] [Medline: [28733271](#)]
61. Almathami HK, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. *J Med Internet Res* 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: [10.2196/16407](https://doi.org/10.2196/16407)] [Medline: [32130131](#)]
62. Tan JR, Coulson S, Keep M. Face-to-face versus video assessment of facial paralysis: implications for telemedicine. *J Med Internet Res* 2019 Apr 12;21(4):e11109 [FREE Full text] [doi: [10.2196/11109](https://doi.org/10.2196/11109)] [Medline: [30977734](#)]
63. Mothes O, Modersohn L, Volk GF, Klingner C, Witte OW, Schlattmann P, et al. Automated objective and marker-free facial grading using photographs of patients with facial palsy. *Eur Arch Otorhinolaryngol* 2019 Dec;276(12):3335-3343. [doi: [10.1007/s00405-019-05647-7](https://doi.org/10.1007/s00405-019-05647-7)] [Medline: [31535292](#)]
64. Ward G, Fielden S, Muir H, Holliday N, Urwin G. Developing the assistive technology consumer market for people aged 50–70. *Ageing Soc* 2016 Feb 22;37(5):1050-1067. [doi: [10.1017/s0144686x16000106](https://doi.org/10.1017/s0144686x16000106)]
65. Kilshaw A, Holmes W, Matteucci P. Funding in facial palsy. *J Plast Reconstr Aesthet Surg* 2016 Nov;69(11):1526-1527. [doi: [10.1016/j.bjps.2016.08.008](https://doi.org/10.1016/j.bjps.2016.08.008)] [Medline: [27637591](#)]
66. Johnson K. Facial Palsy UK Annual Report and Unaudited Financial Statements. 2017 Jun 30. URL: [https://www.facialpalsy.org.uk/wp/wp-content/uploads/2017/11/annual\\_report\\_2017.pdf](https://www.facialpalsy.org.uk/wp/wp-content/uploads/2017/11/annual_report_2017.pdf) [accessed 2020-09-14]
67. Asthana S, Jones R, Sheaff R. Why does the NHS struggle to adopt eHealth innovations? A review of macro, meso and micro factors. *BMC Health Serv Res* 2019 Dec 21;19(1):984 [FREE Full text] [doi: [10.1186/s12913-019-4790-x](https://doi.org/10.1186/s12913-019-4790-x)] [Medline: [31864370](#)]
68. Desveaux L, Soobiah C, Bhatia RS, Shaw J. Identifying and overcoming policy-level barriers to the implementation of digital health innovation: qualitative study. *J Med Internet Res* 2019 Dec 20;21(12):e14994 [FREE Full text] [doi: [10.2196/14994](https://doi.org/10.2196/14994)] [Medline: [31859679](#)]
69. Webster P. Virtual health care in the era of COVID-19. *Lancet* 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30818-7](https://doi.org/10.1016/S0140-6736(20)30818-7)] [Medline: [32278374](#)]

## Abbreviations

**ENT:** ear, nose, and throat  
**FDI:** Facial Disability Index  
**FNP:** facial nerve paralysis



**FRAME:** Facial Remote Activity Monitoring Eyewear  
**GP:** general practitioner  
**HB:** House-Brackmann  
**HRQoL:** health-related quality of life  
**MDT:** multidisciplinary team  
**NHS:** National Health Service  
**NICE:** National Institute for Health and Care Excellence  
**NIHR:** National Institute for Health Research  
**NMR:** neuromuscular retraining  
**QALY:** quality-adjusted life year  
**RR:** response rate  
**SDT:** standard drug treatment

*Edited by G Eysenbach; submitted 18.05.20; peer-reviewed by J Diels, G Volk; comments to author 12.06.20; revised version received 18.06.20; accepted 25.06.20; published 05.10.20.*

*Please cite as:*

*Szczepura A, Holliday N, Neville C, Johnson K, Khan AJK, Oxford SW, Nduka C*

*Raising the Digital Profile of Facial Palsy: National Surveys of Patients' and Clinicians' Experiences of Changing UK Treatment Pathways and Views on the Future Role of Digital Technology*

*J Med Internet Res 2020;22(10):e20406*

URL: <https://www.jmir.org/2020/10/e20406>

doi: [10.2196/20406](https://doi.org/10.2196/20406)

PMID: [32763890](https://pubmed.ncbi.nlm.nih.gov/32763890/)

©Ala Szczepura, Nikki Holliday, Catriona Neville, Karen Johnson, Amir Jahan Khan Khan, Samuel W Oxford, Charles Nduka. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Opportunities and Challenges Surrounding the Use of Data From Wearable Sensor Devices in Health Care: Qualitative Interview Study

Ijeoma Azodo<sup>1</sup>, MD, ChM, MPH; Robin Williams<sup>2</sup>, MA, MSc, PhD; Aziz Sheikh<sup>1</sup>, OBE, FRSE, FMedSci; Kathrin Cresswell<sup>1</sup>, BSc, MSc, PhD

<sup>1</sup>Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

<sup>2</sup>Institute for the Study of Science, Technology and Innovation, University of Edinburgh, Edinburgh, United Kingdom

**Corresponding Author:**

Kathrin Cresswell, BSc, MSc, PhD

Usher Institute

University of Edinburgh

Teviot Place

Edinburgh, EH8 9AG

United Kingdom

Phone: 44 131 651 4151

Email: [kathrin.cresswell@ed.ac.uk](mailto:kathrin.cresswell@ed.ac.uk)

## Abstract

**Background:** Wearable sensors connected via networked devices have the potential to generate data that may help to automate processes of care, engage patients, and increase health care efficiency. The evidence of effectiveness of such technologies is, however, nascent and little is known about unintended consequences.

**Objective:** Our objective was to explore the opportunities and challenges surrounding the use of data from wearable sensor devices in health care.

**Methods:** We conducted a qualitative, theoretically informed, interview-based study to purposefully sample international experts in health care, technology, business, innovation, and social sciences, drawing on sociotechnical systems theory. We used in-depth interviews to capture perspectives on development, design, and use of data from wearable sensor devices in health care, and employed thematic analysis of interview transcripts with NVivo to facilitate coding.

**Results:** We interviewed 16 experts. Although the use of data from wearable sensor devices in health and care has significant potential in improving patient engagement, there are a number of issues that stakeholders need to negotiate to realize these benefits. These issues include the current gap between data created and meaningful interpretation in health and care contexts, integration of data into health care professional decision making, negotiation of blurring lines between consumer and medical care, and pervasive monitoring of health across previously disconnected contexts.

**Conclusions:** Stakeholders need to actively negotiate existing challenges to realize the integration of data from wearable sensor devices into electronic health records. Viewing wearables as active parts of a connected digital health and care infrastructure, in which various business, personal, professional, and health system interests align, may help to achieve this.

(*J Med Internet Res* 2020;22(10):e19542) doi:[10.2196/19542](https://doi.org/10.2196/19542)

**KEYWORDS**

wearable sensor devices; health care; data; qualitative

## Introduction

Many countries now see digital transformation and digitally enabled self-management as a way to tackle key challenges of delivering health and well-being with limited health care resources to an increasingly aging population [1]. The

transformatory potential of wearable devices forms part of a pervasive set of technological innovations characterized as Internet of Things (IoT) technology; these devices are typically mounted on wireless broadband technologies (ie, mediated by smartphones or equivalent devices) and are often linked to social media platforms. IoT refers to interconnected physical objects,

devices, and systems that electronically communicate, receive, process, and transfer digital data with limited direct human input [2-5]. In health care practices, IoT connectivity allows the sharing of data from wearable sensor devices across various contexts, potentially encompassing the user, caregivers, clinicians, and the health record. Wearable sensor devices include tracking devices with capabilities to count steps, measure activity, and check heart rate. Advanced smartwatches have sensing capabilities that include temperature, blood pressure, stress, sleep quality, respiratory rate, physical activity, electrocardiograms, acceleration, and oximetry [6]. These capabilities are also present in wearable sensor devices used in hospitals and clinics, albeit with less investment in the device aesthetics. Technologies are considered “smart” if they monitor, collect, and send data. They differ from mobile devices such as smartphones in that they are wearable and do not require active data input or curation from the user.

The promise of IoT in health care lies in its potential to realize business benefits while providing a valuable, integrated health and social care service and superior experience for citizens and patients [7]. Growing internet connectivity and the increasing availability of smartphones and multiple portable devices facilitate digital information sharing, connected devices, and the possibilities for digital health care [8]. These interrelationships are part of a societal reorientation of new networked organizations that use information and digital data to allow new insights for business, populations, and health care delivery [9]. Here, wearable and sensor devices enable health and care practices through internet connectivity combined with mobile, miniature, pervasive computing [10,11]. It is hoped that emerging technologies will achieve a similar return on investment, productivity, and efficiency savings as seen in the financial and manufacturing industries [12,13]. Existing empirical studies of wearables in health care demonstrate the feasibility of creating technologies [14] and running data analytics [15,16]; small pilot studies have demonstrated the use of wearables in managing specific chronic conditions, including diabetes and high blood pressure, as well as contributing to data analytics [17-20]. There is also some evidence from trials that wearables can promote healthy behavior [21-23]. However, because these technologies have transformative potential, and there is limited adoption experience, their ultimate outcomes and significance cannot reliably be assessed at this moment. Nevertheless, the area is full of potential, with many apps already becoming routinely integrated with everyday life, with 1 in 5 Americans now possessing a smartwatch or fitness tracker [24].

There are a number of challenges associated with the use of wearable and sensor devices. Firstly, they are often conceptualized as consumer goods, which potentially limits their use as health apps. Secondly, they can change health care delivery models and also impact on the collection and dissemination of sensitive personal data. For example, continuous monitoring data collected through IoT technologies has the potential to improve the knowledge base for decision making but stands in contrast to the intermittent nature of clinical consultations. Health care professionals already feel overwhelmed with data, and patients may not wish to be

monitored [25,26]. In addition, studies from the field of dependable computing have shown that systems produce many false-positive alerts, calling the usefulness of these applications into question [27,28]. Such issues may lead to unintended consequences that can, in turn, result in poor patient outcomes and safety risks [29-32]. Health care data are subject to particular concerns and special privacy and confidentiality protections. The use of wearable sensor devices for health care therefore introduces additional layers of complexity surrounding information privacy, data sharing, autonomy, consent, ownership, data access, and data valuation [25].

Industry reports suggest that as a system, IoT, including wearable sensor devices, is an established emerging technology, estimating 5-10 years until it reaches its “plateau of productivity” [33]. The area is, therefore, ripe for scientific investigation. Here, it is crucial to identify emerging risks and challenges in order to anticipate and mitigate them. Therefore, we aimed to explore the opportunities, gaps, and challenges of wearable sensor devices, as one component of an IoT system, in health and care settings.

## Methods

### Overview and Ethical Approval

We conducted a qualitative interview-based study with international experts and stakeholders involved in designing, developing, and implementing wearable sensor devices in health care [34]. This work was theoretically informed drawing on sociotechnical systems theory [35]. Institutional ethical approval was obtained from the University of Edinburgh Usher Research Ethics Group.

### Sampling and Recruitment

We developed a sampling frame by searching the technical, cultural, media, academic, and health care literature to capture the various perspectives and areas of expertise surrounding wearable sensor devices in health care settings. This was used as a basis for recruitment and was supplemented by consulting with local experts at the University of Edinburgh in culture, science and technology, and IoT studies. We used a purposive sampling strategy aiming for maximum variation in order to gain insights into the range of perspectives that need to be taken into account when considering the implementation and adoption of wearable sensor devices in health care settings. In so doing, we sought out experts with experience across a range of health care settings, technological functionalities, and disciplinary backgrounds [36]. This purposive sampling approach was supplemented by snowball sampling using recommendations from the initial purposive sample [37].

Participants were recruited by emails that were sent to publicly available email addresses; each message introduced the project and requested an interview. Interviews were scheduled with respondents who replied and provided informed consent to participate. We sent a total of 79 emails to potential participants. A follow-up email was sent to nonresponders 1 week later, after which we ceased further contact assuming no interest in participation. We stopped recruitment when we reached thematic

saturation (ie, when no new themes emerged in the concurrent analysis) [38].

## Data Collection

We collected data through in-depth, semistructured, one-to-one, qualitative telephone and video interviews from May to July 2019. Identified experts were located in various international settings. Interviews were guided by a topic guide exploring personal and professional backgrounds and experiences of, and involvement with, wearable sensor devices in health and care

settings (see [Textbox 1](#)) [39]. The questions were based on the existing literature in order to understand the current opportunities, gaps, and challenges associated with the use of wearable sensor devices in health and care settings. They were tailored to the specific characteristics of participants and evolved in light of emerging findings. In doing so, we explored visions of technologies; experiences of implementing, using, and researching technologies, including benefits and risks; and features conducive to accelerating the use of wearable sensor devices in health and care settings.

### Textbox 1. Example topic guide.

1. Interviewee's background: current position and role in relation to health information technology.
2. Their vision surrounding the use of wearable sensor devices in health care:
  - What different models of wearable sensor devices in health care exist?
  - Most promising developments to look out for and their benefits
  - Devices and technologies to collect and analyze data
  - Integration with clinical data and interfacing with other technologies.
3. Experiences of technological innovation in health care:
  - Experiences and lessons learned (eg, consent models, data security, motivating users, scale, and data analytics)
  - How does health care differ from other sectors? Anything we can learn?
  - Which factors hinder developments and how might these be addressed?
4. Accelerating innovation in health care:
  - Perceived risks
  - How can this approach be scaled up?
  - What is the role of regulation and government?
5. Anything else?

The interviews were recorded using Zoom conferencing (Zoom Video Communications) or an encrypted digital audio recorder. We read transcripts multiple times annotating with notes to immerse ourselves in the data. We also kept a field journal noting impressions and emerging analytical thoughts.

Following the interviews, the digitally recorded audio files were transferred via a secure virtual private network to a secure site for professional transcription. Interview transcripts, notes, and file names were anonymized.

## Data Analysis

IA, a surgeon and health services researcher, led the data analysis. The analysis drew on sociotechnical systems theory, which assumes reciprocal shaping of technologies and people by their interactions with each other [40]. Through this lens, we viewed the process of integrating wearable sensor devices into health care practices as a sociotechnical practice where technologies shape human behaviors and these, in turn, shape technological design. We analyzed data concurrently with ongoing data collection and stopped collecting new data when no new themes emerged in the analysis.

Completed transcripts and field notes were organized and coded using NVivo 12 (QSR International) [41]. We used a concept-driven thematic approach based on the interview guide

and key concepts in health [42], technology [43,44], unintended consequences in health information technology (HIT) [45], and innovation to analyze the data [46]. In doing so, we adhered to the six key features of an inductive and deductive coding approach, namely (1) developing the codebook with key themes a priori, (2) testing code reliability, (3) summarizing data and identifying key themes, (4) applying coding templates and new codes, (5) connecting codes and identifying themes, and (6) corroborating and legitimating coded themes [47]. We used constant comparison to examine themes, irregularities, and patterns across interviews [48].

## Results

### Overview

We conducted 16 in-depth, semistructured interviews lasting between 30 and 75 minutes with experts in technology, business, health care, and innovation (see [Table 1](#)); 1 interview was conducted in person and 15 interviews were audio recorded. All study participants had at least 5 and up to 20 years' experience in their fields. Contributing participants were located in Europe, the United Kingdom, and the United States.

The analysis generated data around four main themes that demonstrated the significant potential of wearable sensor IoT

in health care, but also highlighted issues around data provenance and quality; transforming health and care relationships; blurring of business, consumer, health, and clinical boundaries; and issues around privacy, confidentiality, and data ownership (see [Textbox 2](#)).

Participants agreed that the application of wearable sensor devices in health and care has significant potential to positively transform health outcomes. Interviewees gave examples of the potential to enhance alerting, increase patient and public involvement, facilitate processing of information, and improve efficiency, while increasing the reliability and reducing the cost of manual data capture. Participants emphasized different issues, according to their professional backgrounds. For example, academics and researchers focused on social implications; technologists and engineers on increased connectivity; and clinicians on patient safety, care processes, and quality of care.

*So a lot of what I'm more focused on is wearables in everyday life, so the way in which they inspire people to adapt particular modes of care and wellness, so the idea that one wears a Fitbit as a motivational tool or a social tool to share their step counts with people as then creating another form of motivation.* [Participant #3, historian, United States]

*By delivering care outside the hospital you save a lot more money. By keeping patients outside the hospital with these devices that can proactively monitor and*

*maybe even prevent certain events. That gives significant cost savings to the industry.* [Participant #12, engineering and business, United States]

Notably, participants expressed little dissonance in conceptualizing wearable sensor devices as a component of an information ecosystem consisting of various technologies and humans. As such, technological functions were seen as vessels that collected and disseminated data that would then travel through the ecosystem and impact human and organizational behavior, as data were interpreted and acted on.

*I realized that the real value in this whole equation is data, and if you can get the data and you can own the data from a wearable device, then we can turn that into meaningful value for someone with an ecosystem, whether that's patients, providers, caregivers, sports teams, someone within the health universe.* [Participant #5, business and innovation, United States]

However, many of these hopes were associated with future developments, interoperability of technologies, improved fidelity to physiology, and enhanced machine learning and predictive capabilities. For the most immediate future, participants highlighted a number of concerns that were perceived to inhibit the potential of wearable sensor devices in health and care settings. These all related to the intersection of social and technological dimensions and will be discussed in turn.

**Table 1.** Study participants' expertise, industries, and work locations.

Participant No.	Expertise	Sector	Location
1	Product developer	Technology	United States
2	Science and technology studies	Academia	United Kingdom
3	Historian	Academia	United States
4	Engineering and business	Academia Technology	United States
5	Neuroscience, psychology, and business	Technology	United States
6	Law	Academia	United States
7	Science and technology studies	Academia	Northern Europe
8	Health care and life sciences	Consulting	United Kingdom
9	Design and innovation Business	Health care and pharmaceuticals	United States
10	Social and behavioral science	Academia	Western Europe
11	Electrical, computer, and biomedical engineering	Academia	United States
12	Business and engineering	Technology	United States
13	Surgery	Health care Academia	United Kingdom
14	Biomedical engineering	Consulting (business)	United Kingdom
15	Rehabilitation Electrical engineering	Public health	United States
16	Quality and improvement Nursing	Health care	United States

**Textbox 2.** Key themes and subthemes identified in our analysis.

Significant potential of wearable sensor devices in health and care:

1. From data to action: the role of data provenance and data quality:

- Issues of objective measurement of activity and translation into signals for effective evidence-based decision making
- Data provenance and quality as central considerations for effective adoption and diffusion
- Risk of data overload among health care professionals.

2. Transforming relationships through wearable sensor devices and associated data:

- Integration of devices and data with existing workflows of health care professionals
- Wearable sensor devices as potentially perceived mechanisms of surveillance and control
- Disruption of the traditional hierarchy of specialist information processing through health and care professionals
- Impact on the provider and patient relationship.

3. Increasing blurring of business and consumer as well as health and care boundaries:

- Tension between consumer product and use for wider public good
- Asymmetry of interests through financial sustainability of wearable sensor devices
- Need for new business models.

4. Privacy, confidentiality, and data ownership:

- Lack of regulation of data flows between health care, commercial, and private spaces
- Need for development of standards, especially for apps that were not designed primarily for use in health- and care-related settings.

## From Data to Action: The Role of Data Provenance and Data Quality

Participants highlighted that incorporating unannotated data into health and care practices was, at present, problematic, as it decoupled social and technological dimensions and thereby assumed that (1) data presented a robust and consistent measurement of activity, (2) data could be translated into recognizable signals, and (3) data were appropriately situated in context for dependable, effective, and evidence-based decision making. A key issue here was establishing baseline, “normal,” pathologic, and “unusual” not otherwise-specified activity. High hopes were placed on algorithms to support the recognition of unusual signals that could then trigger clinical intervention.

*...using algorithms to interpret what normal activity for a patient is, or what a normal baseline is for that patient preoperatively. And then using algorithms and AI [artificial intelligence] to try and refine that and improve, either predicting that patients are deteriorating at an early stage; or purely for providing individual feedback to say, “actually, you’re progressing well,” or “you’re not progressing as much as we’d expect.”* [Participant #13, surgery, United Kingdom]

Data provenance and quality, therefore, became central considerations for effective use of data generated by wearables in health and care settings. This was particularly true where observed data patterns may not have distinct correlates to disease mechanisms or physiology. This makes it difficult to confidently base clinical decisions or interventions on them. Here, participants made distinctions between clinical, health, and

well-being data. For example, variability in data used for individual empowerment was sufficient and acceptable (eg, increasing a person’s daily step count), while variability in clinical parameters (eg, the measurement of blood glucose levels that may indicate the need to give an insulin injection) was viewed as problematic, as resulting human actions had potentially impactful consequences.

*...now we’ve got abundant data and we train algorithms, particularly if I do things like I drive a deep learning algorithm hard, then I’m going to find some pattern and some interesting stuff emerging about how you...you know, is that actually an artefact of the data? Yes, it’s a real pattern, but is it actually meaningful in any true physiological sense that’s a response of causal reality or only correlative?* [Participant #4, engineering and innovation, United States]

In addition, participants raised concerns that health care staff were already overloaded with data and that the integration of further, potentially irrelevant, data from additional devices could unnecessarily burden already-busy professionals and, therefore, fail to mobilize an important actor in the ecosystem. The interviews broadly reflected a presumption that device data would be actioned in consultation with a clinician and that the devices would serve to augment the clinician’s capability in making decisions, rather than eliminate them from the care process.

*The challenge is that, as yet, data that’s collected from wearables is very difficult to integrate into the health record of a patient. So, a patient turning up at their doctor’s with a whole list of data and tracking*



*information from their wearable isn't necessarily going to find a welcome recipient, in terms of the doctor.* [Participant #8, consulting, United Kingdom]

### Transforming Relationships Through Data Generated by Wearable Sensor Devices

Integration of devices and data with existing workflows of health care professionals was viewed as crucial for wearable sensor devices to become active parts of the ecosystem, as this was seen to facilitate the exchange of information between social and technological dimensions. However, participants shared concerns about the impact of new types of data on care processes and relationships in clinical work practices. Some participants viewed the increasing focus on data as challenging the autonomy of health care professionals in making decisions about how and when to provide care and as a method of surveillance and control. This was particularly apparent in interviews with clinicians.

*One of the biggest issues that we faced was with our nursing staff and their concern over sort of this idea of big brother and "people watching over what I'm doing," and "I'm a nurse, I know how to turn a patient, I really don't need a device to tell me how to do that."* [Participant #16, quality improvement and nursing, United States]

There were also some concerns that the wide availability of data generated by wearable sensor devices to a range of actors, including device companies, family members, and caregivers, may disrupt the traditional hierarchy of specialist information processing through health care professionals. Traditionally, clinicians captured and interpreted information in context at discreet moments in the consultation, and wearable sensor devices may disrupt this information flow.

*There are a lot of people who are contacting clinicians or caregivers and saying, "what's this?" On one hand, it's positive, but on the other hand, the health care system isn't structured for that type of engagement. And so, I think you're seeing a range of engagement from people around these devices.* [Participant #5, business and innovation, United States]

Health care providers raised concerns about becoming increasingly distanced from the patient, as they were getting health- and care-related information from sources other than the patient directly, and without recognized professional interpretation. This distance has implications on responsibility and accountability for information collected and the subsequent decisions and actions taken, or missed.

*For doctors, I think their big concern is, well, is my relationship with my patient now going to be diminished 'cause they have another place that they can go to track their symptoms and get information, so maybe I don't become that first-line person that they want to talk to?* [Participant #9, health care design and innovation, United States]

### Increasing Blurring of Business and Consumer as well as Health and Clinical Boundaries

Wearable sensor devices were seen as a key example of blurring boundaries between business and consumer as well as health and care worlds, and this posed challenges associated with the assumptions underlying the design of the technology and the contexts of use. Technology and business actors saw the promotion of a commercial market for consumer technology health and care applications, aligning this unproblematically with promoting public health.

*Companies like Samsung and Alexa, Apple, Google, they see health care as a potential pinpoint that their technology could potentially solve, and so it's a good thing, I think.* [Participant #5, business and innovation, United States]

However, for others, aligning actors around financing and development was a major rate-limiting step for the potential wider public good, as companies came with their own commercial strategies tied to their own proprietary standards. These participants were skeptical that data generated by wearable sensor devices could be used to allow better management of individual health conditions. This indicates a misalignment of higher-level health policy goals and the reality of evidencing improved outcomes for patients and business functions using technology in health and care.

*One of the upshots is it's encouraging healthier behavior overall, I'm for that. I want people to live their healthiest best lives. And if there's a technology that allows us to do that, I'm a little skeptical, but if there is, I'm for it.* [Participant #6, law, United States]

Payments and financial sustainability of wearable sensor devices in health care was raised as a concern about the technology partners' motives, perhaps contributing to a lack of adoption in clinical contexts. In addition to creating products and services to support good health outcomes, uncertainty remained about sustainability. This included issues around maintenance and product iteration, updating equipment and devices, continued use of devices by health care professionals and patients, aligning of different interests over time, and financing of ongoing data analyses. These queries indicated a desire to rethink how goods and services are designed and delivered when technology is closely intertwined with health care. Existing business models reflected in existing technological designs were considered inappropriate.

*Our business model was one which we assume we'd have to make money on subscription income, because in order to deliver a good experience, the bill or materials costs can be high, and to make the economics stack up, you couldn't do it by charging somebody up front for the device. The price would be too high, it could be a couple thousand dollars at that point.* [Participant #4, engineering and innovation, United States]

### Privacy, Confidentiality, and Data Ownership

Participants also raised questions about misuse of data, balancing how much data to collect, data ownership, data portability, and

implications for insurability. It was unclear how the data generated in devices were regulated, and many participants raised concerns of data flows between health care, commercial, and private spaces. The overarching concern of clinicians in this area was the impact of altered data flows on trust, given the lack of clarity over control, sharing, and repurposing of health-related data.

*I think the big question [that] will come about is actually, well, who owns that data? And who uses that data? Because the problem is, if you then have it on a Google server, then is it not only accessible by the researchers and by the clinicians but accessible by the company themselves. Then it only takes, kind of, one example of that industry partner using that data for something else, that potentially patients haven't consented for, that it then starts to become quite a big issue and then trust, kind of, is degraded...* [Participant #13, surgery, United Kingdom]

Important considerations for developers and implementers included the development of technology standards, especially as many applications were not designed primarily for use in health- and care-related settings. This was perceived to promote trust among the user community and ensure a closer fit between technological design and clinical practice.

*What is needed also are some standards. So a lot of these technologies, particularly in the consumer technologies, have been developed as a consumer product and they won't necessarily conform to the health care standards that are needed, either in terms of the technology, the interoperability standards, in terms of the health care side of it, they won't have passed as a medical device for health care.* [Participant #8, consulting, United Kingdom]

Similarly, participants also called for increasing regulatory efforts to ensure that devices that were applied to health and care systems achieved maximum benefit, while minimizing harm.

*It's to work collaboratively with regulators so that we expedite and maximize the potential of the technologies that we have without doing any harm to patients.* [Participant #14, biomedical engineering and consulting, United States]

## Discussion

### Principal Findings

Wearable sensor devices have significant potential, particularly in relation to promoting patient engagement. However, achieving these potential benefits is dependent on addressing the current gap between data created and meaningful interpretation in health and care contexts. Current applications cannot fulfill their potential if they do not yield benefits for clinical users and thereby integrate effectively within the existing ecosystem of social and organizational actors. This may be achieved through modifying technological design to allow data to integrate effectively into health care professional decision making, negotiating of blurring lines between consumer and medical

care, and appropriate regulatory contexts incorporating pervasive monitoring of health.

### Strengths and Limitations

We have drawn on a range of perspectives from various sectors to explore the potential benefits and challenges of incorporating data generated by wearable sensor devices into health and care settings. The breadth of participant expertise across disciplines, projects, and national contexts allowed triangulation and constant comparison of generated data. These in-depth interviews have identified a range of barriers that stakeholders need to negotiate and provide an overview of the extended health care ecosystem of care providers and technology developers that can help to maximize the potential benefits of this technology. Actively anticipating and mitigating risks, before they impact negatively on the safety and quality of health and care, can help to inform future system design, implementation, and optimization.

Some limitations of this study include the diversity of applications and contexts discussed, so that it was at times difficult to extract common themes, and a potential lack of generalizability of findings across settings with different regulatory environments (as most participants came from the United States). Providing accessible and broad coverage of key concerns in the field of applying wearable sensors in health care may also have compromised in-depth insights into the concerns of different stakeholder groups. Unfortunately, none of the insurance companies that we reached out to on multiple occasions responded to our invitation to be interviewed, so we were not able to gain insights into these perspectives. There may also have been pressure on participants to construct knowledge during the interviews. For example, participants responded to a handful of the interview questions with "good question," and their responses likely represented efforts to collate their thoughts and experiences into durable representations of the subject [49].

### Integration of Findings With the Current Literature

The value of this work lies in providing accessible and broad coverage of key concerns in the field of applying wearable sensors in health and care settings. It shows that wearable sensor devices should not be viewed as stand-alone technologies in isolation, but as part of an emerging IoT ecosystem of actors. In doing so, they emerge as a key driver of service redesign in health care innovation. Our analysis revealed tensions in overlapping goals and agendas that challenge the effective integration of data generated by wearable sensor devices into health and care settings. This finding is similar to other sociotechnical studies of HIT where a range of actors with different agendas have to align for successful adoption [50-52].

Wearable sensor devices and associated data, however, do add an extra layer of complexity. Data fluidity means that many stakeholders have to be mobilized in different ways to ensure that data are captured and relayed meaningfully to various actors. In addition, this case also vividly illustrates tensions between visions and real-life challenges of using mobile devices to connect health-related data to meet public health and consumer market agendas. It further highlights the need to

develop new business models that do not exploit certain groups of stakeholders for commercial agendas.

Data generated by wearable sensor devices also create new challenges and opportunities for health care professional work practices and relationships with patients. Previous work has surfaced unintended consequences of HIT on communication, coordination of care, and clinical work practices that our findings confirm [53]. In addition, we have shown how data from wearable sensor devices can introduce new insights for, but also disruptions to, communication and work practices of health and care professionals [54].

Given the potential consequences of acting on data improperly, or of acting on improper data, this study highlights the importance of addressing data provenance and quality and of presenting relevant information. Existing commercial stakeholders are increasingly building on this. Health data aggregation companies, such as 1upHealth [55], Open mHealth [56], Validic [57], Seqster [58], Human API [59], and Apple combine information from many data sources, wearables, apps, sensors, and electronic health records, producing information for health care organizations, clinicians, and users. New start-ups like Conversa and LifeWIRE have added human and nonhuman intermediaries to help patients and providers navigate aggregated health data [60,61]. Generally, efforts promoting solutions that blur lines between consumer and medical care too much are unlikely to be effective, as a functioning ecosystem requires clinicians to be active actors to realize the potential benefits of wearable sensor devices in health and care settings. Nevertheless, changing social contexts, as seen in the COVID-19 pandemic, may change these dynamics. For example, companies like Current Health have expanded their range of sensors and connections to peripheral devices providing patient health data for remote monitoring [62].

Overall, while there is, at present, limited interest in bringing data generated by wearable sensor devices into the care consultation from the provider side [63-65], the ongoing COVID-19 pandemic has mandated a reconsideration for how best to provide care at a distance using wearable sensor devices. Even within this unexpected contextual shift, interconnected devices and data will need to represent the communications, contexts, conversations, and social connectedness of care interactions in order to maximize usefulness and effectiveness for a range of stakeholders [66-68].

### **Policy Recommendations and Implications for Practice Emerging From This Work**

Wearable sensor devices are best conceptualized as data-generating components within a distributed information system. They are not a simple device implementation; rather, such devices are key components of regular input into an IoT platform in a health care ecosystem. Further mixed methods research now needs to demonstrate effects of these devices on cost and health care outcomes, how data are used, and how health care knowledge and practices are presented and represented in the data obtained.

Implications for accelerating the integration of data generated by wearables into health care practices include negotiating the fit with existing health care professional and patient relationships, mitigating adverse unintended consequences, and aligning interrelated agendas. Triangulation of the agendas of providers and those that organize care will be key, as this seems to be crucial for promoting effective self-management. Conceptualizing specific solutions as complex public health interventions with aligned business models could help to achieve this, facilitated by rigorous evaluation to achieve mainstream utility and use [69,70].

Interviewees highlighted that regulatory guidance was lacking in the field; there were also international differences in regulatory contexts. Regulators do not classify most wearable sensors as medical devices, despite increasingly blurring lines [71,72]. There is evidence to support that regulation can both facilitate and stifle innovation [73,74]. The regulatory landscape for wearable sensor devices in health care is challenging because of the distributed stakeholder network, the involvement of technical specialists, tight regulations, and privacy and security concerns. Whether data generated by wearable sensor devices will contribute to the transformation of health and care will, in part, depend on deliberate and consistent regulatory policy to reduce uncertainty around investments [75].

In the United States, such efforts are likely to be most effectively led by national bodies to facilitate widespread diffusion and sustain wearable sensors in routine practices. The United States Centers for Medicaid Services proposed rules to take effect in 2021 that clarify that devices supporting remote physiological monitoring must be defined as medical devices by the United States Food and Drug Administration and must be “reliable,” “valid,” and collect data “electronically” to qualify [76]. This proposal follows expansion of physician reimbursement fees for remote physiological monitoring in 2020. These fee schedules create a favorable environment for expansion of existing and emerging digital health services using devices to take physiological and biometric measurements [77].

### **Conclusions**

Our work suggests that meaningful use of wearable sensor devices in health and care settings occurs through a platform of interconnected devices and data users around a specific use case. This platform represents the sum of knowledge, practices, and contexts of health care, oriented to improve a system of care, rather than a singular focus on a specific device. Engaging the range of relevant stakeholders who participate in design and development of systems is therefore essential to maximize scale, impact, and adoption.

Wearable sensor devices have great potential in improving patient engagement and thereby contributing to preventive, diagnostic, and treatment approaches. Such technologies are, however, most likely to be successful in achieving this potential if systems can align business, professional, personal, and health systems agendas.



## Acknowledgments

We gratefully acknowledge all participants' time and input. This work was funded by a Scottish Government Chief Scientist Research Grant. The views expressed are those of the authors.

## Conflicts of Interest

None declared.

## References

- Lapão LV. The future impact of healthcare services digitalization on health workforce: The increasing role of medical informatics. *Stud Health Technol Inform* 2016;228:675-659. [Medline: [27577470](#)]
- Dimitrov DV. Medical Internet of Things and big data in healthcare. *Healthc Inform Res* 2016 Jul;22(3):156-163 [FREE Full text] [doi: [10.4258/hir.2016.22.3.156](#)] [Medline: [27525156](#)]
- Marques R, Gregório J, Mira Da Silva M, Lapão LV. The promise of the Internet of Things in healthcare: How hard is it to keep? *Stud Health Technol Inform* 2016;228:665-669. [Medline: [27577468](#)]
- Metcalfe D, Millard ST, Gomez M, Schwartz M. Wearables and the Internet of Things for health: Wearable, interconnected devices promise more efficient and comprehensive health care. *IEEE Pulse* 2016 Sep;7(5):35-39. [doi: [10.1109/mpul.2016.2592260](#)]
- Cahill J, Portales R, McLoughlin S, Nagan N, Henrichs B, Wetherall S. IoT/sensor-based infrastructures promoting a sense of home, independent living, comfort and wellness. *Sensors (Basel)* 2019 Jan 24;19(3):485 [FREE Full text] [doi: [10.3390/s19030485](#)] [Medline: [30682864](#)]
- Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: Promises and barriers. *PLoS Med* 2016 Feb;13(2):e1001953 [FREE Full text] [doi: [10.1371/journal.pmed.1001953](#)] [Medline: [26836780](#)]
- Medtech and the Internet of Medical Things. Deloitte. 2020. URL: <https://www2.deloitte.com/uk/en/pages/life-sciences-and-healthcare/articles/medtech-and-the-internet-of-medical-things.html> [accessed 2020-08-24]
- Thümmel C. Digital health. In: Fricker SA, Thümmel C, Gavras A, editors. *Requirements Engineering for Digital Health*. Cham, Switzerland: Springer; 2015:1-23.
- Belliger A, Krieger DJ. The digital transformation of healthcare. In: North K, Maier R, Haas O, editors. *Knowledge Management in Digital Change: New Findings and Practical Cases*. Cham, Switzerland: Springer; 2018:311-326.
- Hung K, Zhang Y, Tai B. Wearable medical devices for tele-home healthcare. In: *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2004 Presented at: 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; September 1-5, 2004; San Francisco, CA p. 5384-5387. [doi: [10.1109/iembs.2004.1404503](#)]
- Price L. The digital health hype cycle 2019. *Healthcare.Digital*. 2020 Mar 31. URL: <https://www.healthcare.digital/single-post/2019/01/12/The-Digital-Health-Hype-Cycle-2019> [accessed 2020-08-24]
- Biesdorf S, Niedermann F. Healthcare's digital future. 2014 Jul 01. URL: <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/healthcares-digital-future> [accessed 2020-08-24]
- McKinsey Global Institute. *The Productivity Puzzle: A Closer Look at the United States*. New York, NY: McKinsey & Company; 2017 Mar. URL: <https://www.mckinsey.com/~media/mckinsey/featured%20insights/employment%20and%20growth/new%20insights%20into%20the%20slowdown%20in%20us%20productivity%20growth/mgi-the-productivity-puzzle-discussion-paper.ashx> [accessed 2020-08-24]
- Qureshi F, Krishnan S. Wearable hardware design for the Internet of Medical Things (IoMT). *Sensors (Basel)* 2018 Nov 07;18(11):3812 [FREE Full text] [doi: [10.3390/s18113812](#)] [Medline: [30405026](#)]
- Athavale Y, Krishnan S. A device-independent efficient actigraphy signal-encoding system for applications in monitoring daily human activities and health. *Sensors (Basel)* 2018 Sep 06;18(9):2966 [FREE Full text] [doi: [10.3390/s18092966](#)] [Medline: [30200566](#)]
- Rathore MM, Ahmad A, Paul A, Wan J, Zhang D. Real-time medical emergency response system: Exploiting IoT and big data for public health. *J Med Syst* 2016 Dec;40(12):283. [doi: [10.1007/s10916-016-0647-6](#)] [Medline: [27796839](#)]
- Bhide A, Muthukumar S, Prasad S. CLASP (Continuous lifestyle awareness through sweat platform): A novel sensor for simultaneous detection of alcohol and glucose from passive perspired sweat. *Biosens Bioelectron* 2018 Oct 15;117:537-545 [FREE Full text] [doi: [10.1016/j.bios.2018.06.065](#)] [Medline: [29982125](#)]
- Istepanian R, Hu S, Philip N, Sungoor A. The potential of Internet of m-health Things. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2011 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 30-September 3, 2011; Boston, MA p. 5264-5266. [doi: [10.1109/IEMBS.2011.6091302](#)]
- Kario K, Tomitani N, Kanegae H, Yasui N, Nishizawa M, Fujiwara T, et al. Development of a new ICT-based multisensor blood pressure monitoring system for use in hemodynamic biomarker-initiated anticipation medicine for cardiovascular disease: The national IMPACT program project. *Prog Cardiovasc Dis* 2017;60(3):435-449. [doi: [10.1016/j.pcad.2017.10.002](#)] [Medline: [29108929](#)]

20. Fagherazzi G, El Fatouhi D, Bellicha A, El Gareh A, Affret A, Dow C, et al. An international study on the determinants of poor sleep amongst 15,000 users of connected devices. *J Med Internet Res* 2017 Oct 23;19(10):e363 [FREE Full text] [doi: [10.2196/jmir.7930](https://doi.org/10.2196/jmir.7930)] [Medline: [29061551](https://pubmed.ncbi.nlm.nih.gov/29061551/)]
21. Kim J, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of personal health records and wearables on health outcomes and patient response: Three-arm randomized controlled trial. *JMIR Mhealth Uhealth* 2019 Jan 04;7(1):e12070 [FREE Full text] [doi: [10.2196/12070](https://doi.org/10.2196/12070)] [Medline: [30609978](https://pubmed.ncbi.nlm.nih.gov/30609978/)]
22. Hardcastle SJ, Hince D, Jiménez-Castuera R, Boyle T, Cavalheri V, Makin G, et al. Promoting physical activity in regional and remote cancer survivors (PPARCS) using wearables and health coaching: Randomised controlled trial protocol. *BMJ Open* 2019 May 24;9(5):e028369 [FREE Full text] [doi: [10.1136/bmjopen-2018-028369](https://doi.org/10.1136/bmjopen-2018-028369)] [Medline: [31129600](https://pubmed.ncbi.nlm.nih.gov/31129600/)]
23. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: A randomized controlled trial. *Telemed J E Health* 2015 Oct;21(10):782-792 [FREE Full text] [doi: [10.1089/tmj.2014.0176](https://doi.org/10.1089/tmj.2014.0176)] [Medline: [26431257](https://pubmed.ncbi.nlm.nih.gov/26431257/)]
24. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2020 Jan 09. URL: <https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/> [accessed 2020-10-09]
25. Mittelstadt B. Ethics of the health-related Internet of Things: A narrative review. *Ethics Inf Technol* 2017 Jul 4;19(3):157-175. [doi: [10.1007/s10676-017-9426-4](https://doi.org/10.1007/s10676-017-9426-4)]
26. Hall A, Walton G. Information overload within the health care system: A literature review. *Health Info Libr J* 2004 Jun;21(2):102-108 [FREE Full text] [doi: [10.1111/j.1471-1842.2004.00506.x](https://doi.org/10.1111/j.1471-1842.2004.00506.x)] [Medline: [15191601](https://pubmed.ncbi.nlm.nih.gov/15191601/)]
27. Dewsbury G, Clarke K, Rouncefield M, Sommerville I. Designing appropriate assistive technology for home users: Developing dependable networks. In: Proceedings of the Inclusive Design and Mobility Response in Indoor/Outdoor Public Buildings and Facilities: CIB Working Group W084 – Building Non-Handicapping Environments. 2002 Presented at: Inclusive Design and Mobility Response in Indoor/Outdoor Public Buildings and Facilities: CIB Working Group W084 – Building Non-Handicapping Environments; October 21-22, 2002; Rome, Italy p. 1-14 URL: <https://tinyurl.com/y3foa8gs>
28. Sommerville I, Dewsbury G. Dependable domestic systems design: A socio-technical approach. *Interact Comput* 2007 Jul;19(4):438-456. [doi: [10.1016/j.intcom.2007.05.002](https://doi.org/10.1016/j.intcom.2007.05.002)]
29. Zheng K, Abraham J, Novak LL, Reynolds TL, Gettinger A. A survey of the literature on unintended consequences associated with health information technology: 2014–2015. *Yearb Med Inform* 2018 Mar 06;25(01):13-29 [FREE Full text] [doi: [10.15265/iy-2016-036](https://doi.org/10.15265/iy-2016-036)]
30. Alami H, Gagnon M, Fortin J. Some multidimensional unintended consequences of telehealth utilization: A multi-project evaluation synthesis. *Int J Health Policy Manag* 2019 Jun 01;8(6):337-352 [FREE Full text] [doi: [10.15171/ijhpm.2019.12](https://doi.org/10.15171/ijhpm.2019.12)] [Medline: [31256566](https://pubmed.ncbi.nlm.nih.gov/31256566/)]
31. Aarts J, Ash J, Berg M. Extending the understanding of computerized physician order entry: Implications for professional collaboration, workflow and quality of care. *Int J Med Inform* 2007 Jun;76 Suppl 1:S4-S13. [doi: [10.1016/j.ijmedinf.2006.05.009](https://doi.org/10.1016/j.ijmedinf.2006.05.009)] [Medline: [16798068](https://pubmed.ncbi.nlm.nih.gov/16798068/)]
32. Ash JS. Some unintended consequences of information technology in health care: The nature of patient care information system-related errors. *J Am Med Inform Assoc* 2003 Nov 21;11(2):104-112. [doi: [10.1197/jamia.m1471](https://doi.org/10.1197/jamia.m1471)]
33. Panetta K. 5 trends emerge in the Gartner hype cycle for emerging technologies. Gartner. 2018 Aug 16. URL: <https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/> [accessed 2020-10-07]
34. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. *BMC Med Res Methodol* 2011 Jun 27;11:100 [FREE Full text] [doi: [10.1186/1471-2288-11-100](https://doi.org/10.1186/1471-2288-11-100)] [Medline: [21707982](https://pubmed.ncbi.nlm.nih.gov/21707982/)]
35. Braithwaite J, Runciman WB, Merry AF. Towards safer, better healthcare: Harnessing the natural properties of complex sociotechnical systems. *Qual Saf Health Care* 2009 Feb;18(1):37-41. [doi: [10.1136/qshc.2007.023317](https://doi.org/10.1136/qshc.2007.023317)] [Medline: [19204130](https://pubmed.ncbi.nlm.nih.gov/19204130/)]
36. Creswell JW. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. 2nd edition. Thousand Oaks, CA: SAGE Publications; 2007.
37. Biernacki P, Waldorf D. Snowball sampling: Problems and techniques of chain referral sampling. *Sociol Methods Res* 2016 Jun 29;10(2):141-163. [doi: [10.1177/004912418101000205](https://doi.org/10.1177/004912418101000205)]
38. Yazan B. Three approaches to case study methods in education: Yin, Merriam, and Stake. *Qual Rep* 2015;20(2):134-152.
39. Seale C, Gobo G, Gubrium JF, Silverman D, editors. *Qualitative Research Practice*. London, UK: SAGE Publications; 2004.
40. Bolstrom RP, Heinen JS. MIS problems and failures: A social-technical perspective, Part I: The causes. *MIS Q* 1977;1(3):17-32 [FREE Full text] [doi: [10.2307/248710](https://doi.org/10.2307/248710)]
41. Download NVivo. QSR International. 2020. URL: <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/support-services/nvivo-downloads> [accessed 2020-10-08]
42. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp0630a](https://doi.org/10.1191/1478088706qp0630a)]



43. Jennex M. Re-visiting the knowledge pyramid. In: Proceedings of the 42nd Hawaii International Conference on System Sciences. 2009 Presented at: 42nd Hawaii International Conference on System Sciences; January 5-8, 2009; Big Island, HI p. 1-7. [doi: [10.1109/hicss.2009.361](https://doi.org/10.1109/hicss.2009.361)]
44. Weinberger D. The problem with the data-information-knowledge-wisdom hierarchy. Harvard Business Review. 2010 Feb 02. URL: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not> [accessed 2020-10-07]
45. Schukat M, McCaldin D, Wang K, Schreier G, Lovell N, Marschollek M, et al. Unintended consequences of wearable sensor use in healthcare. Yearb Med Inform 2018 Mar 06;25(01):73-86. [doi: [10.15265/iy-2016-025](https://doi.org/10.15265/iy-2016-025)]
46. Christensen CM, Bower JL. Customer power, strategic investment, and the failure of leading firms. Strateg Manage J 1996 Mar;17(3):197-218. [doi: [10.1002/\(sici\)1097-0266\(199603\)17:3<197::aid-smj804>3.0.co;2-u](https://doi.org/10.1002/(sici)1097-0266(199603)17:3<197::aid-smj804>3.0.co;2-u)]
47. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92. [doi: [10.1177/160940690600500107](https://doi.org/10.1177/160940690600500107)]
48. Padgett DK. Getting started: Study design and sampling. In: Qualitative and Mixed Methods in Public Health. Thousand Oaks, CA: SAGE Publications; 2012:65-80.
49. Myers MD, Newman M. The qualitative interview in IS research: Examining the craft. Inf Organ 2007 Jan;17(1):2-26. [doi: [10.1016/j.infoandorg.2006.11.001](https://doi.org/10.1016/j.infoandorg.2006.11.001)]
50. Hanseth O, Jacucci E, Grisot M, Aanestad M. Reflexive standardization: Side effects and complexity in standard making. MIS Q 2006;30:563. [doi: [10.2307/25148773](https://doi.org/10.2307/25148773)]
51. Robert G, Greenhalgh T, MacFarlane F, Peacock R. Adopting and assimilating new non-pharmaceutical technologies into health care: A systematic review. J Health Serv Res Policy 2010 Oct;15(4):243-250. [doi: [10.1258/jhsrp.2010.009137](https://doi.org/10.1258/jhsrp.2010.009137)] [Medline: [20592046](https://pubmed.ncbi.nlm.nih.gov/20592046/)]
52. Reeves B, Nass C. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Stanford, CA: CSLI Publications; 1996.
53. Wright ER. Enter health information technology: Expanding theories of the doctor-patient relationship for the twenty-first century health care delivery system. In: Pescosolido B, Martin J, McLeod J, Rogers A, editors. Handbook of the Sociology of Health, Illness, and Healing. New York, NY: Springer; 2011:343-359.
54. Georgiou A, Prgomet M, Lymer S, Hordern A, Ridley L, Westbrook J. The impact of a health IT changeover on medical imaging department work processes and turnaround times. Appl Clin Inform 2017 Dec 19;06(03):443-453. [doi: [10.4338/aci-2015-01-ra-0014](https://doi.org/10.4338/aci-2015-01-ra-0014)]
55. IupHealth. URL: <https://iup.health/> [accessed 2020-10-07]
56. Open mHealth. URL: <https://www.openmhealth.org/> [accessed 2020-10-07]
57. Validic. URL: <https://validic.com/solutions/> [accessed 2020-10-07]
58. Seqster. URL: <https://seqster.com/> [accessed 2020-10-07]
59. Human API. URL: <https://www.humanapi.co/> [accessed 2020-10-07]
60. LifeWIRE. URL: <https://lifewiregroup.com/> [accessed 2020-10-07]
61. Conversa. URL: <https://conversahealth.com/> [accessed 2020-10-07]
62. Current Health. URL: <https://currenthealth.com/platform> [accessed 2020-10-07]
63. Eramo L. Do doctors care about your wearable data? Philips. 2017 Oct 18. URL: <https://www.philips.com/a-w/about/news/archive/future-health-index/articles/20171018-doctors-care-wearable-data.html> [accessed 2020-10-07]
64. Notte C, Skolnik N. Where to go with wearables. MDedge. 2018 Oct 31. URL: <https://www.mdedge.com/endocrinology/article/178625/business-medicine/where-go-wearables> [accessed 2020-10-07]
65. Lim D. HIMSS19: Doctors don't know what to do with data from wearables. MedTech Dive. 2019 Feb 15. URL: <https://www.medtechdive.com/news/himss19-doctors-dont-know-what-to-do-with-data-from-wearables/548575/> [accessed 2020-10-07]
66. Coiera E. When conversation is better than computation. J Am Med Inform Assoc 2000;7(3):277-286 [FREE Full text] [doi: [10.1136/jamia.2000.0070277](https://doi.org/10.1136/jamia.2000.0070277)] [Medline: [10833164](https://pubmed.ncbi.nlm.nih.gov/10833164/)]
67. Vergheze A, Shah NH, Harrington RA. What this computer needs is a physician: Humanism and artificial intelligence. JAMA 2018 Jan 02;319(1):19-20. [doi: [10.1001/jama.2017.19198](https://doi.org/10.1001/jama.2017.19198)] [Medline: [29261830](https://pubmed.ncbi.nlm.nih.gov/29261830/)]
68. Israni ST, Vergheze A. Humanizing artificial intelligence. JAMA 2019 Jan 01;321(1):29-30. [doi: [10.1001/jama.2018.19398](https://doi.org/10.1001/jama.2018.19398)] [Medline: [30535297](https://pubmed.ncbi.nlm.nih.gov/30535297/)]
69. Wight D, Wimbush E, Jepson R, Doi L. Six steps in quality intervention development (6SQuID). J Epidemiol Community Health 2016 May;70(5):520-525. [doi: [10.1136/jech-2015-205952](https://doi.org/10.1136/jech-2015-205952)] [Medline: [26573236](https://pubmed.ncbi.nlm.nih.gov/26573236/)]
70. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: [10.2196/jmir.8775](https://doi.org/10.2196/jmir.8775)] [Medline: [29092808](https://pubmed.ncbi.nlm.nih.gov/29092808/)]
71. Hwang J, Christensen CM. Disruptive innovation in health care delivery: A framework for business-model innovation. Health Aff (Millwood) 2008;27(5):1329-1335. [doi: [10.1377/hlthaff.27.5.1329](https://doi.org/10.1377/hlthaff.27.5.1329)] [Medline: [18780919](https://pubmed.ncbi.nlm.nih.gov/18780919/)]

72. Device software functions including mobile medical applications. US Food and Drug Administration. 2019. URL: <https://www.fda.gov/medical-devices/digital-health/device-software-functions-including-mobile-medical-applications> [accessed 2020-10-07]
73. Medical devices: Software applications (apps). Medicines and Healthcare products Regulatory Agency. 2014. URL: <https://www.gov.uk/government/publications/medical-devices-software-applications-apps> [accessed 2020-10-07]
74. Gann DM, Wang Y, Hawkins R. Do regulations encourage innovation? The case of energy efficiency in housing. *Build Res Innov* 2010 Oct 14;26(5):280-296. [doi: [10.1080/096132198369760](https://doi.org/10.1080/096132198369760)]
75. te Kulve H, Boon W, Konrad KT, Schuitmaker TJ. Influencing the direction of innovation processes: The shadow of authorities in demand articulation. *Sci Public Policy* 2018;45(4):467. [doi: [10.1093/scipol/scy015](https://doi.org/10.1093/scipol/scy015)]
76. Stewart LA. The impact of regulation on innovation in the United States: A cross-industry literature review. 2011 Nov 14. URL: <https://itif.org/publications/2011/11/14/impact-regulation-innovation-united-states-cross-industry-literature-review> [accessed 2020-10-07]
77. Proposed policy, payment, and quality provisions changes to the Medicare physician fee schedule for calendar year 2021. 2020 Aug 03. URL: <https://www.cms.gov/newsroom/fact-sheets/proposed-policy-payment-and-quality-provisions-changes-medicare-physician-fee-schedule-calendar-year-4> [accessed 2020-10-07]

## Abbreviations

**AI:** artificial intelligence

**HIT:** health information technology

**IoT:** Internet of Things

*Edited by G Eysenbach; submitted 23.04.20; peer-reviewed by M Aanestad, D Hoffman, C Morrison; comments to author 12.06.20; revised version received 29.08.20; accepted 14.09.20; published 22.10.20.*

*Please cite as:*

Azodo I, Williams R, Sheikh A, Cresswell K

*Opportunities and Challenges Surrounding the Use of Data From Wearable Sensor Devices in Health Care: Qualitative Interview Study*

*J Med Internet Res* 2020;22(10):e19542

URL: <http://www.jmir.org/2020/10/e19542/>

doi: [10.2196/19542](https://doi.org/10.2196/19542)

PMID: [33090107](https://pubmed.ncbi.nlm.nih.gov/33090107/)

©Ijeoma Azodo, Robin Williams, Aziz Sheikh, Kathrin Cresswell. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Review

# Elaborating Models of eHealth Governance: Qualitative Systematic Review

Anne Granstrom Ekeland<sup>1\*</sup>, PhD; Line Helen Linstad<sup>1\*</sup>, MSc

Norwegian Centre for E-health Research, Tromsø, Norway

\* all authors contributed equally

**Corresponding Author:**

Anne Granstrom Ekeland, PhD

Norwegian Centre for E-health Research

P.o. box 35, University Hospital

Tromsø, 9038

Norway

Phone: 47 95266791

Email: [anne.granstrom.ekeland@ehealthresearch.no](mailto:anne.granstrom.ekeland@ehealthresearch.no)

**Related Article:**

This is a corrected version. See correction statement: <https://www.jmir.org/2020/11/e25853>

## Abstract

**Background:** Large-scale national eHealth policy programs have gained attention not only for benefits but also for several unintended consequences and failed expectations. Given the complex and mixed accounts of the results, questions have been raised on how large-scale digitalization programs are governed to reach health policy goals of quality improvement and equal access along with necessary digital transformations. In this qualitative systematic review, we investigate the following question: How is governance implemented and considered in the studies included in the qualitative review?

**Objective:** The aim of this study is to arrive at informed and recognizable conceptualizations and considerations of models of governance connected to eHealth, as presented and discussed in the scientific literature. In turn, we hope our results will help inform the discussion of how to govern such processes to obtain collectively negotiated objectives.

**Methods:** A qualitative systematic review is a method for integrating or comparing with the findings from qualitative studies. It looks for “themes” or “constructs” that lie in or across individual qualitative studies. This type of review produces a narrative synthesis with thematic analysis and includes interpretive conceptual models. The goal is an interpretation and broadens the understanding of a particular phenomenon. We searched the PubMed database using predefined search terms and selected papers published from 2010 onwards. We specified the criteria for selection and quality assessment.

**Results:** The search returned 220 papers. We selected 44 abstracts for full-text reading, and 11 papers were included for full-text synthesis. On the basis of the 11 papers, we constructed four governance models to categorize and conceptualize the findings. The models are political governance, normally depicting top-down processes; medical governance, which normally depicts bottom-up processes; the internet and global model, emphasizing international business strategies coupled with the internet; self-governance, which builds upon the development of the internet and Internet of Things, which has paved the way for personal governance and communication of one’s own health data.

**Conclusions:** Collective negotiations between the nation-state and global policy actors, medical and self-governance actors, and global business and industry actors are essential. Technological affordances represent both positive and negative opportunities concerning the realization of health policy goals, and future studies should scrutinize this dynamic.

(*J Med Internet Res* 2020;22(10):e17214) doi:[10.2196/17214](https://doi.org/10.2196/17214)

**KEYWORDS**

digital transformations in health care; health policy goals; national and international governance models

## Introduction

### Background

Over the last two decades, large-scale digital health programs and services, such as electronic health records (EHRs), have proliferated with equal access, improved quality, and resource optimization as wider health policy goals. National eHealth policy programs have gained attention not only for their benefits, but also for several unintended consequences and failed expectations. Given the mixed, complex results of these programs, questions are increasingly being asked about the executive power of political-governance strategies and how large-scale digitalization programs are governed to achieve the health policy goals of quality improvement and equal access with the necessary digital transformations.

In this study, based on a qualitative systematic review, we take a closer look at how governance has been implemented and considered in the realm of digitalization in health care by scholars in health care and the health sciences. The objective of the review is to arrive at an informed, recognizable conceptualization and consideration of governance. By summarizing existing and emerging models and the ways they are considered, including suggestions for improvement, we analyze the features of each model and how they possibly interfere with each other. In turn, we hope to inform the discussion on improving governance to obtain eHealth policy goals.

In a 2013 editorial in the *British Medical Journal*, the English national program for information technology (IT) was discussed 10 years after its inception [1]. This program promised to revolutionize care in the English National Health Services (NHS) and was originally scheduled to run for 2 years and 9 months starting in April 2003 [2].

The editorial summarized the results of the national program for IT and discussed its flaws and benefits. We draw attention to the following conclusion on governance: “There have been substantial improvements in the technical knowledge base underpinning information systems in the NHS, in organizational capacity to introduce any new IT system, and in information governance processes and procedures. Ten years on, only a handful of hospitals can be described as paperless, and most communication between NHS organizations still occurs by snail mail, fax, or patient messenger” [1].

Governmental opponents of the program described it as a huge disaster, but in the final benefits statement released by the Department of Health, its proponents largely described it as a success. They predicted that by 2022, financial benefits would be £10.69 bn (US \$13.86 bn), outweighing the cost of £9.78 bn (US \$10.09 bn).

Similar accounts of the contested, intended, and unintended results of eHealth programs were described in the United States in 2019: “The US health system has recently achieved widespread adoption of EHR systems, primarily driven by financial incentives provided by the Meaningful Use (MU) program” [3]. Although successful in promoting EHR adoption and use, this program and other contributing factors also

produced important unintended consequences, such as failed expectations, EHR market saturation, innovation vacuum, physician burnout, and data obfuscation, with far-reaching implications for the United States health system [4]. To avoid unintended consequences, these authors proposed improved governance, including efforts from diverse players such as health care providers, administrators, health information technology (HIT) vendors, policymakers, informatics researchers, funding agencies, and outside developers. The authors also argued for the promotion of new business models, collaboration between academic medical centers and informatics research departments, and improved methods for evaluating HIT.

These accounts of mixed, complex results, challenges to the realistic calculation of benefits and costs, and implementations of intentional changes constitute the background of this paper. Perhaps unintended consequences could have been minimized through stronger, better governance of these processes in England and the United States.

At the outset, we understand health care governance as a political process regarding national eHealth goals. National health care services differ between countries and regions, and they are broadly categorized into 2 systems: single-payer and multi-payer systems [5]. National authorities are responsible for governing their respective national health care services, including eHealth governance strategies that are often top-down through legal regulation and reimbursement schemes, where resources are allocated to prioritized service providers. However, political governance can be driven by evidence-based medical imperatives, which are bottom-up strategies [6]. National eHealth policy adjustment in such cases occurs according to new medical evidence.

Governance is extensively discussed in social and political sciences and is one of the most common social science terms [6]. One prominent definition is “an interactive process through which society and the economy are steered toward collectively negotiated objectives” [6]. Political goals are normally the result of compromises obtained through collective negotiations between stakeholders. “Public governance should thus be considered as composite and mixed with inbuilt tensions between competing concerns, actors, ethics, resources, and time horizons” [6,7].

This paper considers eHealth governance as a dynamic process in which development, decisions, implementation, evaluation, and adjustments overlap and interact. The development grows out of conscious and unconscious interdependencies. To realize deliberate policies in such interdependent situations, collective negotiations might be increasingly important. In Norway, the word “co-management” captures this process, which blurs the strict distinction between bottom-up and top-down governance.

One group of researchers has emphasized the need for more empirical studies in the field of governance to strengthen both conceptual and empirical knowledge [8].

We report on the results of a systematic, qualitative review of how eHealth governance has been conducted and experienced. We investigate how governance has been addressed and has played out in eHealth programs as described in the health



sciences literature. We do not aspire to compare different enactments of governance, but instead, perform a descriptive analysis by conceptualizing the characteristics of governance and the experiences presented in the cited papers. We also discuss how different models interfere with each other and summarize suggestions for improvements.

We have restricted our search to PubMed because digital health is a popular emerging topic, and it is of interest to learn how governance is addressed and considered in health sciences. On the basis of political and social sciences, we contribute to a cross-fertilization of scientific traditions by communicating with practitioners, medical internet researchers, and political health authorities.

## Question

How do the papers included in the qualitative review define and consider governance?

## Objectives

The objective is to arrive at informed, recognizable conceptualizations and considerations of governance models of eHealth, as presented and discussed in the scientific literature. In turn, we hope our results will help inform the discussion on how to govern such processes to obtain collectively negotiated objectives.

## Outline

The remainder of the paper contains an account of the methods used, quantitative and qualitative results, and a discussion section in which we distinguish between and elaborate on models based on our constructs from interpreting the empirical results of the cited papers. We discuss suggestions for governance improvements, present our conclusions, and point to areas for further research. We also discuss the limitations of this review. Due to the coronavirus pandemic, we have provided a postscript that connects our models and conclusions to possible (eHealth) governance in the COVID-19 era.

## Methods

### Qualitative Systematic Review

A qualitative systematic review is a method for integrating or comparing with the findings from qualitative studies. It looks for themes or constructs that lie in or across individual qualitative studies and may employ selective or purposive sampling. This type of review produces a narrative synthesis with a thematic analysis and includes interpretive, conceptual models. The accumulated knowledge resulting from this process may lead to the development of a new theory, an overarching narrative, a wider generalization, or an interpretative translation. The goal is not aggregative in the sense of adding studies together, as with a meta-analysis; rather, it is interpretative and broadens the understanding of a particular phenomenon [9].

### Search Strategy and Information Sources

Our search was performed in November 2018 using the previous version of PubMed. PubMed was updated in the spring of 2020.

We searched “All fields” with the following search criteria:

Search (“governing reforms” OR (telemedicine governance) OR (governance “national telemedicine programme”) OR (governance “national ehealth programme”) OR (governance “national ehealth programme”) OR (governance innovation ICT) OR (“innovative procurement” health\*) OR (“innovative procurement” health) OR (“whole system demonstrator programme” lessons) OR (Sundhedsplatformen) OR (Governance “regional EHR implementation” Denmark)).

The search was restricted to scientific papers, including systematic reviews published in peer-reviewed journals. A systematic review was defined as an overview with an explicit question and a methods section with a clear description of the search strategy and methods used to produce the review. The review was also expected to report on and analyze empirical data. As many papers were retrieved (220), we included only reviews and papers published from 2010 onwards in the final review for pragmatic reasons.

### Inclusion Criteria

#### *Population and Participants*

We included papers on the governance of national, regional, or global eHealth programs focusing on population, public health, hospitals, communities, patients, consumers, health professionals, and family caregivers, regardless of diagnoses or conditions.

#### *Interventions and Issues*

The review included governance issues connected to all eHealth interventions, information and communication technologies (ICT) for communication in health care, and internet-based interventions for diagnosis, self-management, and treatments. Social care was considered relevant if it was an important part of health care and occurred in collaboration with health care for patients with chronic conditions.

#### *Comparisons*

Papers that compared governance of eHealth and governance of standard care or other types of care were included in the review.

#### *Outcomes*

Only papers and reviews that reported relevant outcomes were included. Relevant outcomes were specified as eHealth governance with health-related outcomes (morbidity, mortality, quality of life, and patient satisfaction), process outcomes (quality of care, professional practice, adherence to recommended practice, professional satisfaction, governance strategies, organizational aspects, policy and implementation), and costs or resource use. Systematic reviews and papers reporting on emerging issues, such as unexpected findings or important new insights, were also included.

#### *Languages*

Articles published in English, French, or a Scandinavian language were included.

### Exclusion Criteria

The governance of general health reforms, innovation programs, public reforms, and innovation programs was excluded, as were



conference papers, dissertations, proceedings, and irretrievable papers.

### **Design**

Papers and reviews considered nonsystematic or nonrigorous, such as commentaries, editorials, and proceedings, were excluded, as were systematic reviews with major limitations (low quality) according to a revised checklist for systematic reviews from the Cochrane Effective Practice and Organisation of Care Group. If the same authors had produced several publications of the same review, the most updated and/or the fullest review was selected while other versions were excluded.

### **Participants**

Studies with participants considered irrelevant to the review were excluded, such as studies on the use of ICT outside the health care domain. Animal studies were also excluded.

### **Interventions Considered Irrelevant to the Review**

Studies on the governance of interventions considered irrelevant to the review included internet-based education of students and health professionals, medical technology in clinical practice in general, such as medical and surgical examinations and treatments based on computer technologies, except when used in remote diagnosis and treatment (telehealth); the use of telephones (including cell phones) only; eHealth as only a limited part of an intervention; and the use of the internet for surveys, research, web-based prescriptions, mass-media interventions, and veterinary medicine.

### **Outcomes**

Articles without relevant outcomes were excluded, that is, those not meeting the inclusion criteria.

### **Study Selection**

Articles were stored in the free software Rayyan, where we inserted the selection criteria and read the abstracts [10]. On the

basis of the inclusion and exclusion criteria, two reviewers independently screened the list of the titles and abstracts from the literature searches and identified potentially relevant studies, the full-texts of which were then retrieved. We resolved any disagreement through dialogue between the two reviewers based on the selection criteria. The citations were then exported to EndNote.

### **Data-Collection Process**

Data collection was carried out using a web-based data-extraction form created by the authors in Google Docs, which is enclosed in Appendix 1. The 2 authors collected data based on full-text papers. The following quality domains were assessed to identify, include, and critically appraise the studies: explicit accounts of data-collection methods, explicit accounts of methods used to analyze the findings, and overall assessments of the qualities of the papers and reviews.

### **Analytical Perspective**

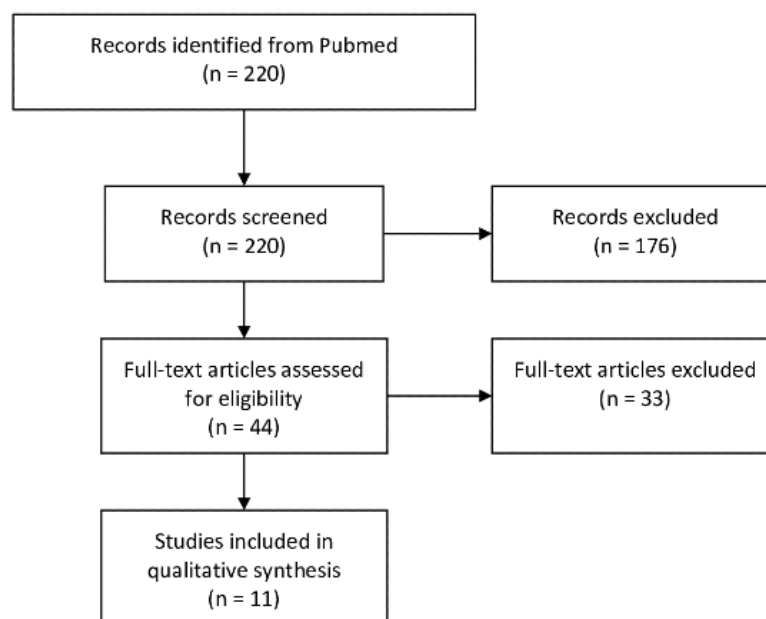
We analyzed the papers from a grounded perspective by interpreting the qualitative data in the included papers. Grounded theory is a form of empiricism that emphasizes inductive reasoning and hypothesis generation, in contrast to the hypothetic co-deductive model of the positivist scientific method [11]. Our interpreted models can be used as working hypotheses for further empirical development and research.

## **Results**

In total, we included 44 of the 220 abstracts extracted from PubMed.

### **Quantitative Results**

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram below (Figure 1) presents the quantitative results [12].

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

### Qualitative Results: Governance Models

In the background section, we briefly indicated 2 broad governance strategies: political governance, which normally depicts top-down processes and medical governance, which normally depicts bottom-up processes. On reviewing the 11 papers, we found it difficult to subsume all of them under either political or medical strategies.

Therefore, we constructed four governance models to categorize and conceptualize our findings. We chose to construct models because “Models seek to simplify phenomena as an aid to conceptualisation and explanation” [13]. They are also tools to realize a narrative synthesis, which includes thematic analysis and interpretive, conceptual models.

The 4 models are: political governance, medical governance, internet and global business governance, and self-governance. We briefly describe the models below before sorting the different papers into them.

1. **Political governance:** Health policy and implementation are nation-state responsibilities. However, nation-states are also members of the World Health Organization, which has a global perspective on governance. Regional or local authorities are also responsible for collectively negotiating goals. National authorities can choose top-down governance approaches or delegate responsibility to the regional and local levels to facilitate bottom-up strategies. Some papers described this up-and-down process as a national “middle-out” approach to eHealth decision-making. We have conceptualized all these strategies as political strategies, implying that the top-down perspective is broad and dynamic, as our background section indicates.
2. **Medical governance:** The medical model implies that governance initiatives and executive agencies rest with medical and health care professionals and professional

organizations, which provide the terms of development. This model also comes with dynamic adaptations, such as adjustments to the political model. In radiology, for instance, which was fully digitalized at an early stage, radiologists saw its national medical potential and consequently influenced political strategies.

3. **Internet and global business governance:** In this category, we include governance mechanisms by nongovernmental organizations, the health care industry, and private internet companies. The structure of the internet facilitates new business models, which may challenge political strategies and national control. The internet is accessible to almost everyone, and no international quality assurer exists, which prohibits the diffusion of information or services that might contradict medical evidence and challenge public health. Illicit drug sales on the internet is an example of a new business model.
4. **Self-governance:** This model builds on the development of the internet and the Internet of Things (IoT), which has paved the way for the personal governance of health by storing, developing, and communicating one’s own health data. This model considers eHHealth to be developed by individuals’ efforts to optimize their own health. The self-management of diabetes through apps, for instance, may facilitate individual governance (by patients and citizens) of eHealth services through demand for new apps and simultaneously empower patients and strengthen their digital literacy.

### The Political-Governance Model

The following 4 papers align with the political-governance model: Atalag addresses a single content model for eHealth interoperability and secondary use in New Zealand [14]; Park and Atalag address the national approach to health care ICT standardization and focus on progress in New Zealand [15];

Kierkegaard addresses interoperability after EHR deployment and describes persistent challenges and how the Danish government applied a combination of bottom-up and top-down strategies to realize full-scale EHR implementation [16]; de Riel et al [17] address success factors for implementing and sustaining a mature electronic medical record, iSante, in a low-resource setting in Haiti (Table 1).

In the political-governance model, eHealth is governed top-down through a national eHealth strategy. The papers included in this model describe national top-down strategies on technical infrastructure, legal frameworks, and institutional structures,

such as national advisory boards or a multi-editorial expert group [15,16] (Table 1).

Atalag describes the national top-down strategy of New Zealand as a “middle-out transitional approach to achieving semantic interoperability in eHealth” [14]. Kierkegaard describes how the Danish authorities combined top-down and bottom-up approaches to realize national eHealth goals: “Changes in the organizational setup and redistribution of responsibilities between the Danish regions and the state play a pivotal role in producing viable and coherent solutions in a timely manner” [16].

**Table 1.** Included papers, governance models, and strategies.

Reference	Country	1: Political governance	2: Medical governance	3: Global internet governance	4: Self-governance
Atalag K, 2013 [14]	New Zealand	Single content model; top-down strategy – framework for standards	N/A <sup>a</sup>	N/A	N/A
Park YT, 2015 [15]	New Zealand	National Health Board and National Health IT Board; top-down strategy	N/A	N/A	N/A
De Riel E, 2018 [17]	Haiti	The Haitian Ministry of Health Management Unit; top-down strategy for a national electronic medical record model	N/A	N/A	N/A
Kierkegaard P, 2015 [16]	Denmark	The National eHealth Authority; top-down strategy – frameworks for standards	N/A	N/A	N/A
Wade, VA 2012 [18]	Australia	Governmental rebates for medical specialists using telehealth No general IT for health top-down strategy	Telehealth as a game-changer in clinical practice in ethical, legal and medical governance aspects	N/A	N/A
Crocker M, 2010 [19]	United Kingdom	N/A	Neuroscience centers organized as a clinically driven tertiary referral service; bottom-up strategy from the medical society (radiology)	N/A	N/A
Sutton LN, 2011 [20]	United Kingdom	N/A	National PACS <sup>b</sup> part of the National Program for IT after reacting to bottom-up strategy on realizing benefits	N/A	N/A
Bagot KL, 2017 [21]	Australia and United Kingdom	N/A	Creating sustainable medical networks in stroke care; cultural differences in medical governance; bottom-up strategy	N/A	N/A
Mackey TK, 2014 (A call for a moratorium) [22]	Global	N/A	N/A	Internet Corporation for Assigned Names and Numbers, nonprofit organization; bidding process and dot-health (health)-industry business strategy	N/A
Mackey TK, 2014 [23]	Global	N/A	N/A	Internet Corporation for Assigned Names and Numbers chose the International Chamber of Commerce to adjudicate dot-health concerns	N/A
Williams SJ, 2015 [24]	Global	N/A	N/A	N/A	Self-management and web-based community member outside national health services; self-health-optimizing strategy

<sup>a</sup>N/A: not applicable.<sup>b</sup>PACS: Picture Archiving and Communication System

### The Medical-Governance Model

The following 4 papers align with this model: Wade et al [18] presented a qualitative study of ethical, medico-legal, and clinical governance matters in Australian telehealth services; Sutton considers a Picture Archive Communication System and diagnostic imaging service delivery from the UK perspective [20]; Bagot et al [21] performed a qualitative analysis comparing with the experience of Australia and the United Kingdom in integrating acute stroke telemedicine consultations into specialists' "usual practice"; and Crocker et al [19] address patient safety and image transfer between referring hospitals and neuroscience centers, exploring possible methods of improvement.

The bottom-up strategy may grow out of local needs [18,20], and sometimes it is picked up and transformed into a national top-down strategy. For instance, Crocker et al [19] claim that "Part of the remit of the National Programme for Information Technology in the NHS (NPfIT) was to improve neuroscience teleradiology." The researchers wanted to evaluate whether their experiences were part of the national top-down strategy; so the paper describes how the medical society experienced their recommendations presented in an NHS report from 2004 being "largely ignored" in the national program. National top-down strategies do not necessarily align with bottom-up strategies [19].

Wade et al [18] state that, from a medical point of view (bottom-up), concerns about data security in telemedicine need to be governed through national top-down guidelines.

### The Internet and Global Business–Governance Model

The following 2 papers align with this model: Mackey et al [23] addressed health domains for sale and argued for the need for global political health governance on the internet, and Mackey et al [23] discussed a call for a moratorium on the top-level domain "generic health" to prevent the commercialization and exclusive control of web-based health information [22].

These authors argue that the national jurisdiction of privacy and data protection is largely ignored and challenges existing international governmental organizations, such as the World Health Organization and European Union [7]. According to the authors, the business interests of large actors and market mechanisms overrule both political and medical governance.

Mackey et al [22] claim that the processes of the Internet Corporation for Assigned Names and Numbers (ICANN) "appear to favour business interests and generation of profits over the future integrity of the Health Internet, failing to make any tangible commitments to protect public health or enforce norms as would be found in a responsible global governance Framework."

However, they also state that "Direct-to-consumer advertising of prescription products is not allowed in the vast majority of countries other than the United States and New Zealand, and it may be unlawful for pharmaceutical manufacturers to engage in multijurisdictional web-based advertising that could occur through future gTLDs (generic top-level domain name)" [23]. There is a lack of global governance in health that may challenge human rights in public health [23].

### The Self-Governance Model

We have included one paper in this category by Williams et al [24], who discussed governance in terms of individual agency as opposed to political governance. They also envision the opposite of individual agency: the ways in which digitalization and the use of technological affordances change individuals and identities. The paper describes a complex model, a dynamic dualism between the governance of individuals by eHealth versus the self-governance of eHealth.

We have presented the included papers under 4 governance models: political governance, which is normally limited to nation-state jurisdiction, international, evidence-based medicine and health care guidelines, large-scale industries governed by business goals, and self-governance by consumers.

### Summary of Empirical Results

Table 1 lists the 11 papers by author, year published, country in which the governance model is discussed, and model or strategy. Some papers do not refer to a specific country because they describe global models or models related to the consumers' use of the IoT products and services.

Table 2 lists the included papers by author, year published, country in which the characteristics, experiences and challenges are described, and the suggested improvements related to a specific governance strategy. Three papers do not relate to a specific country but describe aspects related to global governance and self-governance models.



**Table 2.** Included papers, considerations, and emerging governance trends.

Reference	Country	Experience and performance (descriptive)	Considerations and suggested improvements (descriptive)	Trends and emerging governance models
Atalag K, 2013 [14]	New Zealand	Suggesting a “middle out” approach; a single top-down approach does not produce the expected outcome; success relies on a relationship based on trust between the authorities and the medical society	The national Single Content Model is flexible and “enables smooth transition to a comprehensive solution through gradual replacement over time”	The openEHR trend and Archetypes; governance based upon a top-down strategy for interoperability
Park YT, 2015 [15]	New Zealand	Strong top-down governance strategy and organizational structures and rules has been the most important factors for successful eHealth governance	Next steps will be to analyze how the eHealth structure influence health outcomes and minimize errors	Strong belief in strong government involvement leads to successful goal attainment in eHealth
De Riel E, 2018 [17]	Haiti	A key learning is leadership engagement to create an understanding of how the system works	Haiti’s health care service is dependent on national and third-party funding; realizing a sustainable information and technology communications infrastructure is a muddle-through process	System migration to the OpenMRS platform to take advantage of a global community and ensure sustainability
Kierkegaard P, 2015 [16]	Denmark	Setting national goals and adapting the middle-out approach as part of a national strategy is the best way to realize full-scale implementation of electronic health records, an approach based on cross-regional coordination	A collective phase-out of all systems may be costly, but it may be the only way to create a common national platform of high interoperability; risk of regions working with one vendor, which creates regional dependency on an actor with market monopoly	The Danish framework is flexible and in line with the European Union eHealth Interoperability framework
Wade VA, 2012 [18]	Australia	This study indicates that telehealth can be a tool to realize medical quality of care governed bottom-up and in line with evidence-based medicine	The medico-legal aspects did not seem to be as difficult as anticipated; national reimbursement schemes may increase substantial system benefits	N/A <sup>a</sup>
Crocker M, 2010 [19]	United Kingdom	Image transfer is delayed by an immature technology infrastructure	Bottom-up recommendations from the teleradiology society in United Kingdom has been ignored by the top-down process	N/A
Sutton LN, 2011 [20]	United Kingdom	Creation of a UK reporting “Grid” with remaining organizational and governance challenges	Telehealth challenges around “team organisation” when the reporter and doctor are in separate organizations at a distance	Creation of a medical governance model in radiology across the United Kingdom which harmonize with the national top-down strategy of National Health Services
Bagot KL, 2017 [21]	Australia and United Kingdom	UK model: telemedicine integrated in the specialist work plan; Australia telemedicine was ad hoc	Both networks see telemedicine as part of future organization; can reduce workload by a “follow-the-sun model”	N/A
Mackey TK, 2014: (A call for a moratorium) [22]	Global	The governance of health-related internet domains should be run by international organizations and not by a for-profit company like Internet Corporation for Assigned Names and Numbers, which does not respond to actors who want this policy change	The international society should stop the Internet Corporation for Assigned Names and Numbers governance model	Is it feasible to realize a global internet governance model based on international interstate cooperation?
Mackey TK, 2014 [23]	Global	The governance model of dot-health and other internet health domains is not run by the World Health Organization or another international health organization, which may represent a global threat to public health: access to evidence-based medicine and quality assured health information	The main challenge is to realize a shift from a privately-run model to an international model governed by legitimate international public health actors	The combination of internet and global markets challenges political and medical governance

Reference	Country	Experience and performance (descriptive)	Considerations and suggested improvements (descriptive)	Trends and emerging governance models
Williams, SJ, 2015 [24]	Global	Health apps are a growing trend that realizes “the quantified self”: blurring the lines between health care and wellness through data sharing	N/A	“Quantified-self” apps contribute to a growth of health and wellness data; a “global health and wellness data governance model”?

<sup>a</sup>N/A: not applicable.

## Discussion

### Elaborating Models, Experiences and Suggestions for Improvement

This section focuses on the features and considerations of the different models, the ways in which they interfere with each other, and suggestions for improved governance.

### Features and Considerations of the Different Models

#### *Features and Considerations of Political Governance*

The 4 papers in this category describe top-down governance models and discuss how top-down national policy has used regional eHealth models as tools to implement wider national health strategies through a defined framework and governmentally assigned top-level working groups.

When, for instance, the Danish government wanted to speed up EHR implementation, they focused on regional governance to replace their former top-down strategies. Likewise, when the government found it more suitable to decrease the number of existing EHR systems, they strengthened the top-down national model by using regional middle-out strategies [16]. The government encouraged cross-regional coordination as part of the national eHealth strategy (Table 2).

The history of Denmark’s health care system demonstrates the inherent difficulties of a state-centric approach to technology harmonization. Similar results have been obtained when other countries applied a top-down approach to national-level shared medical records, such as the ambiguities in the NHS National Programme for IT (NPfIT). Moreover, Denmark’s methods for overcoming the interoperability issues between EHRs demonstrated that health information exchange is not merely a technical issue, “but a challenge fraught with organisational and political complexities” [16].

In New Zealand, the government governs the top-down implementation of eHealth using a national eHealth strategy and standards. Atalag argued that it was difficult to achieve a high level of interoperability through a top-down strategy, partly because of resistance from different organizational actors. The model was therefore replaced by “a middle-out (transitional) approach to achieve semantic interoperability in eHealth” [14]. This approach was strategically used temporarily until the top-down approach was resumed (Table 2).

The Haitian model included three multi-site electronic medical records as the cornerstone of the broader planned national eHealth architecture, eventually feeding into an overarching system to aggregate health-indicator reporting (called *Système d’Information Sanitaire National Unique*). The Haitian Ministry of Health set this model as a priority in 2013 [17], a middle-out

approach that included both the top-down governmental strategy, the bottom-up medical strategy, and the donor strategy as a third actor. This middle-out approach helped realize health policy goals related to sustainable health service.

The case from Haiti shows that a top-down strategy may apply in low-resource, donor-dependent settings. “Strong leadership was essential to system continuity and expansion. In addition, many developing countries may not have a legal framework that addresses management or protection of health information” (Table 2).

In New Zealand, Atalag emphasized that “the success of the described approach relies heavily on appropriate governance, and it is imperative to put in place new models of collaboration” [14]. The government of New Zealand, unlike those of other countries, established specific HIT bodies to drive and support standardization, including the National Health IT Board and the Health Sector Architects Group, which actively participated in the standardization processes (Table 1). They collected ideas from the health care industry and the public sector and developed HIT standards. In the second paper, Park and Atalag further emphasize that New Zealand’s governmental structures and processes may be a direct and efficient way of achieving the benefits of HIT standardization because government power has a strong influence on markets, and many health care organizations are under government control [15].

In summary, political-governance models encompass bottom-up, top-down, and middle-out approaches. We chose to label them top-down however, because even bottom-up strategies are a part of top-down governance since health care is a nation-state responsibility. These strategies seem to fluctuate dynamically within the nation-state, following a thesis-and-antithesis curve. Bottom-up and middle-out strategies seem to be shifting to improve top-down governance incrementally and instrumentally. The models enact governance *through* and *of* eHealth: influential human actors use eHealth programs as tools, and shifting strategies are applied to obtain wider health political goals.

#### *Features and Considerations of Medical Governance*

The political-governance model’s top-down focus on governance by standards, legal frameworks, and national boards may lead to an underestimation of other important aspects affecting governance, such as knowledge, identity, roles, cultures, trust, quality, and asymmetric power relations.

Medical governance may shed light on these aspects of governance in eHealth. Medical governance is enacted within medical networks by medical and health care professionals, who see potential in technology and create networks to maximize its benefits and improve medical practice. “Clinical governance is a systematic approach to implementing quality

and safety in health care, which aims to maximize evidence-based practice and reduce risk” [18]. In this paper, Wade et al [18] suggest that various examples have shown that telemedicine can be used to enforce medical governance as a driving force to support the uptake of evidence-based care because it has been difficult to change organizational cultures in health care systems. The literature on the organizational effects of telehealth has been sparse and has focused on the details of implementation. Wade et al [18] further comment on governance within eHealth communities: “Telehealth could be a key factor in quality improvement as it produces immediate contact between providers, is based on the management of real patients and promotes trust in interprofessional relationships” (Table 2).

Bagot et al [21] argue that clinicians must adapt to this new way of delivering services; adaptation subsequently affects the trust and roles and responsibilities between organizations that collaborate on eHealth governance: “Successful telemedicine networks require specialists adapting clinical practice to provide remote consultations.” The authors indicate a knowledge gap with respect to medical governance between different countries (Table 2).

Sutton discusses experiences in medical governance as part of the development of telemedicine networks: “It is important for local health care communities and their patients to ensure teleradiology does not destabilize or de-skill smaller departments. Teleradiology should be complementary and not an alternative to progressing development of the service locally by enhancing the expertise of the local radiology workforce” [20].

Radiology in England has undergone reorganization due to digitalization. Currently, following the merger of clusters into three main areas across England, there are three major Picture Archiving and Communication System providers with associated storage archives [20]. This process includes the medical society bottom-up: “The Radiology Service Improvement Team of the NHS . . . is now part of the NHS Institute for Innovation and Improvement” [20] (Table 2).

Other authors also note that the medical society opined that the teleradiology recommendations of the 2004 NHS Modernization Agency Neuroscience Critical Care Report have been largely ignored, even though “part of the remit of the National Program for Information Technology in the NHS (NPfIT) was to improve neuroscience teleradiology” [19].

The included papers demonstrate that implementing telemedicine services requires inter- and intraorganizational cooperation and raises questions that transcend standards, infrastructures, and legal frameworks. These include implementation challenges, clinicians’ trust, and the associated uncertainty around how best to ensure stable access to skilled personnel at local departments. There is also uncertainty about clinical accountability and responsibility. Clear governance of these aspects is considered crucial to clinicians.

Including high-ranking medical professionals in the implementation process of national programs, such as the NPfIT, is suggested to increase clinicians’ trust in national

implementation. In cases where the medical society considered their input largely ignored, trust was lost in the top-down strategy. However, these papers do not describe how the implementation process is affected by the lack of trust from the medical society.

Other scholars who have evaluated national EHR implementation programs point out similar results: Strategic, organizational, and human challenges are usually more difficult to master than technical aspects” [25].

In summary, the medical model addresses governance within and by professional eHealth communities. Medical governance represents the bottom-up and middle-out perspectives. The papers describe how medical governance systematically outlines how digitalization maximizes evidence-based practice, reduces risk, and thereby increases quality and access to care. Trust, responsibility, role definition, and skill development are considered crucial to obtaining evidence-based results. This model also faces ongoing challenges that require serious effort to succeed.

### ***Features and Considerations of Internet and Global Business Governance***

In the governance literature, the term “organized anarchy” is used to describe a “loosely coupled” organized structure, which refers to a “relatively open and unspecialized” structure [7]. We interpret this model to denote the internet as a technology with affordances that facilitate organized anarchy. It transcends the jurisdictions and economies of the nation-state and allows global discovery and innovation. Technology facilitates the realization of this anarchy, opening the possibilities for actors to create organizations and services beyond national jurisdiction. eHealth can be part of this global development.

The 2 papers included describe how the internet, which consists of a hierarchical domain-naming system for the internet protocol addresses of computers, services, and other digital resources, relies on domain names as an easily recognizable way for users to search and navigate web-based content. “The Internet Corporation for Assigned Names and Numbers (ICANN), a nonprofit corporation founded in 1998 that controls this system, is currently undergoing the largest expansion of the internet in history” [23].

ICANN is a nongovernmental organization established by the United States government but is notionally independent of it. It manages this hierarchical naming system and roughly 500 accredited domain-name registrars. “ICANN relies on an international Board of Directors consisting of various ICANN constituents, a CEO, staff, and advisory committees consisting of stakeholders from national governments, Internet technical experts, and Internet organizations to inform its decisions” [22].

The authors of the two papers are critical of the ways ICANN governs, noting that it has created “the largest expansion of the Internet in history . . . adding over a thousand new generic top-level domain names (gTLD), potentially including a new .health domain and close to 20 other gTLDs related to medicine and health” [23]. They consider that ICANN’s complex, highly political process of awarding health-related gTLDs profoundly impacts information privacy, use, and sale as well as health

marketing and content quality, which could influence future trust, security, and credibility of the Health Internet. Hence, they argue that it is critical that applicants are carefully scrutinized to ensure that they are abiding by ethical principles, practices, and rules with respect to public health and the public interest.

More than 100,000 health-related websites are estimated to exist, and internet users may have difficulty accessing evidence-based sources and often seek information through simple search-engine queries (eg, Google, Yahoo, and Bing) that may prioritize sites of lower quality, undisclosed commercially sponsored content, irrelevant information, and/or misinformation. “For example, illicit online pharmacies have been detected illegally marketing and selling pharmaceuticals without prescriptions, misrepresenting crucial risk information, and not disclosing other risks of their often counterfeit and otherwise dangerous products” [23]. “Because it governs the majority of the domain name system, ICANN bears great responsibility for those standards, and how the Internet can be used to help or harm individual users” [22].

Objections to ICANN’s governance model have come from the World Health Organization, scholars, and international public health organizations. In 2000, the World Health Organization and other stakeholders proposed the formation of health top-level domain names (TLD), but ultimately their proposal was not chosen as one of the seven proof-of-concept names for new TLDs in that round [23]. The authors underline that ICANN lacks “enforceability, because it has no appeal process to take proactive action against websites that violate laws that accredited registrars fail to report. Consequently, many websites feature illicit online content with clear public health and patient safety concerns that registrars take no action against, such as websites selling medicines without a prescription and that also potentially traffic counterfeit or falsified medicines” [22].

The authors give reasons to worry that the situation will be exacerbated by simply awarding new health domains to the highest bidders. “ICANN’s processes appear to favor business interests and generation of profits over the future integrity of the Health Internet, failing to make any tangible commitments to protect public health or enforce norms that would be found in a responsible global governance framework” [22].

According to the United Nations, health is a fundamental human right. More people than ever are using the Health Internet to seek information and make behavioral choices. “Now is not the time either to compromise this legal right or complicate the factual reality, in favor of profit-making interests merely for the sake of unlimited Internet expansion” [22].

The authors argue that governing health-related internet domains should be a priority for international public health organizations as well as global IT organizations, such as ICANN, the World Health Organization, the International Telecommunication Union, the World Summit on the Information Society and its Internet Governance Forum. ICANN, as a nongovernment organization with a global autonomous governance structure, does not seem to make any tangible commitments to protect public health or enforce norms, as would be found in a responsible global governance framework [22].

In summary, this model exceeds the political and medical governance models and puts international business actors in a dominant role. The authors envision a rather dystopian future, especially for the governance of health data and services on the internet, and they argue for action by global health authorities.

### ***Features and Considerations of Self-Governance Models***

We have included one paper in this category. It describes how sleep-monitoring and diagnosing sleep problems and treatment were moved from medical governance to a self-governance model. This “health service” takes place outside health care institutions and is facilitated by the IoT, which may exist at the global level beyond nation-state jurisdictions. In this model, nongovernmental or private actors may deliver services outside the national health system, as discussed in the previous chapter.

The difference between these devices and previous portable devices is that the user–technology relationships configured by new digital sleep-monitoring technologies are primarily between sleepers and devices. “The information about sleep feeds directly back to the user, providing sleepers with new knowledge about their dormant (or not-so-dormant) body/self; knowledge that itself is imbued with a sense of responsibility for them to act to improve their sleep” [24]. The authors argue that sleep, or the sleeping body or self, is yet another site for improvement or optimization in terms of performance, and that health is important well beyond the clinical sphere.

Further, they point out a certain seductive power in tracking, monitoring, and managing ourselves in the interest of self-knowledge, or even self-governance, for self-improvement or self-optimization.

Referring to Deleuze, the authors warn against forms of more or less continuous control [23,26]. Further, they envision submission and postpanoptical surveillance, referring to Massumi, Foucault, and Bauman [27–29].

In summary, self-governance enacts a dynamic between global internet governance that forces or nudges humans, on the one hand, to adhere to technologies and connected health businesses, and on the other, to rely on them as sources of individual self-governance and control. Governance is characterized by a shifting, dynamic tension: governance of the individual by eHealth (submission) versus self-governance within eHealth services, involving empowered, conscious citizens.

### ***The Ways the Models Interfere With Each Other***

Wade et al [18] argue that medical governance needs help from political governance to regulate responsibility, ethics, and accountability by stating that it is unclear whether responsibility should rest with the primary clinician or be divided between the local clinician, the distant clinician, and the technology provider. They substantiate this by asserting that this responsibility is not regulated in Australian law, noting that there is “no case law relevant to telehealth in Australia” [18]. In addition, they expect this to change: “These matters have not been resolved in Australia, although with the advent of universal telehealth rebates, standards for practice are under development by the professional colleges” [18].



Top-down political-governance strategies are used by national authorities to ensure national frameworks of standards and legal aspects, and medical governance seems to be used as a bottom-up model and strategy to ensure that the process is anchored and incorporated into medical practice. In the cases in Denmark and New Zealand (Tables 1 and 2), both governance models were expected to work in tandem. When the tension between them grew or the governments became aware of possible tensions that could slow the process, they used a middle-out approach to increase the legitimacy of the top-down eHealth strategy.

The middle-out approach is often defined by national advisory boards, which include multiple stakeholders. This was the case for Denmark, New Zealand, and Haiti, which recognized tensions between stakeholders and consequently included medical professionals, vendors, patients, and bureaucrats in their eHealth governance models to facilitate and enforce negotiations. eHealth governance is recognized as an arena for collective negotiations to co-manage tensions to realize national health policy goals.

The global internet and business model describes challenges concerning forces that are complex, ubiquitous, unpredictable, and ungovernable by established policy structures and strategies. This model defines governance in terms of challenges in the governance of health-related data and digital services. Such challenges occur because of interactions between technological affordances and international business actors.

According to the 2 papers on ICANN, the internet and global model interferes with the other 3 models because the ICANN overrules World Health Organization efforts.

The political models were based on the abilities of nation-states to regulate and govern eHealth policy alone or through international cooperation in international governmental organizations or supranational organizations. The anarchical structure of the internet adds a challenge to the political and medical governance of eHealth; however, combined with the internet's global outreach, it facilitates business growth and reduces political and medical governance, leaving it up to the consumer to validate services and products. The European Union passed the General Data Protection Regulation to precisely address this challenge and protect consumers [30].

Self-governance may fit into wider patterns of voluntary and involuntary submission, particularly in the digital era of big data, which Williams et al [24] call "creeping forms of monitoring and surveillance that seem to characterize our lives today within and beyond the medical and health domains."

Such submissions may also be used by global web-based businesses to target services to specific user groups and control markets. Thus, there is a tension between self-governance and global internet and business governance concerning control and manipulation.

### Elaborating Suggestions for Governance Improvements

Scholars in the political-governance category articulate the need for more documentation of the successful, locally-led governance of donor-funded systems, including any

capacity-building for local responsible entities and joint system design, planning, and implementation [17]. Experiences with political strategies point to a demand for ongoing efforts to improve them, with stronger political executive power building on medical relevance to achieve wider goals.

In the medical-governance category, scholars point out the need for continued monitoring of a more robust image-routing system across the NHS and audits of its functionality by the Connecting for Health Safety Team and the National Patient Safety Authority to combine bottom-up and top-down strategies [19].

In support of this argument, Sutton claims that "There is much hope that despite the current economic climate, the established IT programmes in the UK are able to continue to facilitate the development of new solutions. In 2010, the Department of Health in England adopted the 'Quality, Innovation, Productivity and Protection' agenda to maximize the benefits of existing IT systems and promote improved patient care and productivity" [20].

In a paper concerning ethical and legal matters in Australian and UK telehealth, the authors reported that Conducting organizational case studies would give a deeper understanding of the matters identified, particularly those of governance and system change" [18]. In their comparison of the integration of acute stroke consultations and specialists' usual practice, the authors suggest that "Future research might investigate the transferability of UK and Australian experiences to broader European, Asian and American networks" [21]. In general, medical culture is similar across national borders.

The medical society is global and agrees on evidence-based medicine being the core driver of medical diagnoses and treatment procedures. However, integrating telemedicine into care requires juggling international medical culture and local cultural variations in medical delivery. Medical-governance strategies need to embed evidence-based medicine and telemedicine into different organizational set-ups. Incremental "juggling" is considered useful for solving organizational challenges.

Regarding internet and global business governance, the authors of the two reviewed papers suggest that "Focusing on the public good can be a first and crucial step to ensure an accurate, reliable, and evidence-based online presence for health for this generation and the next" [23]. They recommend that the "Internet community needs to be vigilant to ensure the reliability and trustworthiness of health information online, and take immediate action to secure the future integrity and proper governance of this important namespace for the health internet" [22]. The authors hope to encourage ICANN to appoint the World Health Organization and a multitude of governmental and health nongovernment organizations as sponsors. This presupposes that the World Health Organization has executive governance power with the ability to overrule the global internet market and the political governance within each member state. They express a need for political governance to dominate global business strategies.

The papers propose solutions for "future e-Health governance approaches to ensure the appropriate management of health



could be accomplished by requesting ICANN to recategorize health as a sponsored gTLD and proactively appoint WHO as its sponsors” [23]. They suggest that the World Health Organization should develop policies to ensure accountability and transparency in gTLD operations that meet the best interests of the global health community and enforce eligibility rules regarding all future health registrants.

They propose that the World Health Organization’s possible appointment as a gTLD sponsor should be governed by a diverse, globally representative board of health stakeholders in partnership with responsible internet service providers. “This governance mechanism can have representation and be organized into subject-specific advisory panels to review and recommend content to be included on for health” [23].

In the self-governance category, consciousness and data activism is important, which resonates with an argument from Alan Peterson, who calls for “algorithmic accountability” which highlights the potential for citizens to create alternative futures—ones oriented to fulfilling human needs rather than techno-utopian visions” [31].

### Limitations of the Review

By limiting our inclusion criteria to 2010 onwards, we might have missed important papers that could enhance the review. Qualitative reviews are not all-embracing, and the results are colored by the search criteria. Nevertheless, our paper provides valuable insights into various governance strategies at play in the realm of eHealth and valuable suggestions for further work to improve governance and scientific knowledge.

### Postscript – Governance and the Coronavirus Pandemic

The current coronavirus pandemic might provide an excellent illustration of our findings and main conclusions.

To govern and control the pandemic, collective, transparent efforts are in demand: political decisions followed by financial incitements, firm global medical knowledge and advice, international business actors’ financial and ethical decisions, and personal decisions by citizens to adhere to the advice. Following a process when the collectively negotiated goal to fight the pandemic is at stake, puts strong demands on each of these “nodes” in dynamic co-governance.

### Conclusions

We have identified 4 different governance models linked to national eHealth programs, national health ICT infrastructures, regional and local professional networks, Health Internet

businesses, consumer-driven self-management solutions, and virtual health services beyond national health services.

The political model depicts governance through various management strategies to influence wider health political goals and innovation. The model also enacts eHealth governance by establishing financing schemes, infrastructures, standards, and laws. Stronger leadership, by involving stakeholders in national health IT boards, is suggested to achieve goals. The medical-governance model enacts governance within and by eHealth communities grounded in evidence and trust. The authors of papers addressing this model argue that evidence-based medicine should be the basis for the development of political-governance strategies. The internet and global business model put international business actors in dominant roles in eHealth governance, challenges the jurisdictions of nation-states, and limits the influence of international health actors, such as the World Health Organization. The authors of papers that address this model envision a dystopian future, especially for the governance of health data and services on the internet, and they argue for action by global health authorities. In the self-governance model, governance is characterized by shifting, dynamic tensions: governance and control of the individual by their submission to eHealth versus self-governance within eHealth by empowered, conscious citizens.

On the basis of this review, we conclude that to achieve health policy goals in large-scale eHealth policy programs, collective negotiations between nation-states, global policy actors, medical and self-governance actors, and global business and industry actors are essential. Digital technology affordances present opportunities for both benefit and harm concerning the realization of health policy goals, a dynamic that future studies should scrutinize. Technological affordances and both optimistic and pessimistic views deserve serious consideration.

According to our findings, further research is needed to produce knowledge about

1. How large-scale eHealth programs are realizing national and international health goals through collective negotiations
2. How the interference dilemmas between models are (creatively) accommodated and dealt with by patients or consumers to obtain quality and equal access
3. How the consequences of interference between models are (creatively) co-managed by global vendors, regional, national, and international governmental actors, and national and international nongovernment organizations to achieve health policy goals?

### Acknowledgments

The authors are grateful to the research librarian Karianne Lind, who performed the search and was an outstanding resource concerning search terms and technical-search issues. The work was funded by the Norwegian Centre for eHealth Research.

### Conflicts of Interest

None declared.

## References

- Greenhalgh T, Keen J. England's national programme for IT. *Br Med J* 2013 Jun 28;346:f4130. [doi: [10.1136/bmj.f4130](https://doi.org/10.1136/bmj.f4130)] [Medline: [23810959](https://pubmed.ncbi.nlm.nih.gov/23810959/)]
- Delivering 21st Century IT Support for the NHS. National Strategic Programme. 2002. URL: <https://www.hsj.co.uk/home/delivering-21st-century-it-support-for-the-nhs-national-strategic-programme/3090.article> [accessed 2020-10-08]
- Public Health and Promoting Interoperability Programs. Centers for Disease Control and Prevention. 2020. URL: <https://www.cdc.gov/ehrmeaningfuluse/introduction.html> [accessed 2019-09-09]
- Colicchio TK, Cimino JJ, Del Fiol G. Unintended consequences of nationwide electronic health record adoption: challenges and opportunities in the post-meaningful use era. *J Med Internet Res* 2019 Jun 3;21(6):e13313 [FREE Full text] [doi: [10.2196/13313](https://doi.org/10.2196/13313)] [Medline: [31162125](https://pubmed.ncbi.nlm.nih.gov/31162125/)]
- Brennan J, McElligott A, Power N. National health models and the adoption of E-health and E-prescribing in primary care - new evidence from Europe. *J Innov Health Inform* 2015 Nov 25;22(4):399-408 [FREE Full text] [doi: [10.14236/jhi.v22i4.97](https://doi.org/10.14236/jhi.v22i4.97)] [Medline: [26855274](https://pubmed.ncbi.nlm.nih.gov/26855274/)]
- Ansell C, Torfing J. *Handbook on Theories of Governance*. Cheltenham, UK: Edward Elgar Pub; 2016.
- Egeberg M, Trondal J. *An Organizational Approach to Public Governance: Understanding and Design*. Oxford, UK: Oxford University Press; 2018.
- Vabo S, Røiseland A. Governance på norsk. Samstyring som empirisk og analytisk fenomen. *Norsk statsvitenskapelig tidsskrift*. (01-02) 2008;24:86-107 [FREE Full text]
- Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* 2009 Jun;26(2):91-108 [FREE Full text] [doi: [10.1111/j.1471-1842.2009.00848.x](https://doi.org/10.1111/j.1471-1842.2009.00848.x)] [Medline: [19490148](https://pubmed.ncbi.nlm.nih.gov/19490148/)]
- Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016 Dec 5;5(1):210 [FREE Full text] [doi: [10.1186/s13643-016-0384-4](https://doi.org/10.1186/s13643-016-0384-4)] [Medline: [27919275](https://pubmed.ncbi.nlm.nih.gov/27919275/)]
- Glaser BG, Strauss AL. *Discovery of Grounded Theory: Strategies for Qualitative Research* Paperback. New York: Aldine; 1967.
- Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
- Knutsen O, editor. *The Nordic Models in Political Science Challenged, But Still Viable?*. Bergen: Fagbokforlaget; 2017.
- Atalag K. Using a single content model for eHealth interoperability and secondary use. *Stud Health Technol Inform* 2013;193:282-296. [Medline: [24018523](https://pubmed.ncbi.nlm.nih.gov/24018523/)]
- Park Y, Atalag K. Current national approach to healthcare ICT standardization: focus on progress in New Zealand. *Healthc Inform Res* 2015 Jul;21(3):144-151 [FREE Full text] [doi: [10.4258/hir.2015.21.3.144](https://doi.org/10.4258/hir.2015.21.3.144)] [Medline: [26279950](https://pubmed.ncbi.nlm.nih.gov/26279950/)]
- Kierkegaard P. Interoperability after deployment: persistent challenges and regional strategies in Denmark. *Int J Qual Health Care* 2015 Apr;27(2):147-153. [doi: [10.1093/intqhc/mzv009](https://doi.org/10.1093/intqhc/mzv009)] [Medline: [25699549](https://pubmed.ncbi.nlm.nih.gov/25699549/)]
- deRiel E, Puttkammer N, Hyppolite N, Diallo J, Wagner S, Honoré JG, et al. Success factors for implementing and sustaining a mature electronic medical record in a low-resource setting: a case study of iSanté in Haiti. *Health Policy Plan* 2018 Mar 1;33(2):237-246. [doi: [10.1093/heapol/czx171](https://doi.org/10.1093/heapol/czx171)] [Medline: [29253138](https://pubmed.ncbi.nlm.nih.gov/29253138/)]
- Wade VA, Elliott JA, Hiller JE. A qualitative study of ethical, medico-legal and clinical governance matters in Australian telehealth services. *J Telemed Telecare* 2012 Mar;18(2):109-114. [doi: [10.1258/jtt.2011.110808](https://doi.org/10.1258/jtt.2011.110808)] [Medline: [22267306](https://pubmed.ncbi.nlm.nih.gov/22267306/)]
- Crocker M, Cato-Addison WB, Pushpanathan S, Jones TL, Anderson J, Bell BA. Patient safety and image transfer between referring hospitals and neuroscience centres: could we do better? *Br J Neurosurg* 2010 Aug;24(4):391-395. [doi: [10.3109/02688697.2010.508847](https://doi.org/10.3109/02688697.2010.508847)] [Medline: [20726748](https://pubmed.ncbi.nlm.nih.gov/20726748/)]
- Sutton LN. PACS and diagnostic imaging service delivery: a UK perspective. *Eur J Radiol* 2011 May;78(2):243-249. [doi: [10.1016/j.ejrad.2010.05.012](https://doi.org/10.1016/j.ejrad.2010.05.012)] [Medline: [21600402](https://pubmed.ncbi.nlm.nih.gov/21600402/)]
- Bagot KL, Cadilhac DA, Bladin CF, Watkins CL, Vu M, Donnan GA, VSTASTUTE investigators. Integrating acute stroke telemedicine consultations into specialists' usual practice: a qualitative analysis comparing the experience of Australia and the United Kingdom. *BioMed Central Health Serv Res* 2017 Nov 21;17(1):751 [FREE Full text] [doi: [10.1186/s12913-017-2694-1](https://doi.org/10.1186/s12913-017-2694-1)] [Medline: [29157233](https://pubmed.ncbi.nlm.nih.gov/29157233/)]
- Mackey TK, Eysenbach G, Liang BA, Kohler JC, Geissbuhler A, Attaran A. A call for a moratorium on the .health generic top-level domain: preventing the commercialization and exclusive control of online health information. *Global Health* 2014 Sep 26;10:62 [FREE Full text] [doi: [10.1186/s12992-014-0062-z](https://doi.org/10.1186/s12992-014-0062-z)] [Medline: [25283176](https://pubmed.ncbi.nlm.nih.gov/25283176/)]
- Mackey TK, Liang BA, Kohler JC, Attaran A. Health domains for sale: the need for global health internet governance. *J Med Internet Res* 2014 Mar 5;16(3):e62 [FREE Full text] [Medline: [24598602](https://pubmed.ncbi.nlm.nih.gov/24598602/)]
- Williams SJ, Coveney C, Meadows R. 'M-apping' sleep: trends and transformations in the digital age. *Sociol Health Illn* 2015 Sep;37(7):1039-1054. [doi: [10.1111/1467-9566.12283](https://doi.org/10.1111/1467-9566.12283)] [Medline: [26182959](https://pubmed.ncbi.nlm.nih.gov/26182959/)]
- Deutsch E, Duftschmid G, Dorda W. Critical areas of national electronic health record programs-is our focus correct? *Int J Med Inform* 2010 Mar;79(3):211-222. [doi: [10.1016/j.ijmedinf.2009.12.002](https://doi.org/10.1016/j.ijmedinf.2009.12.002)] [Medline: [20079685](https://pubmed.ncbi.nlm.nih.gov/20079685/)]
- Deleuze G. Postscript on the Societies of Control. *October* 1992;59:3-7. [doi: [10.4324/9781315242002-3](https://doi.org/10.4324/9781315242002-3)]

27. Massumi B. Requiem for our prospective dead (toward a participatory critique of capitalist power). In: Kaufman E, Heller KJ. editors. Deleuze and Guattari: New mappings in politics, philosophy, and culture. Minneapolis, USA: University of Minnesota Press; 1998:40-64.
28. Foucault M. Discipline and Punish : The Birth of the Prison. New York: Pantheon Books; 1977.
29. Bauman Z. Liquid Modernity. Cambridge: Polity Press; 2000.
30. Goddard M. The EU general data protection regulation (GDPR): European regulation that has a global impact. International Journal of Market Research 2017 Nov;59(6):703-705. [doi: [10.2501/ijmr-2017-050](https://doi.org/10.2501/ijmr-2017-050)]
31. Petersen A. Digital Health and Technological Promise: A Sociological Inquiry. London, UK: Routledge; 2018.

## Abbreviations

**EHR:** electronic health record  
**gTLDs:** generic top-level domain names  
**HIT:** health information technology  
**ICANN:** Internet Corporation for Assigned Names and Numbers  
**ICT:** information and communication technology  
**IT:** information technology  
**NHS:** National Health Services  
**NPfIT:** National Programme for Information Technology  
**TLD:** top-level domain

*Edited by G Eysenbach; submitted 26.11.19; peer-reviewed by D Appelbe, S Fletcher, J Wolff; comments to author 27.01.20; revised version received 23.06.20; accepted 15.09.20; published 28.10.20.*

*Please cite as:*

*Ekeland AG, Linstad LH*

*Elaborating Models of eHealth Governance: Qualitative Systematic Review*

*J Med Internet Res 2020;22(10):e17214*

*URL: <http://www.jmir.org/2020/10/e17214/>*

*doi: [10.2196/17214](https://doi.org/10.2196/17214)*

*PMID: [33112247](https://pubmed.ncbi.nlm.nih.gov/33112247/)*

©Anne Granstrom Ekeland, Line Helen Linstad. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Integrating Genomics and Clinical Data for Statistical Analysis by Using GENome MINing (GEMINI) and Fast Healthcare Interoperability Resources (FHIR): System Design and Implementation

Julian Gruendner<sup>1</sup>, MA, MSc; Nicolas Wolf<sup>1</sup>, MSc; Lars Tögel<sup>2</sup>, PhD; Florian Haller<sup>2</sup>, MD; Hans-Ulrich Prokosch<sup>1</sup>, PhD; Jan Christoph<sup>1</sup>, PhD

<sup>1</sup>Department of Medical Informatics, Friedrich-Alexander University, Erlangen-Nürnberg, Erlangen-Tennenlohe, Germany

<sup>2</sup>Diagnostic Molecular Pathology, Institute of Pathology, Friedrich-Alexander University, Erlangen-Nürnberg, Erlangen, Germany

**Corresponding Author:**

Julian Gruendner, MA, MSc

Department of Medical Informatics

Friedrich-Alexander University, Erlangen-Nürnberg

Wetterkreuz 15

Erlangen-Tennenlohe, 91058

Germany

Phone: 49 9131 8567787

Email: [julian.gruendner@fau.de](mailto:julian.gruendner@fau.de)

## Abstract

**Background:** The introduction of next-generation sequencing (NGS) into molecular cancer diagnostics has led to an increase in the data available for the identification and evaluation of driver mutations and for defining personalized cancer treatment regimens. The meaningful combination of omics data, ie, pathogenic gene variants and alterations with other patient data, to understand the full picture of malignancy has been challenging.

**Objective:** This study describes the implementation of a system capable of processing, analyzing, and subsequently combining NGS data with other clinical patient data for analysis within and across institutions.

**Methods:** On the basis of the already existing NGS analysis workflows for the identification of malignant gene variants at the Institute of Pathology of the University Hospital Erlangen, we defined basic requirements on an NGS processing and analysis pipeline and implemented a pipeline based on the GEMINI (GENome MINing) open source genetic variation database. For the purpose of validation, this pipeline was applied to data from the 1000 Genomes Project and subsequently to NGS data derived from 206 patients of a local hospital. We further integrated the pipeline into existing structures of data integration centers at the University Hospital Erlangen and combined NGS data with local nongenomic patient-derived data available in Fast Healthcare Interoperability Resources format.

**Results:** Using data from the 1000 Genomes Project and from the patient cohort as input, the implemented system produced the same results as already established methodologies. Further, it satisfied all our identified requirements and was successfully integrated into the existing infrastructure. Finally, we showed in an exemplary analysis how the data could be quickly loaded into and analyzed in KETOS, a web-based analysis platform for statistical analysis and clinical decision support.

**Conclusions:** This study demonstrates that the GEMINI open source database can be augmented to create an NGS analysis pipeline. The pipeline generates high-quality results consistent with the already established workflows for gene variant annotation and pathological evaluation. We further demonstrate how NGS-derived genomic and other clinical data can be combined for further statistical analysis, thereby providing for data integration using standardized vocabularies and methods. Finally, we demonstrate the feasibility of the pipeline integration into hospital workflows by providing an exemplary integration into the data integration center infrastructure, which is currently being established across Germany.

(*J Med Internet Res* 2020;22(10):e19879) doi:[10.2196/19879](https://doi.org/10.2196/19879)



**KEYWORDS**

next-generation sequencing; data analysis; genetic databases; GENome MINing; Fast Healthcare Interoperability Resources; data standardization

## Introduction

### Background

Combining omics data with other clinical patient data has been the focus of multiple studies in the past years [1-3]. In particular, the emergence and widespread use of next-generation sequencing (NGS) for the identification of pathological gene variants has led to an increase in the amount of data available for diagnosis. This has directly improved the quality of medical care for many diseases [4-6]. The meaningful combination of NGS data with other patient data from varying sources has been challenging for a number of reasons. First, the amount of data generated by NGS is vast compared to other data such as laboratory and clinical data, data derived from classical diagnostic tests, and demographic data, for example, age and gender. Second, NGS data typically contains many data points, which might not be relevant for subsequent analysis. Third, NGS data are highly sensitive in terms of privacy and more difficult to anonymize in a meaningful way. Therefore, any analysis within and across institutions will have to be carefully considered and crafted in a privacy-preserving manner. Many different studies have been conducted to investigate the feasibility of directly integrating omics data into clinical data repositories such as OMOP (Observational Medical Outcomes Partnership) [7], i2b2 (informatics for integrating biology and the bedside), and transMART [2]. The OHDSI-OMOP (Observational Health Data Sciences and Informatics-OMOP) common data model (CDM) focuses on observational research, and i2b2 focuses on the integration of different types of data into 1 clinical repository. These systems require an extra genomics pipeline to be run before the data can be loaded and integrated into the data repositories. Many of the proposed methods are not optimized for initial data investigation, automatic annotation of the data, or privacy-preserving cross-hospital analysis.

This study focusses on bridging this gap by creating a system that supports the whole process. It starts from variant call format (VCF) files. The data are then annotated and preanalyzed. In the final step, selected genomics data are combined with other structured and standardized patient data in 1 data set in a table format for further statistical analysis. We achieve this by extending the open source framework GEMINI (GENome MINing) [8]. It allows a user to load VCF files and places genetic variants, sample phenotypes and genotypes, as well as genome annotations into 1 database. This supports powerful exploration of genetic variations for disease and population genetics. GEMINI makes it possible to analyze the NGS data using structured query language (SQL), which allows researchers to filter variants by clinical relevance, rarity, and read quality. Our proposed system then combines GEMINI with the Fast Healthcare Interoperability Resources (FHIR) [9], a standard for health care data exchange. FHIR describes the clinical information in so-called resources. These modular components describe different elements found in electronic

health records, for example, a patient or a clinical observation such as a laboratory result. It was developed to address the shortcomings of the previously developed HL7 clinical care document standard. The FHIR standard aims to improve interoperability, and its lightweight nature and direct use of common data formats such as JSON and XML allows it to easily integrate with lightweight web services, similar to the ones created in the pipeline described here.

In order to create a pipeline on standardized data such as FHIR, heterogenous hospital data need to be transformed into a standardized format first. The German Medical Informatics Initiative (MI-I) [10] has recently funded 4 consortia of mainly university hospitals across Germany to investigate how heterogenous clinical data, including omics data, can be integrated into clinical data repositories. One aim of the MI-I is to establish data integration centers (DICs), which are the backbone of the cross-hospital and cross-consortia communication. The DICs will provide different services, including data integration, data harmonization, standardized data repositories, consent management, and ID management [11-14]. This study builds on efforts of 2 use cases from the MIRACUM (Medical Informatics in Research and Care in University Medicine) [14], particularly on 1 use case, which aims to establish a genomics pipeline to support NGS data interpretation and clinical decision making at molecular tumor boards. This MIRACUM-Pipe [15], similar to other genomics pipelines, creates VCF files in the process. These VCF files provide a good point of integration for further data analysis beyond the use for the molecular tumor board. Another use case of the MIRACUM consortium focuses on analyzing data within and across hospitals in a privacy-preserving manner, as well as deploying prediction and decision support models in a clinical context.

### Aim of the Study

In order to treat individual patients within the molecular tumor board, individual sequencing data are generated in VCF. Our study supports research by building on these data, aggregating them across multiple patients and enriching them with clinical data. These integrated data sets are analyzed with open-source analysis tools and frameworks to generate or validate hypotheses. Our proposed system extends GEMINI with multiple web services and a user interface to support the whole data flow as well as the subsequent integration with the already established FHIR [9] data repository, KETOS [16], and DataSHIELD [17] platforms for statistical analysis. To validate our approach, we implemented and applied this data analysis pipeline to data from the 1000 Genomes Project, which yielded highly consistent results. We then applied our pipeline to high-throughput gene panel sequencing data from a cohort of 206 patients from the Institute of Pathology of the University Hospital Erlangen and compared the findings with the results of the established Illumina genomics pipeline [18]. Finally, as a proof of concept, we combined the selected data from this patient cohort with diagnosis codes from the local FHIR server



and analyzed the correlation between locations based on the diagnoses and gene mutations to demonstrate one potential use of the system.

## Methods

### Common Requirements for NGS Analysis

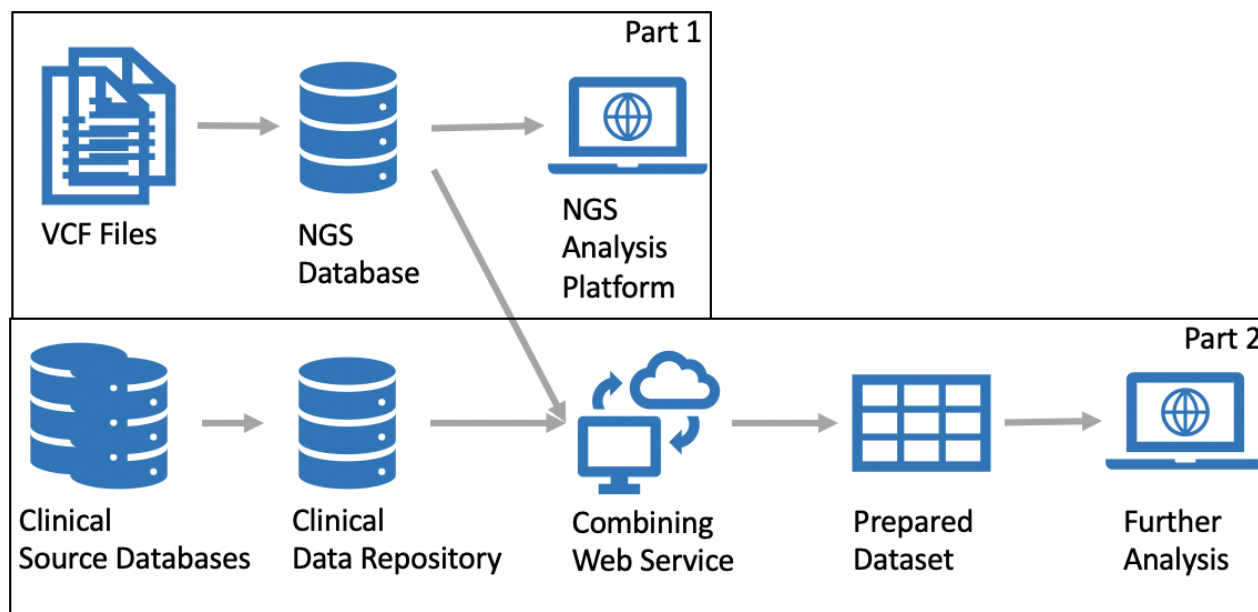
To understand the requirements for NGS analysis, we studied the status quo of the Institute of Pathology of the University Hospital Erlangen. The institute analyses gene mutations by using a commercial pipeline provided by Illumina and presents its findings to the local molecular tumor board. A detailed description of the DNA and RNA library preparations from patient tissues can be found in [Multimedia Appendix 1](#). The pipeline currently used by the institute has limited capabilities when it comes to performing cross-patient analysis and is not easily integrated with other open-source solutions. Therefore, the following requirements were defined, which need to be met: be on-premises, be open-source, deliver the same results as a commercial product, allow analysis across multiple patients, and possibility of web-based integration with existing and new infrastructure to be developed.

### Data Analysis

The biggest drawback of the current system is that it does not support cross-patient analysis and integration of additional

patient data, as the patient data are distributed across multiple systems in the hospital. The system therefore does not currently support a way to, for example, correlate a patient's diagnosis with particular gene mutations as is done in our exemplary analysis shown below. We therefore propose a new pipeline, which improves on these drawbacks. [Figure 1](#) depicts the multistep analysis workflow of such a genomics pipeline. In the first step, VCF files are uploaded into a database for gene annotation and, subsequently, gene variant filtering and analysis. To allow further processing of the data, it is crucial to filter variants that are irrelevant for a particular research question. The amount of data and noise must be reduced. For this, researchers need to be able to query and filter the genomic database. This would satisfy any research only focusing on NGS data; however, in a clinical setting, NGS data become truly meaningful when clinical patient data are available. To achieve this, a standardized way to combine genomics data with clinical data is essential. Finally, combined data needs to be transformed to a common format that can easily be analyzed, processed further, and integrated into existing data analysis frameworks. The resulting system described in [Figure 1](#) can be split into 2 larger parts: (1) Part 1: an NGS processing and analysis pipeline and (2) Part 2: a system for integrating clinical patient data for further analysis.

**Figure 1.** Genomics pipeline and analysis system. NGS: next-generation sequencing; VCF: variant call format.

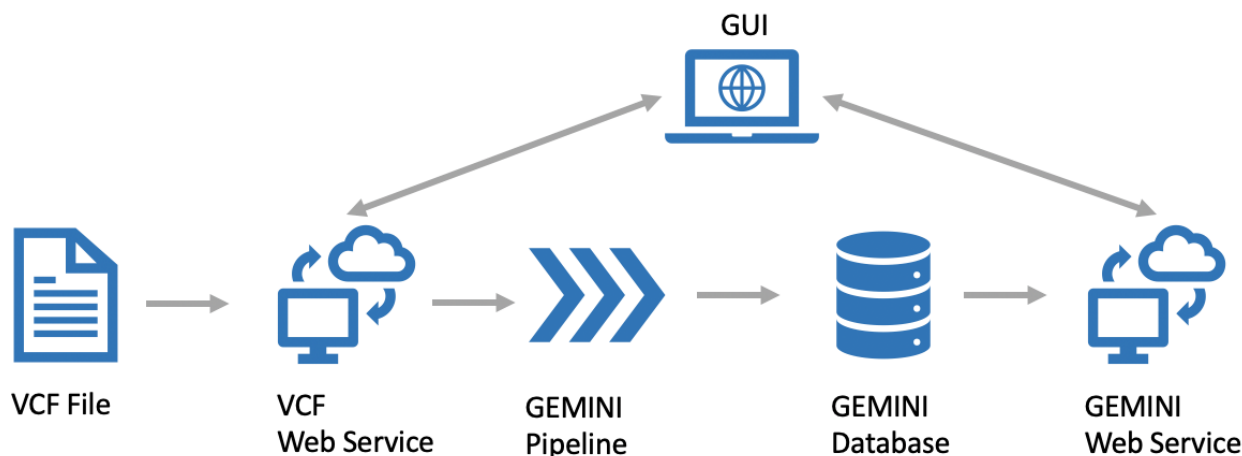


### Part 1: NGS Processing and Analysis Pipeline

[Figure 2](#) shows the first part of the architecture of the implemented genomics pipeline and analysis system. The proposed system includes a graphical user interface (GUI, [Figure](#)

[2](#)), which supports all steps. The GUI is used to upload a VCF file via the VCF web service, which is then processed and loaded into GEMINI via the GEMINI web service. The web service then uses our integrated GEMINI pipeline ([Figure 3](#)) to process and load the data into GEMINI.

**Figure 2.** Part 1: Next-generation sequencing processing and analysis pipeline architecture. GUI: graphical user interface; VCF: variant call format; GEMINI: GEnome MINIng.



**Figure 3.** GEMINI pipeline schema. SnpEff: single nucleotide polymorphism effect; GEMINI: GEnome MINIng.



This GEMINI pipeline (Figure 3) first decomposes the VCF file using the Variant Tools software [19] decompose functionality, which splits up multiallelic variants of a VCF file, resulting in a separate row for each reference sequence/alternative sequence pair. The decomposed VCF file is then normalized using the normalize function of the Variant Tools software. This step ensures that each VCF entry is left aligned and parsimonious, as described by Tan et al [20]. The resulting VCF file is then automatically annotated using the SNP (single nucleotide polymorphism) effect (SnpEff) tool, reported by Cingolani et al [21]. The SnpEff annotation process enriches the VCF file with information about possible effects or malignancies. One important functionality of GEMINI is the addition of multiple extra annotations during the load process.

This is the last step of our GEMINI pipeline (GEMINI-Load, Figure 3) and includes the following annotation sources: 1000 Genomes Project (population data, allele frequencies) [22], dbSNP (reference snp IDs according to the National Center of Biotechnology Information) [23], and ClinVar (disease information, eg, disease name or clinical relevance) [24].

Once loaded, data can be analyzed using the user interface (GUI, Figure 4) and the GEMINI web service. The user interface allows one to download a GEMINI database in the form of an SQLite file for further use. It further implements a web-based query tool for the GEMINI database. The advantage of this web-based query tool is that the researcher does not need to learn how to use the command line tool, while still providing full GEMINI query functionality.

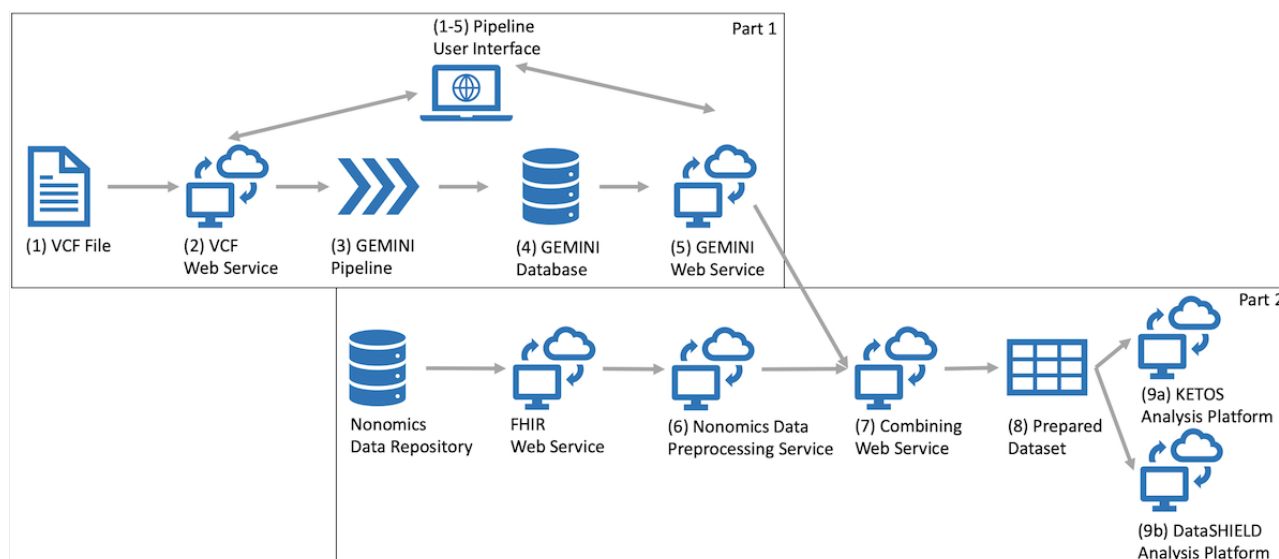
**Figure 4.** Exemplified user interface.

chrom	start	end	gene
chr3	138413708	138413709	PIK3CB
chr17	7578211	7578212	TP53
chr2	30143492	30143499	ALK

## Part 2: A System for Integrating Clinical Patient Data for Further Analysis

Figure 5 depicts the structure of the complete system (numbers indicate the order of the executable steps). A VCF file (Step 1) is read, annotated, and loaded into GEMINI (steps 2-4). Then, NGS data are combined with nonomics, clinical patient data in an FHIR format using the patient ID as a link in a prepared data set (steps 5-8) for the final analysis. The combining of the NGS and clinical patient data is performed by the combining web service, which is described in more detail with an example in

the Results section. One of the advantages of an on-premises open source solution is the straightforward extendibility with the other already established web services. Here, we used the KETOS analysis platform established as part of the local DIC. Further, the creation of a prepared data set in the form of a comma-separated values (CSV) file allows a direct import into the DataSHIELD platform for privacy-preserving cross-hospital analysis (steps 8, 9a, and 9b). DataSHIELD supports this as its underlying data warehouse supports the loading of CSV files for further analysis.

**Figure 5.** Combining genomics and patient data. VCF: variant call format; GEMINI: GENome MINIng; FHIR: Fast Healthcare Interoperability Resources.

## Results

### Overview

We implemented and deployed the above-described architecture at the University Hospital Erlangen. Its functionality was then validated by comparing the results of our analysis system to the results from the 1000 Genomes Project [22] and subsequently by assessing the ability of the system to identify relevant tumor-associated gene variants. Finally, we used the system to combine gene mutation data from 206 patients with diagnosis codes from the local FHIR data repository. We analyzed the resulting prepared data set by using KETOS and Jupyter

Notebook (interactive cell-based code development in a web browser) [25].

### Comparison with the 1000 Genomes Project

To demonstrate the accurate functionality of the pipeline, we investigated the genomic variations of the X-chromosome in a cohort of 1092 individuals, which were studied in the initial phase of the 1000 Genomes Project [26]. First, we loaded the publicly available VCF file data supplied by the study into a GEMINI database by using our pipeline. Then, we compared the variation data in our database to the results from the 1000 Genomes Project [27] (Table 1).

**Table 1.** Comparison of the statistical evaluation of the 1000 Genomes Project and GEMINI (GEnome MINIng) pipeline.

Mutations	1000 Genomes Project (n)	GEMINI (n)
<b>Single nucleotide polymorphisms</b>		
Total variants	~1,300,000	1,275,275
Average per sample	~105,000	104,757
<b>Indels</b>		
Total variants	~59,000	59,157
Average per sample	~13,000	12,715
<b>Large deletions</b>		
Total variants	432	432
Average per sample	26	26

As shown in Table 1, results generated with our system were identical to the results generated by the 1000 Genomes Project. The analysis, including converting and loading the data into GEMINI, took approximately 3 hours to run on a MacBook Pro (15-inch, 2016)-System with 2.7 GHz Intel Core i7-Processor and 16 GB 2133 MHz LPDDR3-RAM. Once loaded, the SQL queries took approximately 120 seconds to complete. This clearly demonstrates that the bottleneck is loading the data rather than the subsequent analysis using GEMINI.

### Comparison With Illumina Pipeline and Genomic Analysis

#### Single Patient Analysis

We loaded individual patient-derived VCF files into the GEMINI database and prepared an SQL query according to

established filter criteria used in the Illumina pipeline at the Institute of Pathology. In the first iteration, the resulting SQL query (Multimedia Appendix 2) filtered the following: all variants that failed to pass any of the variant quality filters, all variants that displayed a low impact in regard to protein functionality, and all variants that were not considered pathogenic as they did not change the coding protein (eg, synonymous variants, intronic variants, upstream/downstream variants). In addition, all variants with a population-based allele frequency of  $\geq 2\%$  were also excluded. Table 2 shows an exemplary comparison of the results from the Illumina versus the GEMINI pipeline. Illumina VariantStudio yielded 16 variants, while GEMINI yielded 15 variants. The Illumina VariantStudio included chromosome 4 position 10085736. GEMINI also included chromosome 9 position 21970916.

**Table 2.** Comparison of the filtered results of the mutations in the GEMINI (GEName MINIng) and Illumina pipeline.

Chromosome	Position	Codon change (according to the Human Genome Variation Society coding)	Illumina	GEMINI
1	40366658	c.539A>G	✓	✓
1	40366659	c.537_538insCG	✓	✓
2	215632255	c.1518_1519delTGinsCA	✓	✓
3	12645693	c.776C>G	✓	✓
4	106156187	c.1088C>T	✓	
5	112178795	c.7504G>A	✓	✓
5	176520270	c.1189_1190delGGinsAC	✓	✓
5	176522605	c.1702_1704delCCAinsGCC	✓	✓
7	116411990	c.3029C>T	✓	✓
9	21970916	c.442G>A		✓
16	3778424	c.6624A>C	✓	✓
16	3779338	c.5709dupG	✓	✓
16	3779338	c.5709delG	✓	✓
16	3779361	c.5687A>C	✓	✓
17	7579408	c.277_278delCT	✓	✓
22	41546158	c.2773C>A	✓	✓

Based on this result, we revised the filter criteria ([Multimedia Appendix 3](#)) so that only variants with the filter impact\_severity *high* or one of the impact types (disruptive\_inframe\_deletion, disruptive\_inframe\_insertion, missense\_variant) were included in the results. Furthermore, the ClinVar database was used to assess the pathogenicity of individual gene alterations, which is readily available in the annotated GEMINI database. All mutations that were classified as benign, likely\_benign, and benign/likely\_benign according to the ClinVar database were deleted from the data set. Mutations without known ClinVar status were kept and manually classified by interrogating web-based variant repositories. The adjusted SQL query yielded the same results as the Illumina pipeline, which qualifies the open source solution for routine use in clinical practice. Finally, the automation of the filtering steps saves hands-on time of approximately 30 minutes per patient analysis.

### Multiple Patient Analysis

We used gene sequencing data of 206 patients from the University Hospital Erlangen. To analyze the data, the 206 individual VCF files were first merged and then loaded into GEMINI. Each file had a size of 57 MB, which adds up to a total file size of 12 GB. The merged VCF file had a size of 4.8 GB, and the GEMINI database a size of 664 MB. To analyze the whole patient cohort, we first created an SQL query to provide simple descriptive statistics of the cohort, such as the overall number of mutations by *impact\_severity* and *type* (snp, indel). As expected, most mutations were classified with a low impact severity. We then further analyzed the merged data with another SQL query to determine the occurrence of gene mutations, impact, and codon\_change ([Multimedia Appendix 4](#)). This analysis revealed that impact\_severity is not a suitable filter for dichotomizing high impact variants from synonymous

variants. Therefore, specific impacts (impact column) were used instead. In addition, the same filter criteria as described for the adjusted query were used for further analysis ([Multimedia Appendix 3](#)). Looking at the results in more detail, it became clear that many of the mutations were sequencing artefacts, which would have to be excluded before further analysis. Especially critical is the fact that merging of multiple VCF files leads to information loss and reduces the quality of the sequencing data. We discovered that the merge function vcf-merge of VCFtools only keeps the lowest passed quality filter for each variant. This means that a higher quality grade will only be listed in the FILTER column if it was passed by each sample for a variant. Other merge functions such as the merge function of BCFtools allow the user to choose between either keeping the lowest passed or the highest passed filter for each variant. However, in all cases, the quality criteria of each individual sample cannot be reconstructed. This poses a potential problem for further analysis as 1 low-quality read for a variant in 1 sample would obscure potential good results in others, leading to the exclusion of good-quality reads; vice versa, 1 high-quality read may mask low-quality reads, which might impact the data analysis. A solution for this problem is to either exclude particularly poor-quality samples before joint analysis is executed or to load and analyze samples individually in an automated way. This will preserve sample quality information in a combined data set.

### Creating a Prepared Data Set for Combined Analysis

To create a combined prepared data set and to make it available for further analysis, we integrated the combining web service with the KETOS platform by making the web service, which combines the NGS data with the nonomics data, available to the KETOS host by using internet protocol address restriction



as well as password protection. The combined prepared data set is a combined subset of all the available NGS and clinical data, which were prepared for 1 specific research question. The combining web service data for a particular research question is identified using JSON syntax. It includes an *fhir* part to specify patient data in the form of the required FHIR resources and a GEMINI part to specify the filtered variants to be included. The specification also requires patient IDs to identify the patient cohort for further analysis (see [Multimedia Appendix 5](#)). In an initial combining web service specification, we required the user to specify the exact variant positions and alleles. This

returned a data set in which the variant-specific fields are connected to the variant via a prefix, which is added to all extracted variant fields. The prefix contains the chromosome, chromosomal location of the variant, as well as the nonreference allele, for example, chrX#154158284#154158285#C#. The web service collects the data for the prepared data set and returns the data either as a JSON array or as a CSV file. The resulting data set in the CSV format is then created as shown in [Table 3](#). It contains 1 row for each patient. Each row includes the patient ID as well as other fields specified by the user.

**Table 3.** Example of a prepared data set.<sup>a</sup>

Patient_ID	Gender	Date of birth	Disease	<prefix> <sup>b</sup> ref	<prefix> <sup>b</sup> alt	<prefix> <sup>b</sup> gts
28	male	01.10.41	1	G	C	C

<sup>a</sup>This table shows only the examples of values.

<sup>b</sup><prefix> is the concatenation of the chromosome, chromosomal location of the variant, and nonreference allele, eg, chrX#154158284#154158285#C#—this would, for example, result in the following column name: chrX#154158284#154158285#C#ref.

In our further analysis, we found this specification of the combining web service to be limiting. We therefore extended the web service to allow the researcher to specify the GEMINI query used to collect the NGS data directly in SQL, rather than being limited to extracting fields by variant position only. The resulting request specification is analogous to the one described in [Multimedia Appendix 5](#), except for the following extending fields inside the GEMINI part: *sql\_mode*, *sql*, *columns*, *rows*. This allowed us to create an example of a data set by using a query similar to that shown in [Multimedia Appendix 6](#), which

resulted in a data set similar to [Table 4](#). This allowed us to avoid information loss regarding read quality, as the combining web service can query an individual database for each patient, thereby omitting data-merging steps. The resulting data set could then be imported into the KETOS analysis environment where further analysis could be conducted within a Jupyter Notebook [25] by using a researcher's favorite tools. Further, the comma-separated nature of the data set means that it could also be imported into DataSHIELD, which supports further privacy-preserving analysis across institutions.

**Table 4.** Format of the initial raw data set.<sup>a</sup>

Patient_ID <sup>b</sup>	Gender <sup>c</sup>	Date of birth <sup>d</sup>	Diagnosis <sup>e</sup> (ICD <sup>f</sup> 10 Code)	Gene <sup>h</sup>
PSEUDO-ID-1	Female	01.01.50	2019-01-01T00:00:00+00:00	<i>PIK3CB</i>
PSEUDO-ID-2	Male	01.01.50	2019-01-01T00:00:00+00:00	<i>PIK3CB</i>
PSEUDO-ID-3	Male	01.01.50	-	-
PSEUDO-ID-4	Male	01.01.50	2019-01-01T00:00:00+00:00	-

<sup>a</sup>Since the combined data set comprised 206 patients, 135 diagnoses, and 152 genes, values shown in this table are only examples, as the entire data cannot be represented here.

<sup>b</sup>Examples of IDs of patients.

<sup>c</sup>Examples of genders.

<sup>d</sup>Examples of birth dates of patients.

<sup>e</sup>Example value, C20 or C61; timestamp diagnosis. (One column per diagnosis - if there is no diagnosis for a patient, the column will be empty in this patient's row.)

<sup>f</sup>International Classification of Diseases.

<sup>h</sup>Gene name examples. (One column per Gene - if there is no gene mutation for a patient, the column will be empty in this patient's row.)

## Exemplary Analysis of Tumor Entity and Diagnosis Inside KETOS by Using Jupyter Notebook

As a proof of concept for our system, we created a data set with combined diagnostics and tumor mutational data for further analysis. The initial data set included 206 patients. In the first step, we identified 135 cancer-related diagnosis codes across our cohort. We then used the revised GEMINI query ([Multimedia Appendix 3](#)) to collect the gene mutations across our cohort. The combined data set comprised 206 patients, 135

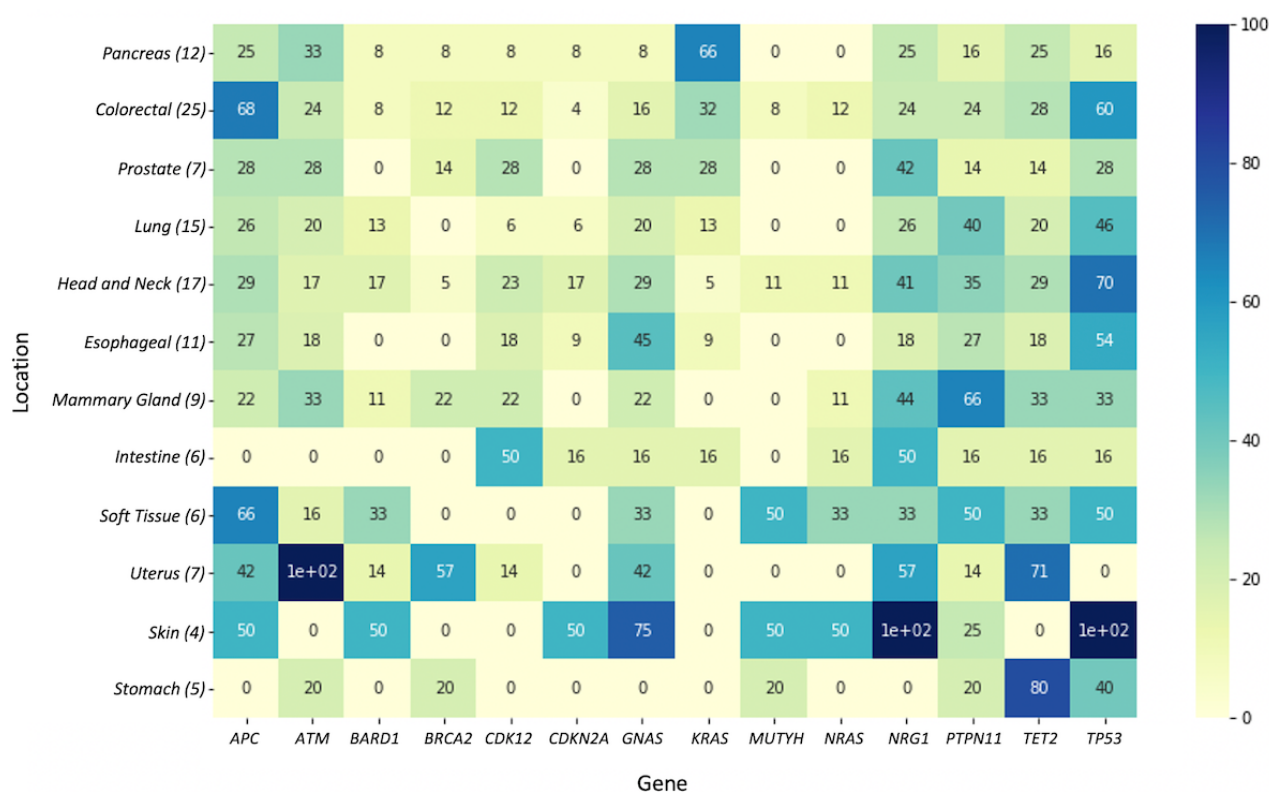
diagnoses, and 152 genes. The final raw data set had the format shown in [Table 4](#).

Using Jupyter Notebook in one of the analysis environments inside the locally installed KETOS data analysis platform; we requested the data from our combining web service. Patients without diagnosis were removed. We removed all secondary tumor diagnoses from our data set and focused on primary tumor locations. Tumor entities represented by less than 4 patients were regarded as underrepresented and were also removed prior

to further analysis. This yielded a sample group of 124 patients across 12 tumor entities. A total of 142 unique genes with mutations were identified across this population. As a proof of principle, we aimed to identify the driver mutations for each tumor entity. Therefore, we included only genes that were mutated in more than 50% of the patients, resulting in the heat map depicted in Figure 6. This filtering step identified 14 frequently mutated genes within the patient cohort, mainly *bona fide* tumor suppressor genes such as *APC*, *ATM*, *BRCA2*, or *TP53* and well-known protooncogenes such as *GNAS*, *KRAS*, and *NRAS*. Notably, this analysis revealed tumor characteristic

mutational profiles. For instance, the colorectal cancer cohort displayed a characteristic rate of *APC*, *TP53*, *KRAS*, and *NRAS* mutations when compared to previously reported data [28]. Likewise, *KRAS* mutations were highly prevalent in the pancreatic cancer cohort, a malignancy with an observed *KRAS* mutation rate of up to 95% [29]. Since patients with cancers were distributed across multiple sites and the number of patients for some locations was small, a very accurate distribution by location might have been lost in the analysis process. This could be remedied using a larger sample size and factoring the multisite cancers into the analysis.

**Figure 6.** Distribution (%) of the gene mutations by location. Y-axis: location (number of patients); x-axis: gene, eg, 80% for stomach for *TET2* means that 4 of 5 patients with stomach cancer had a mutation in *TET2*.



## Discussion

### Overview

We presented an open-source web-based NGS annotation pipeline based on the GEMINI database, which produced the same results as the established Illumina pipeline used in routine diagnostics. We then extended our pipeline to combine the annotated NGS data with other clinical data and demonstrated the feasibility of our approach in an exemplary analysis, which combined gene mutation data with clinical diagnosis codes.

The new pipeline improved on the current pipeline by supporting this cross-patient analysis. This allowed us to combine the NGS data with other patient data from other systems across the hospital in 1 data set. The source code of this project is open to the public and is available on GitHub [30].

Previous studies have investigated the integration of omics with other clinical data. In particular, 2 larger systems, which have recently seen efforts to achieve this integration, are the OHDSI-OMOP CDM and i2b2. Murphy et al [2] compared 3 different strategies for integrating omics data into i2b2, 2 of which combined omics and clinical data in the same database. The third approach is similar to the one described in this study. It kept omics data in a separate database. For the OMOP CDM, NGS data were integrated into 1 database with clinical data by creating additional tables for the omics data. All approaches had 1 thing in common that they did not provide a user interface for loading data into the respective databases. In contrast, our pipeline significantly improves usability by providing a user interface, which allows a user to process VCF files and load them into the GEMINI analysis database. Additionally, using an established omics database allowed us to fast-track querying and preanalysis of the genomics data, thereby significantly reducing the amount of data at an early stage in the process.

GEMINI is part of the pipeline used by the MIRACUM consortium and therefore provides us with a clear point of integration. This means that once the MIRACUM-Pipe has been established across the different DICs, it can potentially replace the Illumina pipeline and integrate directly with the solution described here.

### Generalizability of the Pipeline

In this paper, we focus on the integration with the DICs currently being established across Germany. However, as the system relies on common formats (VCF for NGS data and FHIR for clinical patient data), it could easily be applied to other hospitals. The FHIR format, in particular, is currently seeing a rise in popularity and as electronic health record vendors will provide more and more data in the FHIR format, establishing a local FHIR store with patient data should be feasible for many institutions. However, as with any standardized format, studies that have captured data in their own format or another standard format such as OMOP would have to convert their data into FHIR formatted data in order to take advantage of this solution. As the whole solution can be deployed on premises, acceptance for the solution should be high as the deployment can be tailored to the specific needs of individual institutions and no data has to leave the institution at any time. In our example, we show how the prepared data sets, which result from the pipeline, can be analyzed in Jupyter Notebook and KETOS, a web-based analysis platform. Generally, the prepared data sets in the CSV format could be read into many other analysis platforms such as DataSHIELD, converted into table-like formats, and analyzed further.

### Potential Use of the System

We have demonstrated in this paper how our system can be used to combine diagnoses with NGS data for further analysis. Connecting the system to an FHIR repository of standardized data means that there is a great opportunity to undertake further analyses in the future with other data available in an FHIR store. The MIRACUM FHIR store at the University Hospital in Erlangen currently has information about patients, encounters, diagnoses, laboratory results, procedures, and medications. The system described here would, for example, support further investigation of the correlation between the diagnosis of COVID-19, disease outcome from hospital discharge information, and gene mutations. The prerequisite for this would be that the gene sequences of respective patients are sequenced as part of a wider investigation.

### Lessons Learned

We were able to augment the existing GEMINI software with web services. The resulting pipeline, which is based on the open source project, could be hosted on premises. It fulfilled the requirement of producing the same results as the established commercial pipeline. One of the biggest drawbacks of the GEMINI database is that it does not allow the uploading of the data of multiple patients directly into 1 database. Instead, data from multiple patients have to be merged first, which leads to a loss of information regarding the sequencing quality for a particular variant. This may impact analysis and data interpretation as low-quality reads and high-quality reads would

be interpreted in the same way. However, this limitation can be augmented by preselecting files with good-quality reads or loading each file individually and focusing only on variants that are relevant to the particular research question. The combining web service in `sql_mode` described above can also be used for this type of analysis by simply requesting GEMINI data from individual databases. Once loaded into the GEMINI database, queries for a particular variant were very fast and results were returned within seconds, making the analysis across multiple patients feasible.

### Limitations

The web-based user interface provides direct access to different databases. However, exporting prepared data sets is currently not supported and has to be triggered using a web service representational state transfer call. Therefore, the platform presented here requires the user to be proficient with SQL and basic programming skills in order to extract and analyze the data, making it unsuitable for users with no programming or SQL skills. The chosen method of prefiltering data for further analysis using an established, fast open source tool allows one to avoid large data volumes. However, it also limits the explorative analysis of NGS with other patient data. The medical device regulations in Germany approves the use of the system described above for research purposes only. The MIRACUM DICs follow these regulations. The reliance on FHIR requires the infrastructure to provide an FHIR server. However, the MI-I initiative has already set FHIR as the format of choice for interconsortia communication [31].

### Future Directions

In this study, we focused on 1 hospital to show how a potential analysis can be made possible within a hospital. In the future, it would be of interest to duplicate the platform across multiple hospitals to establish cross-hospital analysis pipelines and run analyses across institutions by using DataSHIELD. Pipeline automatization, ie, automated variant annotation and execution of predefined variant filtering/classification steps as well as automated inclusion of results in clinical reports could lead to significant time savings. Another prospective extension would be the integration of therapy suggestions from different web-based databases such as Somatic Mutations In Cancer [32], My Cancer Genome [33], and ClinicalTrials.gov [34]. The prepared data set or combining web service should be integrated further into existing workflows to automate data selection and preparation as well as to make the data provision easier. The FHIR standard for patient data is constantly being developed and has been standardized to integrate omics data directly. Our approach could be extended to load selected data directly into an FHIR database, which would make explorative analysis easier.

### Conclusion

In this study, we successfully demonstrated how NGS genomics data can be combined with FHIR clinical data to provide accessory analysis to existing gene variant analysis solutions in clinical settings. The chosen method of prefiltering data for further analysis by using an established database, which is based on the fast GEMINI open source tool, means that large data

volumes can be avoided. In addition, we showed that a pipeline and web services built on open-source tools delivered the same results as a commercial product and could be hosted on the premises, and be integrated well within a clinical DIC, building

on existing structures, and benefiting from the data standardization DICs provide. Finally, we showed how the system could be used to create and analyze a data set, which included gene mutation and diagnosis data.

## Acknowledgments

The present work was performed in fulfilment of the requirements for obtaining the degree “Dr. rer. biol. hum.” from the Friedrich-Alexander-Universität Erlangen-Nürnberg (JG). This work was supported by grants from MIRACUM project, which is funded by the German Ministry for Education and Research (BMBF) (FKZ 01ZZ1801A).

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Institute of Pathology library creation pipeline.

[[PDF File \(Adobe PDF File\), 114 KB](#) - [jmir\\_v22i10e19879\\_app1.pdf](#)]

### Multimedia Appendix 2

Initial GEMINI query.

[[PDF File \(Adobe PDF File\), 68 KB](#) - [jmir\\_v22i10e19879\\_app2.pdf](#)]

### Multimedia Appendix 3

Revised GEMINI query.

[[PDF File \(Adobe PDF File\), 70 KB](#) - [jmir\\_v22i10e19879\\_app3.pdf](#)]

### Multimedia Appendix 4

Example of multipatient GEMINI query.

[[PDF File \(Adobe PDF File\), 67 KB](#) - [jmir\\_v22i10e19879\\_app4.pdf](#)]

### Multimedia Appendix 5

Example of combiner webservice JSON query.

[[PDF File \(Adobe PDF File\), 73 KB](#) - [jmir\\_v22i10e19879\\_app5.pdf](#)]

### Multimedia Appendix 6

Example of query request: SQL extension.

[[PDF File \(Adobe PDF File\), 67 KB](#) - [jmir\\_v22i10e19879\\_app6.pdf](#)]

## References

1. Shin SJ, You SC, Park YR, Roh J, Kim J, Haam S, et al. Genomic Common Data Model for Seamless Interoperation of Biomedical Data in Clinical Practice: Retrospective Study. *J Med Internet Res* 2019 Mar 26;21(3):e13249 [[FREE Full text](#)] [doi: [10.2196/13249](#)] [Medline: [30912749](#)]
2. Murphy SN, Avillach P, Bellazzi R, Phillips L, Gabetta M, Eran A, et al. Combining clinical and genomics queries using i2b2 - Three methods. *PLoS One* 2017;12(4):e0172187 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0172187](#)] [Medline: [28388645](#)]
3. López de Maturana E, Alonso L, Alarcón P, Martín-Antoniano IA, Pineda S, Piorno L, et al. Challenges in the Integration of Omics and Non-Omics Data. *Genes (Basel)* 2019 Mar 20;10(3):238 [[FREE Full text](#)] [doi: [10.3390/genes10030238](#)] [Medline: [30897838](#)]
4. Schwaederle M, Parker BA, Schwab RB, Daniels GA, Piccioni DE, Kesari S, et al. Precision Oncology: The UC San Diego Moores Cancer Center PREDICT Experience. *Mol Cancer Ther* 2016 Apr;15(4):743-752. [doi: [10.1158/1535-7163.MCT-15-0795](#)] [Medline: [26873727](#)]
5. Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba II, et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *JAMA* 2014 May 21;311(19):1998-2006 [[FREE Full text](#)] [doi: [10.1001/jama.2014.3741](#)] [Medline: [24846037](#)]



6. Stockley TL, Oza AM, Berman HK, Leighl NB, Knox JJ, Shepherd FA, et al. Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. *Genome Med* 2016 Oct 25;8(1):109 [FREE Full text] [doi: [10.1186/s13073-016-0364-2](https://doi.org/10.1186/s13073-016-0364-2)] [Medline: [27782854](https://pubmed.ncbi.nlm.nih.gov/27782854/)]
7. Shin SJ, You SC, Roh J, Park YR, Park RW. Genomic Common Data Model for Biomedical Data in Clinical Practice. *Stud Health Technol Inform* 2019 Aug 21;264:1843-1844. [doi: [10.3233/SHTI190676](https://doi.org/10.3233/SHTI190676)] [Medline: [31438371](https://pubmed.ncbi.nlm.nih.gov/31438371/)]
8. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput Biol* 2013;9(7):e1003153 [FREE Full text] [doi: [10.1371/journal.pcbi.1003153](https://doi.org/10.1371/journal.pcbi.1003153)] [Medline: [23874191](https://pubmed.ncbi.nlm.nih.gov/23874191/)]
9. HL7 FHIR. URL: <https://www.hl7.org/fhir/> [accessed 2020-02-04]
10. Semler S, Wissing F, Heyder R. German Medical Informatics Initiative. *Methods Inf Med* 2018 Jul 17;57(S 01):e50-e56. [doi: [10.3414/me18-03-0003](https://doi.org/10.3414/me18-03-0003)]
11. Winter A, Stäubert S, Ammon D, Aiche S, Beyan O, Bischoff V, et al. Smart Medical Information Technology for Healthcare (SMITH). *Methods Inf Med* 2018 Jul 17;57(S 01):e92-e105. [doi: [10.3414/me18-02-0004](https://doi.org/10.3414/me18-02-0004)]
12. Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, et al. HiGHmed – An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. *Methods Inf Med* 2018 Jul 17;57(S 01):e66-e81. [doi: [10.3414/me18-02-0002](https://doi.org/10.3414/me18-02-0002)]
13. Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn K. Data Integration for Future Medicine (DIFUTURE). *Methods Inf Med* 2018 Jul 17;57(S 01):e57-e65. [doi: [10.3414/me17-02-0022](https://doi.org/10.3414/me17-02-0022)]
14. Prokosch H, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. *Methods Inf Med* 2018 Jul 17;57(S 01):e82-e91. [doi: [10.3414/me17-02-0025](https://doi.org/10.3414/me17-02-0025)]
15. Metzger P, Scheible R, Hess M, Boeker M. MIRACUM-Pipe. URL: <https://github.com/AG-Boerries/MIRACUM-Pipe> [accessed 2020-03-23]
16. Gruendner J, Schwachhofer T, Sippl P, Wolf N, Erpenbeck M, Gulden C, et al. Correction: KETOS: Clinical decision support and machine learning as a service - A training and deployment platform based on Docker, OMOP-CDM, and FHIR Web Services. *PLoS One* 2019;14(11):e0225442 [FREE Full text] [doi: [10.1371/journal.pone.0225442](https://doi.org/10.1371/journal.pone.0225442)] [Medline: [31721815](https://pubmed.ncbi.nlm.nih.gov/31721815/)]
17. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones E, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. *Int J Epidemiol* 2014 Dec;43(6):1929-1944 [FREE Full text] [doi: [10.1093/ije/dyu188](https://doi.org/10.1093/ije/dyu188)] [Medline: [25261970](https://pubmed.ncbi.nlm.nih.gov/25261970/)]
18. Illumina. URL: <https://www.illumina.com/> [accessed 2020-02-02]
19. GitHub. vt, A tool set for short variant discovery in genetic sequence data. URL: <https://github.com/atks/vt> [accessed 2020-02-02]
20. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics* 2015 Jul 01;31(13):2202-2204 [FREE Full text] [doi: [10.1093/bioinformatics/btv112](https://doi.org/10.1093/bioinformatics/btv112)] [Medline: [25701572](https://pubmed.ncbi.nlm.nih.gov/25701572/)]
21. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w<sup>1118</sup>; iso-2; iso-3. *Fly (Austin)* 2012;6(2):80-92 [FREE Full text] [doi: [10.4161/fly.19695](https://doi.org/10.4161/fly.19695)] [Medline: [22728672](https://pubmed.ncbi.nlm.nih.gov/22728672/)]
22. 1000 Genomes | A Deep Catalog of Human Genetic Variation. URL: <https://www.internationalgenome.org/home> [accessed 2020-01-10]
23. Database of Single Nucleotide Polymorphisms (dbSNP). National Center for Biotechnology Information. URL: <https://www.ncbi.nlm.nih.gov/snp/> [accessed 2019-01-10]
24. ClinVar. National Center for Biotechnology Information. URL: <https://www.ncbi.nlm.nih.gov/clinvar/> [accessed 2019-01-10]
25. Project Jupyter | Home. URL: <https://jupyter.org/> [accessed 2020-02-04]
26. Index of /vol1/ftp/phase1/analysis\_results/integrated\_call\_sets/. URL: [ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis\\_results/integrated\\_call\\_sets/](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/integrated_call_sets/) [accessed 2020-02-02]
27. 1000 Genomes Project Consortium T, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012 Nov 01;491(7422):56-65 [FREE Full text] [doi: [10.1038/nature11632](https://doi.org/10.1038/nature11632)] [Medline: [23128226](https://pubmed.ncbi.nlm.nih.gov/23128226/)]
28. Fearon ER. Molecular genetics of colorectal cancer. *Annu Rev Pathol* 2011;6:479-507. [doi: [10.1146/annurev-pathol-011110-130235](https://doi.org/10.1146/annurev-pathol-011110-130235)] [Medline: [21090969](https://pubmed.ncbi.nlm.nih.gov/21090969/)]
29. Bryant KL, Mancias JD, Kimmelman AC, Der CJ. KRAS: feeding pancreatic cancer proliferation. *Trends Biochem Sci* 2014 Feb;39(2):91-100 [FREE Full text] [doi: [10.1016/j.tibs.2013.12.004](https://doi.org/10.1016/j.tibs.2013.12.004)] [Medline: [24388967](https://pubmed.ncbi.nlm.nih.gov/24388967/)]
30. Wolf N, Gruendner J. julianguendner/genomics\_data\_service. URL: [https://github.com/julianguendner/genomics\\_data\\_service](https://github.com/julianguendner/genomics_data_service) [accessed 2020-05-10]
31. mf. Medizininformatik-Initiative. DFZ 2017 Aug 30;61(9):25-25. [doi: [10.1007/s12614-017-6946-6](https://doi.org/10.1007/s12614-017-6946-6)]
32. COSMIC | Catalogue of Somatic Mutations in Cancer. URL: <https://cancer.sanger.ac.uk/cosmic> [accessed 2020-02-02]
33. Home - My Cancer Genome. URL: <https://www.mycancergenome.org/> [accessed 2020-02-02]
34. NIH, US National Library of Medicine. Home - ClinicalTrials.gov. URL: <https://www.clinicaltrials.gov/> [accessed 2020-02-02]



## Abbreviations

**CDM:** common data model  
**CSV:** comma-separated values  
**DIC:** data integration center  
**FHIR:** Fast Healthcare Interoperability Resources  
**GEMINI:** GEnome MINIng  
**GUI:** graphical user interface  
**i2b2:** informatics for integrating biology and the bedside  
**MI-I:** Medical Informatics Initiative  
**MIRACUM:** Medical Informatics in Research and Care in University Medicine  
**NGS:** next-generation sequencing  
**OHDSI:** Observational Health Data Sciences and Informatics  
**OMOP:** Observational Medical Outcomes Partnership  
**SNP:** single nucleotide polymorphism  
**SnPEff:** single nucleotide polymorphism effect  
**SQL:** structured query language  
**VCF:** variant call format

*Edited by G Eysenbach; submitted 07.05.20; peer-reviewed by G Geleijnse, E Chukwu; comments to author 01.06.20; revised version received 26.07.20; accepted 17.08.20; published 07.10.20.*

*Please cite as:*

Gruendner J, Wolf N, Tögel L, Haller F, Prokosch HU, Christoph J  
*Integrating Genomics and Clinical Data for Statistical Analysis by Using GEnome MINIng (GEMINI) and Fast Healthcare Interoperability Resources (FHIR): System Design and Implementation*  
*J Med Internet Res* 2020;22(10):e19879  
URL: <http://www.jmir.org/2020/10/e19879/>  
doi: [10.2196/19879](https://doi.org/10.2196/19879)  
PMID: [33026356](https://pubmed.ncbi.nlm.nih.gov/33026356/)

©Julian Gruendner, Nicolas Wolf, Lars Tögel, Florian Haller, Hans-Ulrich Prokosch, Jan Christoph. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 07.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Application of Big Data Technology for COVID-19 Prevention and Control in China: Lessons and Recommendations

Jun Wu<sup>1</sup>; Jian Wang<sup>2</sup>, Prof Dr; Stephen Nicholas<sup>3,4</sup>, Prof Dr; Elizabeth Maitland<sup>5</sup>, Prof Dr; Qiuyan Fan<sup>1</sup>

<sup>1</sup>Dong Fureng Institute of Economic and Social Development, Wuhan University, Wuhan, China

<sup>2</sup>Dong Fureng Institute of Economic and Social Development, Wuhan University, Beijing, China

<sup>3</sup>Australian National Institute of Management and Commerce, Sydney, Australia

<sup>4</sup>Newcastle Business School, University of Newcastle, Newcastle, Australia

<sup>5</sup>School of Management, University of Liverpool, Liverpool, United Kingdom

## Corresponding Author:

Jian Wang, Prof Dr

Dong Fureng Institute of Economic and Social Development

Wuhan University

No.54 Dongsì Lishi Hutong, Dongcheng District

Beijing, 100010

China

Phone: 86 13864157135

Email: [00032954@whu.edu.cn](mailto:00032954@whu.edu.cn)

## Abstract

**Background:** In the prevention and control of infectious diseases, previous research on the application of big data technology has mainly focused on the early warning and early monitoring of infectious diseases. Although the application of big data technology for COVID-19 warning and monitoring remain important tasks, prevention of the disease's rapid spread and reduction of its impact on society are currently the most pressing challenges for the application of big data technology during the COVID-19 pandemic. After the outbreak of COVID-19 in Wuhan, the Chinese government and nongovernmental organizations actively used big data technology to prevent, contain, and control the spread of COVID-19.

**Objective:** The aim of this study is to discuss the application of big data technology to prevent, contain, and control COVID-19 in China; draw lessons; and make recommendations.

**Methods:** We discuss the data collection methods and key data information that existed in China before the outbreak of COVID-19 and how these data contributed to the prevention and control of COVID-19. Next, we discuss China's new data collection methods and new information assembled after the outbreak of COVID-19. Based on the data and information collected in China, we analyzed the application of big data technology from the perspectives of data sources, data application logic, data application level, and application results. In addition, we analyzed the issues, challenges, and responses encountered by China in the application of big data technology from four perspectives: data access, data use, data sharing, and data protection. Suggestions for improvements are made for data collection, data circulation, data innovation, and data security to help understand China's response to the epidemic and to provide lessons for other countries' prevention and control of COVID-19.

**Results:** In the process of the prevention and control of COVID-19 in China, big data technology has played an important role in personal tracking, surveillance and early warning, tracking of the virus's sources, drug screening, medical treatment, resource allocation, and production recovery. The data used included location and travel data, medical and health data, news media data, government data, online consumption data, data collected by intelligent equipment, and epidemic prevention data. We identified a number of big data problems including low efficiency of data collection, difficulty in guaranteeing data quality, low efficiency of data use, lack of timely data sharing, and data privacy protection issues. To address these problems, we suggest unified data collection standards, innovative use of data, accelerated exchange and circulation of data, and a detailed and rigorous data protection system.

**Conclusions:** China has used big data technology to prevent and control COVID-19 in a timely manner. To prevent and control infectious diseases, countries must collect, clean, and integrate data from a wide range of sources; use big data technology to analyze a wide range of big data; create platforms for data analyses and sharing; and address privacy issues in the collection and use of big data.

**KEYWORDS**

big data; COVID-19; disease prevention and control

**Introduction**

Big data are complex data sets that traditional data processing systems cannot efficiently and economically store, manage, or process. Compared to traditional data, big data has five “V” characteristics: volume, variety, velocity, veracity, and value [1]. In the digital economy era, data are a fundamental strategic resource for countries, enhancing the government’s social governance capacity and public service levels. Big data technology supports a wide range of health care functions, including clinical decision support, population health management, and disease monitoring [2,3]. By discovering correlations in data and understanding patterns and trends, big data technology can improve health care, save lives, and reduce health system costs. Through the analysis of patient characteristics and patient nursing costs, the most clinical cost-effective treatment methods can be determined; the application of big data analysis technology to patient files can identify individuals who may benefit from preventive care or lifestyle changes; the collection and analysis of medical procedure data can determine the most valuable patient nursing programs; and through analysis and drug treatment data, the health status of the population can be monitored and the health status of patients maximized through drug treatments [4]. In the prevention and control of pandemics, large data technology can be used for epidemic prediction, pandemic alerts, tracking and tracing of infected individuals, identifying potential pharmacological treatments, and optimal resource allocations within the health system [5-10].

Big data analytics is fast becoming a crucial component for the modeling of virus transmission, aiding infection control measures and emergency response analyses required during local or international disease outbreaks [11]. In the prevention and control of infectious diseases, previous research on the application of big data technology has mainly focused on the early warning and monitoring of infectious diseases. Four data streams are used for early warning and monitoring infectious diseases: medical health data, participatory syndromic data, internet data, and nonhealth digital data [12]. Medical health data include electronic records of medical institutions, medical insurance claims, discharge records, and death certificates. Such data can provide information about disease conditions and can be monitored at various levels or aggregated by geographic location before reporting [13-15]. Participatory syndromic data are data from crowdsources, where volunteers report a series of symptoms on their own. These data streams do not provide the confirmed infection status of specific pathogens but provide personal-level health data in almost real time [16-18]. Internet data does not depend on a specific patient or medical condition but originates from the use of internet search engines, social media, or online consumption. In addition to self-reporting health results, these data streams also provide information about health-related behaviors, including contact and travel patterns,

and vaccine status, which are key elements to understanding the spread of diseases [19-24]. Nonhealth digital data mainly include social and natural factors such as weather, temperature, humidity, population movement, transportation, infrastructure, and medical environment [25-28]. With the development of computer and space technology, geographic information system remote sensing is increasingly used in infectious disease monitoring and early disease warning research because of its powerful geospatial data acquisition, management, processing, analysis, and display capabilities [29,30].

Before the outbreak of COVID-19, China mainly used infectious disease case data reports for disease early warning and monitoring. In recent years, China has been working to advance the informationization process of medical institutions, store information related to medical services in computer network systems, and accumulate a large amount of medical service data. For example, the hospital information system (HIS) is an important source of medical health data. HIS mainly includes a hospital management information system, laboratory information system (LIS), medical image archive and communication system (picture archiving and communication system), radiation information management system (radiology information system), and clinical decision support system. The electronic medical record system (EMRS) in the medical and health departments includes data on patient name, treatment data, illnesses, test results, orders, operation records, and nursing records.

After experiencing the severe acute respiratory syndrome (SARS) outbreak in 2003, the Chinese government formulated plans to identify the early signs of infectious diseases, setting in law the requirement under the Emergency Regulations for Public Health Emergencies that units that discover infectious diseases must report the disease one level up within a specified time frame. These reports contain information on patient’s name, ID number, age, occupation, residential address, date of disease onset, date of diagnosis, type of infectious disease, and route of transmission. In addition, China has established a service network that links disease control institutions, hospitals, and primary medical and health institutions. Under the National Infectious Disease Report Information Management System (NIDRIMS), institutions immediately report new infectious disease cases online. Those institutions without direct network reporting capabilities were instructed to immediately report the case to the local district-level disease control institution and complete an infectious disease report card within 2 hours. Based on the newly entered data on the direct network reporting system and the historically accumulated infectious disease data, an automatic national infectious disease early warning and surveillance information system was created in April 2008, and an infectious disease monitoring data system trailed in December 2009. For example, a doctor diagnosing a patient with tuberculosis can use the test results to extract the information needed to complete the infectious disease report from the

patient's electronic medical records, which is submitted to the NIDRIMS. NIDRIMS can query the detailed information of the case to monitor and provide an early warning of infectious diseases. The national infectious disease early warning system and surveillance system significantly facilitated the prevention and control of infectious diseases in China. In the H7N9 epidemic in 2013, although the virus spread widely and affected more than ten Chinese provinces and cities, the actual number of infected people was only 132, with 29 deaths and no medical staff infected.

After the outbreak of COVID-19, the Chinese government and social organizations actively used big data technology to prevent and control the disease. At the Scheduling Meeting on Big Data Supporting the Prevention and Control of Corona Virus Disease 2019 held by the Ministry of Industry and Information Technology on January 26, 2020, a joint antiepidemic prevention and control mechanism was proposed, and big data analysis was mobilized to guide research and predict the epidemic's developments, including the deployment of antiepidemic work and monitoring the health of mobile personnel. On February 14, 2020, the government further ordered the use of digital technologies including big data, artificial intelligence (AI), and cloud computing to include epidemic monitoring and analysis, tracing of virus sources, epidemic prevention and treatment, and resource allocation.

Existing papers on COVID-19 big data provide limited discussion and analysis on specific big data applications [10,31]; address the application of big data in a subdivision of COVID-19 prevention and control, and discuss the application of specific types of data in the prevention and control of COVID-19

[6,7,30,32]; or mainly discuss the application methods and achievements of big data technology outside of China [8,9]. Adding to the existing literature, this paper reviews the data collection methods and data information that existed in China before the outbreak of COVID-19. We introduce China's post-COVID-19 data collection methods and data information analyses. Based on the increased data collected, China's experiences in applying big data technology for personnel tracking, epidemic surveillance, early warning, tracing of virus sources, drug screening, medical treatment, resource allocation, and production recovery are detailed. We also analyzed the issues, challenges, and responses encountered by China in the application of big data technology to prevent and control COVID-19 from four perspectives: data access, data use, data sharing, and data protection. Suggestions for further improvements are made from four perspectives, data collection, data circulation, data innovation, and data security, to help understand China's response to the epidemic and to provide lessons for other countries facing the COVID-19 pandemic.

## Methods

By June 2019, the number of mobile internet users in China was 847 million [33]. In recent years, the degree of informatization by the Chinese government and various social organizations has continuously improved. China's data collection methods and data reserves also increased. [Textbox 1](#) summarizes the pre-COVID-19 key data collection methods and data information that helped prevent and control COVID-19, while [Textbox 2](#) provides an outline of the newly added data collection methods and data information implemented after the outbreak of COVID-19.

**Textbox 1.** Pre-COVID-19 disease prevention and control data from government, commercial and public welfare sources and websites, academic papers, public reports, institutional reports, and fieldwork.

**Location information data**

- Telecommunication operators
  - Mobile phone signaling generates location data allowing the user's location to be identified and tracked based on the location of the base station. The acquisition of these data requires the authorization of the Public Security Department. In addition to operators, some government departments, large internet companies, and map service providers can apply for the acquisition of these data.
  - When applying for a mobile phone number, you need to provide the user's name, ID number, residential address, and other information, which can be accessed by the operator and government departments.
- Mobile payment
  - The electronic wallet is bound to the user's ID card and mobile phone number. The payee is usually a store or an enterprise and has location information on the map. Mobile payment is widely used in China.
- Takeaway
  - Can provide the user's residential address
- Online shopping
  - Requires mobile phone number and residential address
- Courier
  - Courier requires customer's ID card number, mobile phone number, and residential address.
- Internet Protocol address
  - Users will leave location information when they go online.

**Travel data**

- Ticket purchase information
  - Purchase of train tickets, plane tickets, ferry tickets, and bus tickets require the user's name, ID number, and mobile phone number information.
- Toll station vehicle information
  - At the entrance and exit of the road, the license plate number is bound to the ID card information.
- Taxi software
  - Online car-hailing software (Didi, Shenzhou special car, Shouqi car-hailing, etc) collect the departure and destination addresses.
- Map data
  - Automatically locates the user's location when running map software
- Card swipe information data
  - The bus and subway provide payment records.

**Medical health data**

- Electronic medical records
  - Includes data on patient name, treatment data, illnesses, test results, orders, operation records, and nursing records
- National Health Commission
  - Name, ID number, age, occupation, residential address, date of onset, date of diagnosis, and type of infectious disease when encountering infectious disease cases

**Internet Data**

- Consumer data
  - Online shopping



- News data
  - Online news media information
- Social data
  - Communicated by the public through social software
- Retrieve data
  - User search history

**Government data**

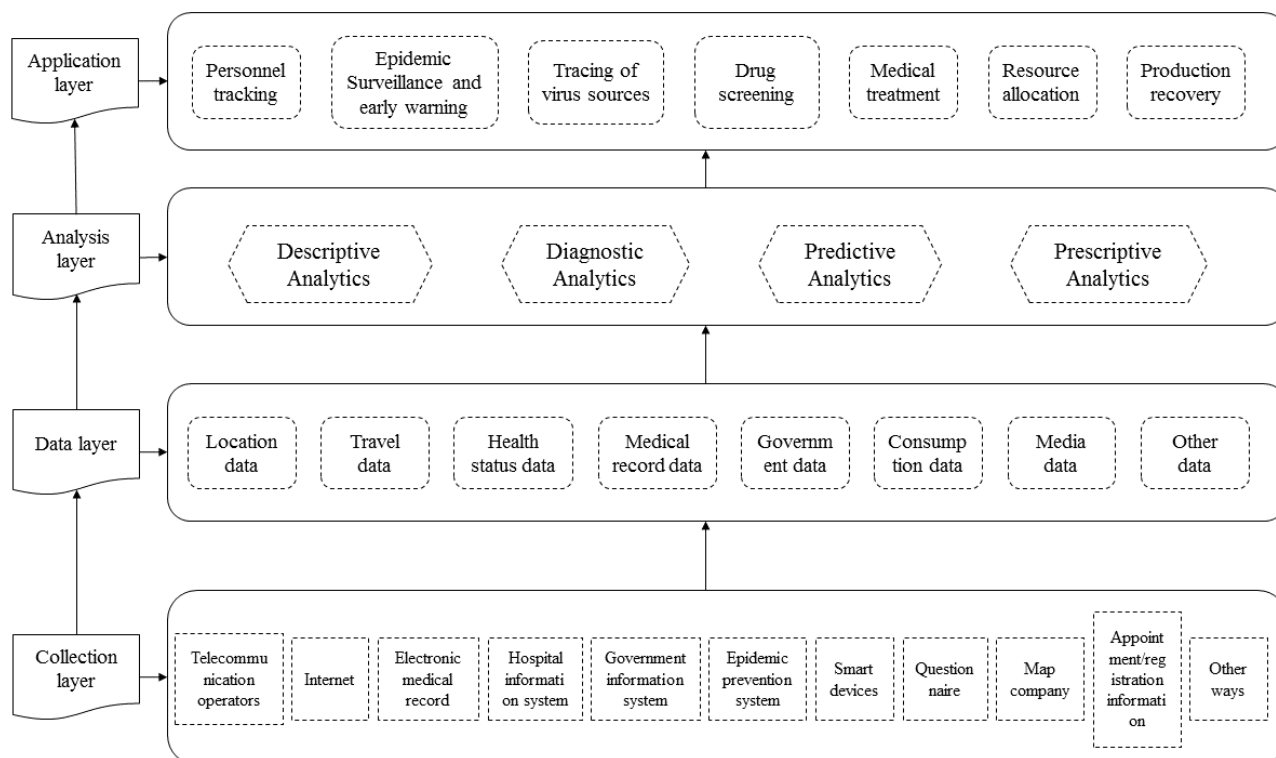
- Tax data
  - Tax department continues to collect
- Electricity data
  - The energy department will continue to collect and record
- Legal data
  - The legal department will continue to collect and record

**Textbox 2.** Newly added data collected in China after the outbreak of COVID-19 from government, commercial and public welfare sources and websites, academic papers, public reports, institutional reports, and fieldwork.

<p><b>Location information data</b></p> <ul style="list-style-type: none"> <li>Questionnaire data from schools and various organizations, including current residential address, through an online questionnaire (WeChat or Alipay applet)</li> </ul> <p><b>Consumption data</b></p> <ul style="list-style-type: none"> <li>When purchasing special medicines (treatment of colds, fever, and coughing, and other medicines), ID number and residential address information</li> </ul> <p><b>Reservation data</b></p> <ul style="list-style-type: none"> <li>To go to hospitals, parks, and other places, you need to use your ID number and mobile phone number to make an appointment</li> </ul> <p><b>Registration data</b></p> <ul style="list-style-type: none"> <li>Temporary access to shopping malls, restaurants, and other public places requires registration of ID card number and mobile phone number</li> </ul> <p><b>Travel data</b></p> <ul style="list-style-type: none"> <li>Questionnaire <ul style="list-style-type: none"> <li>When you travel, you need to provide departure address and destination information.</li> </ul> </li> </ul> <p><b>Patient data</b></p> <ul style="list-style-type: none"> <li>National Health Commission <ul style="list-style-type: none"> <li>Information on confirmed and suspected COVID-19 cases (name, ID card number, age, occupation, residence address, date of onset, date of diagnosis) by the National Health Commission</li> </ul> </li> </ul> <p><b>Health status data</b></p> <ul style="list-style-type: none"> <li>Questionnaire survey <ul style="list-style-type: none"> <li>Daily report of health status (temperature and the presence of cold, fatigue, coughing symptoms) on time every day by internet applet</li> </ul> </li> <li>Drug purchase information <ul style="list-style-type: none"> <li>When purchasing special drugs (treatment of colds, fever, and coughing, and other drugs) data on name, ID, and address, and data on body temperature and cold, fatigue, and coughing symptoms</li> </ul> </li> <li>Smart devices <ul style="list-style-type: none"> <li>Thermal imaging human body temperature measurement devices in public places such as train stations, airports, shopping malls, hospitals, and pharmacies automatically collect data.</li> </ul> </li> </ul> <p><b>Treatment data</b></p> <ul style="list-style-type: none"> <li>Hospital <ul style="list-style-type: none"> <li>Statistics on the treatment effect on patients with COVID-19 treated with different drugs</li> </ul> </li> </ul> <p><b>Prevention and control materials data</b></p> <ul style="list-style-type: none"> <li>The government department <ul style="list-style-type: none"> <li>Supply and demand information, price, and quantity of various epidemic prevention materials into a system platform</li> </ul> </li> </ul> <p><b>Data of designated</b></p> <ul style="list-style-type: none"> <li>Hospitals, government, and map suppliers <ul style="list-style-type: none"> <li>Government's listed designated hospitals have special identification label on map</li> </ul> </li> </ul>
---

Figure 1 sets out a diagrammatic representation of the conceptual basis of big data technology for COVID-19 prevention and control on four levels: application, analysis, data, and collection.

In the next section, we discuss the key elements within each collection, data, analysis, and application layer of Figure 1.

**Figure 1.** Conceptual structure chart of prevention and control of COVID-19 with big data technology.

## Results

### Location Data and Travel Data

Turning to the collection, data, analysis, and application layers in Figure 1, the scale of population migration is crucial for disease space propagation predictions, risk area identification, and control measure decisions to control infectious diseases [34]. Studies have shown that there is a close relationship between the number of train trips and the number of COVID-19 cases [35]. Before Wuhan was quarantined on January 23, 2020, more than 5 million people left Wuhan for other parts of China [36]. Internet companies used ticket purchase information and toll booth information to create population migration maps. The maps show the movement of population between cities in specific time ranges. For January 22, 2020, the day before the closure of Wuhan, the data found that 74.77% of the total number of people who moved out of Wuhan migrated to other parts of Hubei Province including 14.56% in Xiaogan City, 14.08% in Huanggang City, and 7.65% in Jingzhou City; 5.67% moved to Henan Province; and 3.24% moved to Hunan Province [37]. In the early stage of the epidemic, the travel data identified cities and regions susceptible to potential future outbreaks, allowing government organizations to make timely preparations for epidemic prevention and control.

According to the patient information data provided by the National Health Commission, internet companies and government official websites produced epidemic maps showing the number of patients diagnosed with COVID-19, the number of suspected patients, and the number of cured patients in each province and city [38]. The death toll allowed the public to keep abreast of the epidemic's development. For example, based on a unified geographic framework, research institutions quickly

absorbed and integrated geographic big data including internationally published World Health Organization data, daily family health and disease control data, professional population health platform data, Tencent site selection request data, Baidu migration data, patient spatial-temporal trajectory data, international airline data, census data, remote sensing images, and other multisource data. Multi-scale comprehensive spatial-temporal dynamic visualization technology allowed provinces, cities, counties, communities, and individual epidemic data to be unified in a spatial-temporal data visualized on a multidimensional "one map." The map service provider identified the residential areas with confirmed or suspected cases on the map, and the map users can interrogate the information around their specific locations on the map. Individuals can avoid activities in high-risk areas, and hospitals that can receive suspected cases are highlighted on the map so that people can find COVID-19 treatment hospitals. Map service providers also used mobile phone signaling data to measure the density of people in specific areas and differentiate the display on the map through red, yellow, and green color differences to remind the public of areas to avoid to reduce the risk of infection.

Based on the card swipe information data, systems developed by the public transportation management departments allowed passenger flow and full load of subways and buses data to be queried in a timely manner to select the best travel arrangements and times, to avoid congestion, and to reduce the risk of infection. The epidemic prevention department can quickly confirm the flight, train, or bus information of travelers infected with COVID-19 within the past 2 weeks and inform fellow travelers who have been in close proximity to be tested and take self-isolation measures. To inform the traveling public about high-risk travel, internet companies entered the information of

flights, trains, or buses used by patients with COVID-19 into a tracking system and developed a “close contact measurement instrument.” By entering their ID number and name, a traveler can immediately check whether they have taken the same flight, train, or bus as a patient with COVID-19.

Through epidemiological surveys based on questionnaires, the staff of the epidemic prevention department can identify the areas and the specific times patients with COVID-19 visited before becoming sick and can promptly notify others who were active at the same time and place to pay attention to their health. For example, if a confirmed case occurs in a hospital, the hospital can notify other patients and people entering the hospital to isolate themselves through their mobile phone number. The operator’s mobile phone signaling data can timely and accurately identify the country where the user stayed in the previous 14 days and the domestic city where the user stayed for more than 4 hours to determine whether the user has been in a key epidemic area or country.

Most provinces and cities in China have established a big data health code system. When travelers scan their official Quick Response (QR) code, register, and fill in personal relevant information (including name; ID number; mobile phone number; body temperature; cities and travels in the past 14 days; and whether they have symptoms such as fever, coughing, and fatigue), the system will automatically generate a personal “health QR code,” which is divided into three color levels of red, yellow, and green [39]. Green means that the person did not appear in the virus-infected area during the quarantine period and can carry out cross-regional movement. When people with green codes have been to high-risk areas and in contact with high-risk people, their code will turn red. People with red or yellow codes will be quarantined, and the codes will turn green when certain conditions are met.

The digital epidemic prevention system, jointly developed by Alibaba’s Bodhidharma, Dingding, Alipay, and Alibaba Cloud, summarizes information from the hospital’s diagnosed cases and identifies people who have purchased fever-reducing drugs in pharmacies within the past month. Big data technology is then used to analyze the activity trajectories of confirmed cases and close contacts, and an epidemic spread model is constructed in combination with the positioning system. Based on the mobile phone positioning system of the diagnosed cases and close contacts, the expert team can find other mobile phone numbers within a 3-meter transmission range and more than a specified contact time based on the physical distance between the mobile phones to analyze the infection risks between parties. The big data positioning information can identify a large number of close contacts and quickly establish a province-wide big data outbreak early warning mechanism. For example, major epidemic transmission centers have been identified, including the Baodi Mall, a Tianjin hospital, the Tulong shopping mall in Harbin, and the Yintaidao shopping mall in Wenzhou [30].

## Medical and Health Data

For epidemic prevention and control, big data in health care can promote the timely detection and reporting of cases, improve the probability of finding diagnosis and treatment methods quickly, and improve the efficiency of hospital management in

a pressure environment. There were initial failures in fully identifying and reporting COVID-19 cases in the EMRS disease reporting system. EMRS information should be complete and accurate. In the early stages of the epidemic, frontline doctors were not efficient in collecting patient data, identifying patients’ conditions, and reporting infectious diseases. One way of addressing this problem was to link EMRS data to epidemic decision making through China’s Center for Disease Control and Prevention (CDC), the government’s public welfare institution responsible for the technical management of disease control and public health [40]. By connecting the CDC’s monitoring and early warning system to hospitals’ EMRS, big data technology was applied in a timely manner to extract and analyze medical big data. First, through automatic matching, CDC’s AI knowledge base interrogated EMRS for keywords such as pneumonia. When suspected hospitals cases were identified, the EMRS immediately monitored front-end doctors’ computers, prompting doctors to verify the completeness and accuracy of the EMRS information and generate an infectious disease report. Such monitoring systems attenuate concealing or underreporting infections. Second, CDC monitoring improved the discovery and testing of EMRS big data to ensure the effective transmitting of health data and to avoid late or underreporting. For example, the CDC monitoring system reduced the average time required for doctors to report a COVID-19 case from the previous 5-8 minutes to within 40 seconds and the time required for online reporting via the CDC web-based infectious disease reporting system from 2-3 minutes to a few seconds [41]. CDC monitoring also supported COVID-19 diagnosis both within China and worldwide. One example was the successful isolation of the first COVID-19 virus strain in China, with the publication of the virus’ scientific information, electron micrographs, specific primers, and probes for global use. In the early stages of COVID-19 in China, the failure to provide timely data allowed the virus to spread both locally and worldwide. In the middle stages of the disease, timely release of scientific resources and data provided COVID-19 genes and gene sets, and big data information for COVID-19 research platforms for Chinese and international researchers [42]. CDC monitoring also supported hospital management. For example, the CDC promoted the integration of EMRS, HIS, and LIS data, and the hospital operations data to manage protective supplies, health status reports, epidemic developments, and telecommuting. By analytically and visually processing the integrated data, managers made decisions and assessments, which supported hospital treatment regimens [40]. When medical resources were overwhelmed by patients, big data technology helped develop a hierarchical diagnosis and treatment system, which allocated scarce health resources efficiently. In epidemic prevention and control, big data predictive analytics were mainly applied to predicting the impact of epidemic developments on medical resources. Based on big data models such as communication dynamics and risk level distribution of the incidence rate and close contacts, predictive analytics estimated epidemic peaks and inflection points, which allowed the differential allocation of resources to regional hospitals. Big data searches on internet platforms and susceptible-exposed-infectious-removed (SEIR) transmission modeling can predict COVID-19 transmission trends [43]. Using

population migration data to fill in the dynamic propagation SEIR model combined with AI methods trained with SARS data, Yang et al [44] predicted a COVID-19 pandemic curve. The authors showed that if the Chinese authorities had postponed the implementation of strict public health measures for 5 days, the scale of the epidemic would have tripled. A loosening or removal of lockdown interventions would have caused Hubei Province to peak again from mid-March to late April [44].

Based on the predicted epidemic trends and risk level information, health departments deployed prevention resources in advance of the epidemic in specific areas to contain the spread of the virus and avoid second outbreaks. Relevant government departments also used big data predictive analytics to identify the epidemic's peak and inflection point to determine the approximate time to resume normal work. Finally, big data predictive analytics was applied to integrate HIS medical costs and insurance information to estimate epidemic trends, proportion of patients of different histology types, the cost of diagnosis and treatment, and the allocation of health resources.

Further, big data diagnostic analytics can screen existing clinical pneumonia drugs to treat patients with COVID-19. For example, big data on patients prescribed traditional Chinese medicine were used to explore the efficacy of Chinese medicinal materials and their composition to treat COVID-19. Zhang et al [45] systematically screened natural compounds used in Chinese treatment and found 13 of them to exert potential anti-COVID-19 benefits in terms of the regulation of viral replication, modulation of immune and inflammatory pathways, and hypoxia cascade.

### News Media and Social Data

During the outbreak, there were inaccurate statements and misleading information on the internet, such as rumors that high alcohol content drinks and gargling with salt water could resist COVID-19, drinking Banlangen and smoked vinegar could prevent COVID-19, and that children could not be infected with COVID-19 [46]. Such false information will lead the public to relax their vigilance and take incorrect COVID-19 epidemic prevention measures. After identifying this type of false COVID-19 information, internet companies alerted epidemic prevention experts and national authoritative epidemic prevention organizations to take action to correct such false or inaccurate information. China took action to correct false information through authoritative national media, short films, and live broadcasts [47].

When a major epidemic occurs, the negative impact of uncertainty and panic on social activities may exceed the negative impact of viral diseases. Social media data can track and evaluate the spatial spread of public sentiment. By tracking topics and sentiments that appeared in microblog users' timelines in response to COVID-19, a user semantic behavior evolution model was introduced to measure and analyze changes in public opinion [30,43]. The results indicated that from January 9 to February 10, 2020, more than 60% of posts related to science popularization of disease prevention and government COVID-19 announcements were positive and stable. For example, the posts with the topic of "help-seeking" were concentrated in the key epidemic area of Wuhan, and the posts

related to "donation information" were widely distributed throughout the country [30].

Through real-time monitoring of keyword information searched by users, such as the frequency of the words "fever," "epidemic," "cough," "pneumonia," and "infection" on internet platforms, health officials assessed the health status of people in various regions. When the frequency of such keyword searches increased rapidly in a certain area, the risk of a large-scale outbreak in that area was increased accordingly [5]. Such tracking allowed relevant health departments to formulate epidemic prevention and treatment measures in advance. For example, Qin et al [43] used big data to predict the number of suspicious or confirmed new cases of COVID-19. Using a series of lagging "social media search indexes" for various keywords including clinical symptoms of COVID-19 (such as dry cough, fever, chest pain, and pneumonia), the authors found that COVID-19 outbreaks could be found 6-9 days in advance.

### Data Collected by Intelligent Equipment

Using thermal imaging human body temperature measurement equipment in public places such as train stations, airports, shopping malls, hospitals, and pharmacies to automatically collect data, individuals with an abnormal body temperature in a large number of passing people can be identified using face recognition technology [48]. Once identified, people with COVID-19 can be isolated, containing the spread of the virus. This big data early warning system can automatically collect and analyze information to diagnose people with COVID-19 in real time.

### Data on Epidemic Prevention Materials

The spatial distribution of medical resources is usually balanced according to factors such as population density, but the uneven spatial outbreak and spread of COVID-19 creates large regional imbalances in the supply and demand for medical resources. The key to epidemic prevention and control is to understand the spatial and temporal dynamics of the supply and demand of medical resources to optimize the distribution of materials, resources, and medicine. Sharing data with emergency suppliers and public resource national trading institutions underpins a system that provides information on supply and demand for resources and maintains unified prices to minimize unfair competition, ensure material security, and guarantee material quality [49]. Through telephone and online inquiries, the dynamic status of medical protection equipment in China was analyzed. Data analytics on material needs identified the demand from multiple locations and allowed health organizations to coordinate the release and tracking of medical supplies by their name and quantity of required materials, contact information, and transport mode [50]. By stopping the release of material supplies through scattered, uncoordinated channels, big data allowed the managed allocation of supplies to hospitals and cities with the greatest need, balanced supply and demand, and ensured the efficient allocation of medical resources.

A stable and efficient national material supply and transportation system provides important support for successful epidemic prevention work. For example, using provincial epidemiological data, online consumption data and postal service data identifies



each provinces' supply and demand status for necessities and food, preventing shortages and price changes [51]. At the same time, by tracking the transportation of materials, highly sensitive nodes that may cause virus transmission during transportation were identified, and early warning and decision support were provided to prevent and control the regional spread of COVID-19.

### Online Consumer Data

Big data can help enterprises to target production, marketing, logistics, and safe work resumption. Amid the epidemic, enterprises producing medical products and basic necessities can assess public demand using big data to ensure supply and demand equilibrium by scheduling optimal production and efficient distribution of products. Big data, for example, identified electric pressure cookers as one of the most talked about products during the epidemic, which allowed firms to target the production and distribution of pressure cookers [52].

### Government Data

During the epidemic, the use of big data technology for population information management and decision making was a fundamental approach by government service departments [53]. Tax data, especially value-added tax invoice data, have the unique advantages of wide coverage, strong timeliness, and fine granularity, helping to schedule the resumption of work, production, and the operation of enterprises [54]. Gathering data on industry, region, scale, and economic type, national and local governments analyzed the dynamics of the economy's operation, planned the resumption of work, determined the production levels and sales of enterprises, identified and solved problems along the industrial supply chain, and provided targeted assistance to individual business and business sectors. For example, the Shandong Provincial Taxation Administration used big data to help a local pharmaceutical glass company to distribute their epidemic prevention supplies by screening logistics companies that had a high degree of credibility and carrier capacity, ensuring the smooth delivery of their supplies [55].

Through big data analysis, governments can assess the impact of the epidemic on socioeconomic operations in the near and long term; establish economic emergency response systems; implement various incentive measures such as tax cuts, fee reductions, and special subsidies; mitigate the risk of financial chain breakage; and promote the continuous operation of small- and medium-sized enterprises. For example, the Shanghai Municipal Tax Administration used tax data to identify potential targets for tax incentives and subsequently launched more than 30 tax incentive programs, covering more than 200,000 tax companies [56].

### Deep Learning Data

Used in a variety of fields, deep learning is a machine learning analytic method frequently used to interrogate big data, performing tasks that are difficult for conventional analysis methods [57-60]. Mobilizing big data, deep learning offers a method to better forecast disease trends, including COVID-19. For Hong Kong, Xu et al [61] showed deep learning yielded better prediction performance for flu-like diseases than the

generalized linear model, the least absolute shrinkage and selection operator model, and the autoregressive integrated moving average (ARIMA) model. Wang et al [62] found deep learning provided a convenient tool for fast screening COVID-19 and finding potential high-risk patients, which may be helpful for medical resource optimization and early prevention before patients show severe COVID-19 symptoms. Various non-China studies indicate that infectious disease can be predicted more effectively when large amounts of data including weather variables and internet big data are used to predict infectious diseases. Using deep learning, Chae et al [63] found that deep neural network and long-short term memory learning models outperformed ARIMA approaches when predicting for three infectious diseases in Korea 1 week into the future. Using data mining algorithms to predict COVID-19 outbreak trends, Ayyoubzadeh et al [8] showed that the most effective predictive factors, besides the previous day incidence, was the search frequency of handwashing, hand sanitizer, and antiseptic topics. Togacar et al [64] found a deep learning model using data on COVID-19, pneumonia, and normal x-ray imaging could detect COVID-19 efficiently.

## Discussion

### Issues and Challenges in COVID-19 Big Data

#### Data Collection and Access

During a pandemic, the way data are collected and the quality of the collected data pose challenges. When data were collected by government workers exposed to disease environments, causing physical and mental dangers, data standards were not uniform. For organizations, different information reporting systems and data reporting formats were adopted, resulting in inconsistent standards, unreliable structures, and untimely reporting, especially in the process of statistical aggregation [65]. Frequently, data were not updated in a timely manner. Against the backdrop of the rapid spread of the COVID-19 virus, immediate data updates were essential to assess the state of the epidemic and its dangers, and to avoid panic among the public. However, due to insufficient awareness of the epidemic's risk in the early stages, the lack of advanced means of data collection, the large number of subjects involved in data collection and release, and the inevitable "fragmentation" of big data platforms, much of the data were not updated timely or optimally.

#### Data Use

Data use depends heavily on visualization, and data mining techniques were found inadequate. Data source, quality, and scale all influence data integration, and raw data collection issues contribute to the difficulty of data integration. Converting low-value density big data into high-value density knowledge through data integration is the purpose of big data management applications [66]. Given existing data development capabilities, there has been a general lack of effective integration of data sources. This has been complicated by the rapid development of internet and information technology, and the rapidly expanding types and volume of data resources. Data quality and degree of data analysis determines big data's role in epidemic

prevention and control. Much of the data used for epidemic prevention and control were available through official channels or accumulated by enterprises during their operations but were of widely varying quantity, quality, and variety. It is often the case that data holders do not have sufficient skills to exploit good data, or those with good data development skills do not have sufficient data.

The visual presentation of spatial, temporal, and quantitative features of epidemic data in the form of maps is the most common means of using epidemic big data. Epidemic surveillance is about spatial governance through spatial-temporal modeling and analysis, including the control of people flow, logistics, information flow, and technology flow. Although geographic information technology has unique advantages in data collection, processing, and management, most visualization products simply present, rather than analyze, information about the epidemic. In the immediate COVID-19 period, there were insufficient data on transportation networks, service facilities, public opinion, prevention and control effects, and the information on their mutual relationship. More importantly, there was inadequate innovation in the way the data were used.

### **Data Sharing**

The key to the role of digital government and smart cities in managing major public emergencies is to enable data sharing. In 2015, China's State Council issued the *Action to Promote the Development of Big Data* [67], which proposed that big data should be used as an important means to enhance the government's governance capacity, improve the level of government decision making, and manage risk prevention through efficient collection and integration of government and social data. By the end of 2017, China had introduced a series of policies to promote effective government data sharing and completed the construction of a national data sharing platform to promote the sharing of data resources at the provincial and ministerial levels. In practice, data sharing has displayed a silo effect. Data sharing among Chinese departments and regional governments remains inadequate, and the phenomenon of "block islands" still exists. For example, the health code is a two-dimensional risk assessment code generated from personal travel information and health status information. During the use of these health codes, some regions did not recognize the health codes generated by other regions [39]. Partly, this is due to inadequate infrastructure for urban data collection, lack of community data resources, and lack of big data management expertise, which led to a disconnect between spatial and social governance. However, there was a continuing unwillingness to share data across provinces and municipalities, and insufficient means of information interaction, resulting in inaccurate data analysis and weak regional joint prevention capacity. Only by fully realizing data flow across regions, levels, and sectors can data be efficiently integrated and better applied to epidemic management.

### **Data Protection**

In terms of data protection, there is inadequate privacy protection and undisclosed compliance risks. The value of big data lies in its disclosure and concrete application. In the age of the internet and electronic information, people's lifestyles

generate big data on all aspects of their life, but privacy may be compromised. In the face of the rapidly spreading epidemic, government investigates and closely monitors the health of individuals, collecting a large amount of personal information on COVID-19 confirmed, suspected, and potential cases and their close contacts. Information on people leaving Wuhan to visit their home villages and cities have been exposed on the internet, which caused a serious negative impact on individuals and families who were subjected to unwarranted harassment, discrimination, and physical threats [68].

## **Strategies and Recommendations**

### **Data Collection and Access**

From China's big data experience during the first stages of the COVID-19 pandemic, data collection should have been more timely, comprehensive, and accurate. To improve data collection and access, electronic questionnaires should be promoted to conduct information surveys, including the use of the WeChat applet or Alipay applet. For COVID-19 tracking, after scanning their QR code, individuals can efficiently input their information such as ID card number, household registration, residential address, and itinerary of the past 2 weeks, and then update their body temperature every day. Of course, manual collection of data can lead to delays, concealment, inaccuracies, and underreporting. One solution is to collect these data automatically by apps, which will improve data accuracy and use. Accurate tracking data can be collected using portable electronic equipment, cameras, access control, license plates, and other information.

To conduct more in-depth scientific analyses, data on weather, news, transportation, business, and health are currently collected in addition to collecting accurate personal information data. The credibility of these data is important. For example, when the mobile phone number registered according to an individual's ID card information is used by others, there will be errors in predicting a person's movements. In addition, individuals deliberately concealing their own information or falsely reporting their own information leads to data errors. Data sharing will allow cross-comparisons of different data sources, including mutual authentication of government, operator, traffic, public security, network, and other information. For example, the National Health Commission and the national information security department could work together to formulate common standards of data structure, format, attributes, content, and scope for individuals, and release these data to the institutions or departments with investigation rights to ensure the quality of the data. Cross-comparisons of data raise issues of privacy and whether intentionally concealing or falsely reporting data should be legally punished.

### **Data Use and Sharing**

Data use and sharing require the integration of data sets and the strengthening of data sharing. One way to integrate data is to give full play to the technical advantages of scientific research institutions or commercial units. For example, the functions of the Guiyang big data exchange center, Shanghai Data Trading Center, and commercial data package trading platforms (such as COFCO, Youyi, Yaxin, and shuduoduo) can integrate, use,

and share data. In addition, one way to innovate data use is to use blockchain technology to build a neutral third-party security housing platform to realize the separation of ownership and use rights [69]. Secure house platforms allow the demand side to use data that it does not own. For the data supplier, secure house platforms provide massive data access to multiple data sources. For the data demander, it uses data fusion technology to provide multiparty high-quality data resources that the data demander did not have previously. In addition, it supports the third-party algorithm and user-defined algorithms. For example, AI companies can carry out AI analysis and algorithm modeling for pathological sections of patients in the secure house, which not only ensures patients' privacy and prevents the leakage of medical data but also optimizes the use of the AI. To fully realize big data sharing in a secure way, governments can set an identity for big data assets and establish a data query mechanism to realize cross-platform data tracing and extraction.

### **Data Innovation**

The value and efficiency of big data use can be improved by making full use of big data mining and data analysis methods. Based on big data resources and big data analysis methods such as machine learning and data simulation, a more comprehensive and rapid assessment of the risk and impact of the epidemic could be carried out. Strengthening data collaboration, for example, between data owners and big data technology platforms, and integrating multidisciplinary knowledge and analysis methods will yield innovative big data solutions. The health code is a successful example. Information on individuals' ID, residential address, health status, whether they have been to the infected area, and contact history of key personnel were entered into the applet and a QR code generated. Spatially, the risk level of the epidemic in defined areas were assessed temporally by the number of visits, length of stay, and interpersonal relations between individuals in epidemic areas, allowing three risk states (red, green, and yellow) to be measured, analyzed, and assessed [39].

A further innovation would be the expanded use of the QR code, where individuals would be required to scan their QR code to show their health code status before entering public places. At the same time, the health code would also record the current position of the individual. When their QR code is red or yellow, they would be denied entry to public areas. In addition, when there are COVID-19 cases in specific public areas, the health department can quickly inform individuals not to enter or to isolate through information automatically reported by their health code. In addition to obtaining basic health information of residents during the epidemic, the health code can obtain medical insurance information and other disease information, realize a series of digital services in the hospital system from registering appointments to picking up medicines, and fully exploit new opportunities for online appointments and online consultations while accumulating data assets for the analysis of medical and health big data intelligence applications.

### **Data Security and Privacy**

Attention should be paid to big data privacy issues [70]. Privacy issues are fundamentally country-specific legal requirements, where China's privacy framework will vary from that of other

countries. Guarantees of nondisclosure and proper use of personal privacy information are the most basic requirements for data collection and use. Statistical processes involving the collection, aggregation, sharing, and disclosure of personal information should require the protection of personal information to prevent data leakage, loss, and abuse. Governments should establish institutional safeguards, listing situations that require the public to declare private information data. The government should also specify the list of institutions that can request personal privacy information and the requirements allowing third-party access to personal data. In China, the law on the prevention and control of infectious diseases stipulates that "any unit or individual within the territory of China must accept the prevention and control measures such as the investigation, inspection, collection of samples, isolation and treatment of infectious diseases by the Department of Disease Prevention and Control, and truthfully report the relevant information" [71]. On February 9, 2020, China issued a notice protecting personal information and regulating the use of big data to support joint prevention and control of diseases, specifying that personal information collected for epidemic and disease prevention and control should not be used for other purposes. No unit or individual may disclose information such as name, age, ID card number, telephone number, or home address without the consent of the collector, except for those who have been desensitized due to the need for epidemic prevention and control [72]. Government departments only release information of public concern, such as "digitizing" individual patients. Each confirmed patient has a code, which only shows their gender, age, disease severity, and frequent activity area. There are legal penalties for failing to follow the legal requirements on data handling.

Privacy concerns might be further addressed by enhancing big data's privacy framework, including setting access and use permissions for the encrypted data collected, making the data available only to institutions or organizations with a clear purpose, ensuring only the minimum personal information for epidemic prevention and disease control is collected, and setting retention periods that limit the time collected personal information is retained.

Big data helped contain, suppress, and control COVID-19. The main lesson is that COVID-19 big data technology can be applied to epidemic alerts, tracking the spread of viruses in real time, planning public health interventions, monitoring the effectiveness of treatment regimens, identifying potential vaccine candidates, strengthening community and regional government responses to epidemics, and guiding social and economic recovery in the postepidemic period. In the long run, big data technology helps build smart, healthy, and resilient cities and countries.

### **Conclusion**

In China, big data technology displayed its potential in containing, suppressing, and controlling COVID-19, with the Chinese government and nongovernmental organizations making full use of big data information from the major areas of medical institutions, telecommunication operators, internet companies, government departments, commercial companies, and AI



equipment. Through the analysis of big data technology to prevent and control COVID-19, we found that big data technology played an important role in epidemic surveillance and early warning, tracing of virus sources and personal tracking, planning public health interventions, monitoring drug screening and medical treatment, strengthening community and regional government responses to epidemics, and guiding social and economic recovery in the postepidemic period. These big data applications were central for the public, enterprises, and governments in their response to COVID-19.

We also identified gaps in the use of big data technology to prevent and control COVID-19, including problems in data acquisition, data use, data circulation, and data protection. The application of big data technology in the prevention and control of infectious diseases like COVID-19 requires further development in the collection, analysis, distribution, and privacy around personal and public information. Although China's big data application to fighting the COVID-19 epidemic was a success story, it also provides lessons for the continuous improvement in the application of big data technology to epidemics in China and for other countries managing the COVID-19 pandemic.

## Conflicts of Interest

None declared.

## References

- Sharma S, Mangat V. Technology and trends to handle big data: survey. 2015 Presented at: 2015 Fifth International Conference on Advanced Computing & Communication Technologies; Feb 21-22 2015; Haryana, India. [doi: [10.1109/acct.2015.121](https://doi.org/10.1109/acct.2015.121)]
- Burghard C. Big data and analytics key to accountable care success. IDC Health Insights. 2012. URL: <https://www.coursehero.com/file/14345607/Big-Data-and-Analytics-Key-to/>
- Feldman B, Martin E. Big data in healthcare hype and hope. Semantic Scholar. 2012. URL: [https://pdfs.semanticscholar.org/241e/3e0657ff97d5e2beafaf5eb87085eef40cf8.pdf?\\_ga=2.219768095.740413039.1595810682-1144616737.1595400014](https://pdfs.semanticscholar.org/241e/3e0657ff97d5e2beafaf5eb87085eef40cf8.pdf?_ga=2.219768095.740413039.1595810682-1144616737.1595400014)
- Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;2:3 [FREE Full text] [doi: [10.1186/2047-2501-2-3](https://doi.org/10.1186/2047-2501-2-3)] [Medline: [25825667](https://pubmed.ncbi.nlm.nih.gov/25825667/)]
- Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Feb 19;457(7232):1012-1014. [doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634)] [Medline: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)]
- Qiu H, Yuan L, Wu Q, Zhou Y, Zheng R, Huang X, et al. Using the internet search data to investigate symptom characteristics of COVID-19: a big data study. *World J Otorhinolaryngol Head Neck Surg* 2020 May 19;1 [FREE Full text] [doi: [10.1016/j.wjorl.2020.05.003](https://doi.org/10.1016/j.wjorl.2020.05.003)] [Medline: [32837757](https://pubmed.ncbi.nlm.nih.gov/32837757/)]
- Guraya SY. Transforming laparoendoscopic surgical protocols during the COVID-19 pandemic; big data analytics, resource allocation and operational considerations. *Int J Surg* 2020 Aug;80:21-25 [FREE Full text] [doi: [10.1016/j.ijssu.2020.06.027](https://doi.org/10.1016/j.ijssu.2020.06.027)] [Medline: [32590073](https://pubmed.ncbi.nlm.nih.gov/32590073/)]
- Ayyoubzadeh S, Ayyoubzadeh S, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: data mining and deep learning pilot study. *JMIR Public Health Surveill* 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: [10.2196/18828](https://doi.org/10.2196/18828)] [Medline: [32234709](https://pubmed.ncbi.nlm.nih.gov/32234709/)]
- Haleem A, Javaid M, Khan I, Vaishya R. Significant applications of big data in COVID-19 pandemic. *Indian J Orthop* 2020 May 07;1-3 [FREE Full text] [doi: [10.1007/s43465-020-00129-z](https://doi.org/10.1007/s43465-020-00129-z)] [Medline: [32382166](https://pubmed.ncbi.nlm.nih.gov/32382166/)]
- Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How big data and artificial intelligence can help better manage the COVID-19 pandemic. *Int J Environ Res Public Health* 2020 May 02;17(9):3176 [FREE Full text] [doi: [10.3390/ijerph17093176](https://doi.org/10.3390/ijerph17093176)] [Medline: [32370204](https://pubmed.ncbi.nlm.nih.gov/32370204/)]
- Garattini C, Raffle J, Aisyah DN, Sartain F, Kozlakidis Z. Big data analytics, infectious diseases and associated ethical impacts. *Philos Technol* 2019;32(1):69-85 [FREE Full text] [doi: [10.1007/s13347-017-0278-y](https://doi.org/10.1007/s13347-017-0278-y)] [Medline: [31024785](https://pubmed.ncbi.nlm.nih.gov/31024785/)]
- Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. *J Infect Dis* 2016 Dec 01;214(suppl\_4):S375-S379 [FREE Full text] [doi: [10.1093/infdis/jiw400](https://doi.org/10.1093/infdis/jiw400)] [Medline: [28830113](https://pubmed.ncbi.nlm.nih.gov/28830113/)]
- Simonsen L, Gog JR, Olson D, Viboud C. Infectious disease surveillance in the big data era: towards faster and locally relevant systems. *J Infect Dis* 2016 Dec 01;214(suppl\_4):S380-S385 [FREE Full text] [doi: [10.1093/infdis/jiw376](https://doi.org/10.1093/infdis/jiw376)] [Medline: [28830112](https://pubmed.ncbi.nlm.nih.gov/28830112/)]
- HealthMap. 2016. URL: <https://healthmap.org/about/>
- Chowell G, Cleaton JM, Viboud C. Elucidating transmission patterns from internet reports: Ebola and Middle East respiratory syndrome as case studies. *J Infect Dis* 2016 Dec 01;214(suppl\_4):S421-S426 [FREE Full text] [doi: [10.1093/infdis/jiw356](https://doi.org/10.1093/infdis/jiw356)] [Medline: [28830110](https://pubmed.ncbi.nlm.nih.gov/28830110/)]
- Guerrisi C, Turbelin C, Blanchon T, Hanslik T, Bonmarin I, Levy-Bruhl D, et al. Participatory syndromic surveillance of influenza in Europe. *J Infect Dis* 2016 Dec 01;214(suppl\_4):S386-S392. [doi: [10.1093/infdis/jiw280](https://doi.org/10.1093/infdis/jiw280)] [Medline: [28830105](https://pubmed.ncbi.nlm.nih.gov/28830105/)]

17. MacFadden DR, Fisman D, Andre J, Ara Y, Majumder MS, Bogoch II, et al. A platform for monitoring regional antimicrobial resistance, using online data sources: ResistanceOpen. *J Infect Dis* 2016 Dec 01;214(suppl\_4):S393-S398. [doi: [10.1093/infdis/jiw343](https://doi.org/10.1093/infdis/jiw343)] [Medline: [28830108](#)]
18. Salathé M. Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health. *J Infect Dis* 2016 Dec 01;214(suppl\_4):S399-S403 [FREE Full text] [doi: [10.1093/infdis/jiw281](https://doi.org/10.1093/infdis/jiw281)] [Medline: [28830106](#)]
19. Zhang Y, Milinovich G, Xu Z, Bambrick H, Mengersen K, Tong S, et al. Monitoring pertussis infections using internet search queries. *Sci Rep* 2017 Sep 05;7(1):10437. [doi: [10.1038/s41598-017-11195-z](https://doi.org/10.1038/s41598-017-11195-z)] [Medline: [28874880](#)]
20. Rohart F, Milinovich GJ, Avril SMR, Lê Cao KA, Tong S, Hu W. Disease surveillance based on internet-based linear models: an Australian case study of previously unmodeled infection diseases. *Sci Rep* 2016 Dec 20;6:38522. [doi: [10.1038/srep38522](https://doi.org/10.1038/srep38522)] [Medline: [27994231](#)]
21. Lamos V, Miller AC, Crossan S, Stefansen C. Advances in nowcasting influenza-like illness rates using search query logs. *Sci Rep* 2015 Aug 03;5:12760. [doi: [10.1038/srep12760](https://doi.org/10.1038/srep12760)] [Medline: [26234783](#)]
22. Cho S, Sohn CH, Jo MW, Shin S, Lee JH, Ryoo SM, et al. Correlation between national influenza surveillance data and google trends in South Korea. *PLoS One* 2013;8(12):e81422 [FREE Full text] [doi: [10.1371/journal.pone.0081422](https://doi.org/10.1371/journal.pone.0081422)] [Medline: [24339927](#)]
23. Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, et al. Dynamic forecasting of zika epidemics using Google Trends. *PLoS One* 2017;12(1):e0165085 [FREE Full text] [doi: [10.1371/journal.pone.0165085](https://doi.org/10.1371/journal.pone.0165085)] [Medline: [28060809](#)]
24. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. *PLoS One* 2013;8(2):e56176 [FREE Full text] [doi: [10.1371/journal.pone.0056176](https://doi.org/10.1371/journal.pone.0056176)] [Medline: [23457520](#)]
25. Song Y, Wang F, Wang B, Tao S, Zhang H, Liu S, et al. Time series analyses of hand, foot and mouth disease integrating weather variables. *PLoS One* 2015;10(3):e0117296 [FREE Full text] [doi: [10.1371/journal.pone.0117296](https://doi.org/10.1371/journal.pone.0117296)] [Medline: [25729897](#)]
26. Hii YL, Rocklöv J, Ng N. Short term effects of weather on hand, foot and mouth disease. *PLoS One* 2011 Feb 11;6(2):e16796 [FREE Full text] [doi: [10.1371/journal.pone.0016796](https://doi.org/10.1371/journal.pone.0016796)] [Medline: [21347303](#)]
27. Lopman B, Armstrong B, Atchison C, Gray JJ. Host, weather and virological factors drive norovirus epidemiology: time-series analysis of laboratory surveillance data in England and Wales. *PLoS One* 2009 Aug 24;4(8):e6671 [FREE Full text] [doi: [10.1371/journal.pone.0006671](https://doi.org/10.1371/journal.pone.0006671)] [Medline: [19701458](#)]
28. Huang X, Williams G, Clements ACA, Hu W. Imported dengue cases, weather variation and autochthonous dengue incidence in Cairns, Australia. *PLoS One* 2013;8(12):e81887 [FREE Full text] [doi: [10.1371/journal.pone.0081887](https://doi.org/10.1371/journal.pone.0081887)] [Medline: [24349148](#)]
29. Allen C, Tsou M, Aslam A, Nagel A, Gawron J. Applying GIS and machine learning methods to Twitter data for multiscale surveillance of influenza. *PLoS One* 2016;11(7):e0157734 [FREE Full text] [doi: [10.1371/journal.pone.0157734](https://doi.org/10.1371/journal.pone.0157734)] [Medline: [27455108](#)]
30. Zhou C, Su F, Pei T, Zhang A, Du Y, Luo B, et al. COVID-19: challenges to GIS with big data. *Geography Sustainability* 2020 Mar;1(1):77-87. [doi: [10.1016/j.geosus.2020.03.005](https://doi.org/10.1016/j.geosus.2020.03.005)]
31. Lin L, Hou Z. Combat COVID-19 with artificial intelligence and big data. *J Travel Med* 2020 Aug 20;27(5):taaa080 [FREE Full text] [doi: [10.1093/jtm/taaa080](https://doi.org/10.1093/jtm/taaa080)] [Medline: [32437541](#)]
32. Million M, Gautret P, Colson P, Roussel Y, Dubourg G, Chabriere E, et al. Clinical efficacy of chloroquine derivatives in COVID-19 infection: comparative meta-analysis between the big data and the real world. *N Microbes N Infections* 2020 Jun:100709. [doi: [10.1016/j.nmni.2020.100709](https://doi.org/10.1016/j.nmni.2020.100709)]
33. The 45th China statistical report on internet development. China Internet Network Information Center. 2020. URL: [http://www.cac.gov.cn/2020-04/27/c\\_1589535470378587.htm](http://www.cac.gov.cn/2020-04/27/c_1589535470378587.htm)
34. Wang Y, Dong L, Liu Y, Huang Z, Liu Y. Migration patterns in China extracted from mobile positioning data. *Habitat Int* 2019 Apr;86:71-80. [doi: [10.1016/j.habitatint.2019.03.002](https://doi.org/10.1016/j.habitatint.2019.03.002)]
35. Zhao S, Zhuang Z, Ran J, Lin J, Yang G, Yang L, et al. The association between domestic train transportation and novel coronavirus (2019-nCoV) outbreak in China from 2019 to 2020: a data-driven correlational report. *Travel Med Infect Dis* 2020;33:101568 [FREE Full text] [doi: [10.1016/j.tmaid.2020.101568](https://doi.org/10.1016/j.tmaid.2020.101568)] [Medline: [32006656](#)]
36. Mayor of Wuhan: more than 5 million people left Wuhan. News China. 2020. URL: [http://news.china.com.cn/2020-01/26/content\\_75650784.htm](http://news.china.com.cn/2020-01/26/content_75650784.htm)
37. Baidu Migration. URL: <https://qianxi.baidu.com/2020/>
38. Real-time big data report of the epidemic. Baidu. URL: [https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari\\_pc\\_3](https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_pc_3)
39. The big data behind the health code is revealed. Today Science and Technology. 2020. URL: <http://qikan.cqvip.com/Qikan/Article/Detail?id=7101192711>
40. Let big data be the early warning of the epidemic-all aspects of big data in the application of disease control. China Net Medical Channel. 2020 Mar 13. URL: <http://med.china.com.cn/content/pid/164670/tid/1026>
41. Jiao Y. Shuangshuo Technology. 2020 Mar 28. URL: [https://www.sohu.com/a/383736468\\_133140](https://www.sohu.com/a/383736468_133140)
42. COVID-19 strain. China Center for Disease Control and Prevention. 2020 Jan 24. URL: <http://nmcdc.cn/coronavirus>



43. Qin L, Sun Q, Wang Y, Wu K, Chen M, Shia B, et al. Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. *Int J Environ Res Public Health* 2020 Mar 31;17(7):2365 [FREE Full text] [doi: [10.3390/ijerph17072365](https://doi.org/10.3390/ijerph17072365)] [Medline: [32244425](https://pubmed.ncbi.nlm.nih.gov/32244425/)]
44. Yang Z, Zeng Z, Wang K, Wong S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* 2020 Mar;12(3):165-174. [doi: [10.21037/jtd.2020.02.64](https://doi.org/10.21037/jtd.2020.02.64)] [Medline: [32274081](https://pubmed.ncbi.nlm.nih.gov/32274081/)]
45. Zhang D, Wu K, Zhang X, Deng S, Peng B. In silico screening of Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus. *J Integr Med* 2020 Mar;18(2):152-158 [FREE Full text] [doi: [10.1016/j.joim.2020.02.005](https://doi.org/10.1016/j.joim.2020.02.005)] [Medline: [32113846](https://pubmed.ncbi.nlm.nih.gov/32113846/)]
46. Banlangen was snapped up after the outbreak of Wuhan pneumonia! Expert interpretation. Shuangshuo Technology. 2020 Jan 03. URL: [https://www.sohu.com/a/364568004\\_693031](https://www.sohu.com/a/364568004_693031)
47. Details, help seeking and rumor refutation: mainstream media promote the reporting and dissemination of epidemic situation by short video. Shuangshuo Technology. 2020 Feb 04. URL: [https://www.sohu.com/a/370518635\\_120325](https://www.sohu.com/a/370518635_120325)
48. Epidemic prevention measures of subway equipment inspection personnel under COVID-19 epidemic and intelligent inspection prospect. Shanxi Science and Technology. 2020. URL: [http://yuxiqbs.cqvip.com/Qikan/Article/detail?id=7101902137&from=Qikan\\_Search\\_Index](http://yuxiqbs.cqvip.com/Qikan/Article/detail?id=7101902137&from=Qikan_Search_Index)
49. Huang J, Cao F. China's epidemic prevention and control emergency procurement behavioral economics elaboration and enlightenment. *Chinese Government Procurement* 2020;3:76-80 [FREE Full text]
50. Questionnaire on community health and resource allocation management during the epidemic. Wenjuan.com. URL: <https://www.wenjuan.com/slp/UvEvErb/>
51. 43 logistics and supply chain management experts make suggestions for the fight against the new. Zibo Comprehensive Logistics Information Platform. 2020 Feb 11. URL: <http://www.zbwlb.net/plus/view.php?aid=2173>
52. Report of home improvement users during the COVID-19 epidemic. Leju Finance. 2020 Mar 05. URL: <http://news.leju.com/2020-03-05/6641239750554729875.shtml>
53. Zhou X, Chen L. Improving government public health emergency management capabilities driven by big data: inspirations of local government population information management during the New Coronary Pneumonia epidemic. *Price Theory Pract* 2020 Apr 08;1. [doi: [10.19851/j.cnki.CN11-1010/F.2020.02.001](https://doi.org/10.19851/j.cnki.CN11-1010/F.2020.02.001)]
54. Xing D. The topic of 'Stabilizing the economy against the epidemic' (5) Taxation responsibility for stabilizing the economy under the new coronary pneumonia epidemic. *Guangdong Economics* 2020;3:28-31 [FREE Full text]
55. Make good use of tax data to empower the fight against the epidemic. Shandong Provincial Tax Service, State Tax Administration. 2020 Mar 05. URL: [http://shandong.chinatax.gov.cn/art/2020/3/5/art\\_23\\_174298.html](http://shandong.chinatax.gov.cn/art/2020/3/5/art_23_174298.html)
56. Working together to inject tax power and tax wisdom to fight against the epidemic. Shanghai Taxation Bureau. 2020 Mar 16. URL: <https://baijiahao.baidu.com/s?id=1661305448361223482&wfr=spider&for=pc>
57. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *J Big Data* 2015 Feb 24;2(1):1. [doi: [10.1186/s40537-014-0007-7](https://doi.org/10.1186/s40537-014-0007-7)]
58. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J Pathol Inform* 2016;7:29 [FREE Full text] [doi: [10.4103/2153-3539.186902](https://doi.org/10.4103/2153-3539.186902)] [Medline: [27563488](https://pubmed.ncbi.nlm.nih.gov/27563488/)]
59. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017 Jan 25;542(7639):115-118. [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)]
60. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* 2018 Feb 21;8(1):3395. [doi: [10.1038/s41598-018-21758-3](https://doi.org/10.1038/s41598-018-21758-3)] [Medline: [29467373](https://pubmed.ncbi.nlm.nih.gov/29467373/)]
61. Xu Q, Gel YR, Ramirez Ramirez LL, Nezafati K, Zhang Q, Tsui K. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. *PLoS One* 2017;12(5):e0176690 [FREE Full text] [doi: [10.1371/journal.pone.0176690](https://doi.org/10.1371/journal.pone.0176690)] [Medline: [28464015](https://pubmed.ncbi.nlm.nih.gov/28464015/)]
62. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J* 2020 Aug;56(2):2000775 [FREE Full text] [doi: [10.1183/13993003.00775-2020](https://doi.org/10.1183/13993003.00775-2020)] [Medline: [32444412](https://pubmed.ncbi.nlm.nih.gov/32444412/)]
63. Chae S, Kwon S, Lee D. Predicting infectious disease using deep learning and big data. *Int J Environ Res Public Health* 2018 Jul 27;15(8):1596 [FREE Full text] [doi: [10.3390/ijerph15081596](https://doi.org/10.3390/ijerph15081596)] [Medline: [30060525](https://pubmed.ncbi.nlm.nih.gov/30060525/)]
64. Toğaçar M, Ergen B, Cömert Z. COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Comput Biol Med* 2020 Jun;121:103805 [FREE Full text] [doi: [10.1016/j.compbiomed.2020.103805](https://doi.org/10.1016/j.compbiomed.2020.103805)] [Medline: [32568679](https://pubmed.ncbi.nlm.nih.gov/32568679/)]
65. Wu Q, Yang L. Zhejiang experience and suggestions on the application of big data in epidemic prevention and control. *Zhejiang Economy* 2020;3:54-55 [FREE Full text]
66. Jin C, Chen J, Liu W, Zhang Z. Sharing, integration and integration of government governance big data. *Big Data* 2020;6(2):27-40 [FREE Full text]
67. The State Council released the "Outline of Action to Promote the Development of Big Data". China Paper NewsLetter. 2015. URL: [http://yuxiqbs.cqvip.com/Qikan/Article/detail?id=666542015&from=Qikan\\_Search\\_Index](http://yuxiqbs.cqvip.com/Qikan/Article/detail?id=666542015&from=Qikan_Search_Index)

68. Many cases of leaking epidemic information were notified: what did they leak? Ifeng News. 2020 Mar 27. URL: <http://feng.ifeng.com/c/7vAM1eli8QQ>
69. Think In Cloud 2018. Bag Event. URL: <https://www.bagevent.com/event/1250836>
70. Ioannidis JPA. Informed consent, big data, and the oxymoron of research that is not research. *Am J Bioeth* 2013;13(4):40-42. [doi: [10.1080/15265161.2013.768864](https://doi.org/10.1080/15265161.2013.768864)] [Medline: [23514395](https://pubmed.ncbi.nlm.nih.gov/23514395/)]
71. Law of the People's Republic of China on prevention and control of infectious diseases. China.com. 2004. URL: [http://www.china.com.cn/law/flfg/txt/2006-08/08/content\\_7063888.htm](http://www.china.com.cn/law/flfg/txt/2006-08/08/content_7063888.htm)
72. Notice on the protection of personal information and the use of big data to support joint prevention and control. Office of the Central Cybersecurity and Informatization Committee. 2020 Feb 10. URL: [http://www.gov.cn/xinwen/2020-02/10/content\\_5476711.htm](http://www.gov.cn/xinwen/2020-02/10/content_5476711.htm)

## Abbreviations

**AI:** artificial intelligence

**ARIMA:** autoregressive integrated moving average

**CDC:** Center for Disease Control and Prevention

**EMRS:** electronic medical record system

**HIS:** hospital information system

**LIS:** laboratory information system

**NIDRIMS:** National Infectious Disease Report Information Management System

**QR:** Quick Response

**SARS:** severe acute respiratory syndrome

**SEIR:** susceptible-exposed-infectious-removed

*Edited by G Eysenbach; submitted 30.06.20; peer-reviewed by N Bragazzi, T Igbe; comments to author 16.07.20; revised version received 28.07.20; accepted 14.09.20; published 09.10.20.*

*Please cite as:*

Wu J, Wang J, Nicholas S, Maitland E, Fan Q

Application of Big Data Technology for COVID-19 Prevention and Control in China: Lessons and Recommendations

*J Med Internet Res* 2020;22(10):e21980

URL: <http://www.jmir.org/2020/10/e21980/>

doi: [10.2196/21980](https://doi.org/10.2196/21980)

PMID: [33001836](https://pubmed.ncbi.nlm.nih.gov/33001836/)

©Jun Wu, Jian Wang, Stephen Nicholas, Elizabeth Maitland, Qiuyan Fan. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 09.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Interactive Web-Based Resource for Annotation of Genetic Variants Causing Hereditary Angioedema (HADA): Database Development, Implementation, and Validation

Alejandro Mendoza-Alvarez<sup>1</sup>, MSc; Adrián Muñoz-Barrera<sup>2</sup>, BSc; Luis Alberto Rubio-Rodríguez<sup>2</sup>, BSc; Itahisa Marcelino-Rodríguez<sup>1</sup>, PhD; Almudena Corrales<sup>1,3</sup>, CLT; Antonio Iñigo-Campos<sup>2</sup>, CLT; Ariel Callero<sup>4</sup>, MD, PhD; Eva Perez-Rodriguez<sup>4</sup>, MD; Jose Carlos Garcia-Robaina<sup>4</sup>, MD; Rafaela González-Montelongo<sup>2</sup>, PhD; Jose Miguel Lorenzo-Salazar<sup>2</sup>, MSc; Carlos Flores<sup>1,2,3,5</sup>, PhD

<sup>1</sup>Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain

<sup>2</sup>Genomics Division, Instituto Tecnológico y de Energías Renovables, Santa Cruz de Tenerife, Spain

<sup>3</sup>CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain

<sup>4</sup>Allergy Unit, Hospital Universitario Nuestra Señora de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain

<sup>5</sup>Instituto de Tecnologías Biomédicas, Universidad de La Laguna, Santa Cruz de Tenerife, Spain

**Corresponding Author:**

Carlos Flores, PhD

Research Unit

Hospital Universitario Nuestra Señora de Candelaria

Universidad de La Laguna

Carretera del Rosario s/n

Santa Cruz de Tenerife, 38010

Spain

Phone: 34 922602938

Email: [cflores@ull.edu.es](mailto:cflores@ull.edu.es)

## Abstract

**Background:** Hereditary angioedema is a rare genetic condition caused by C1 esterase inhibitor deficiency, dysfunction, or kinin cascade dysregulation, leading to an increased bradykinin plasma concentration. Hereditary angioedema is a poorly recognized clinical entity and is very often misdiagnosed as a histaminergic angioedema. Despite its genetic nature, first-line genetic screening is not integrated in routine diagnosis. Consequently, a delay in the diagnosis, and inaccurate or incomplete diagnosis and treatment of hereditary angioedema are common.

**Objective:** In agreement with recent recommendations from the International Consensus on the Use of Genetics in the Management of Hereditary Angioedema, to facilitate the clinical diagnosis and adapt it to the paradigm of precision medicine and next-generation sequencing-based genetic tests, we aimed to develop a genetic annotation tool, termed Hereditary Angioedema Database Annotation (HADA).

**Methods:** HADA is built on top of a database of known variants affecting function, including precomputed pathogenic assessment of each variant and a ranked classification according to the current guidelines from the American College of Medical Genetics and Genomics.

**Results:** HADA is provided as a freely accessible, user-friendly web-based interface with versatility for the entry of genetic information. The underlying database can also be incorporated into automated command-line stand-alone annotation tools.

**Conclusions:** HADA can achieve the rapid detection of variants affecting function for different hereditary angioedema types, and further integrates useful information to reduce the diagnosis odyssey and improve its delay.

(*J Med Internet Res* 2020;22(10):e19040) doi:[10.2196/19040](https://doi.org/10.2196/19040)

**KEYWORDS**

genetic cause; hereditary angioedema; knowledge database; precision medicine; variant interpretation

## Introduction

Hereditary angioedema (HAE) is a rare genetic disease caused by an increase of vascular permeability, generating recurrent acute swelling episodes commonly localized on the face, trunk, and extremities. HAE can also be life-threatening when the upper airways or the tissues from the oral cavity are affected [1,2]. Because of its nonspecific clinical signs, HAE is poorly recognized by physicians. Therefore, delayed or ambiguous diagnoses are common, increasing the risk of patient morbidity and mortality [3]. With exceptions for particular subtypes, the recent international consensus guidelines recommend the use of genetic testing to reach a definitive HAE diagnosis in clinical practice [4]. However, with increasing recognition that the genetic cause of HAE is more complex than previously anticipated [5], the tasks involved in the identification of the genetic defect in each patient are increasingly demanding [6].

With the decreasing cost and increasing sequencing throughput of next-generation sequencing (NGS), the screens for gene defects can now be accomplished through a simultaneous evaluation of gene sets, the exome, or the whole genome as indicated [7]. However, access to either high-performance computational equipment and trained bioinformatics personnel or, alternatively, to paywalled cloud-based all-inclusive solutions is the most limiting factor for NGS to be efficiently used in clinical settings [5]. To facilitate the identification of pathogenic variants in HAE patients, Kalmár et al [8] developed HAEdb, an online locus-specific database to centralize the information of the genetic alterations, which allows researchers to retrieve mutation information and contribute new detected variants. However, HAEdb focuses only on the most frequently affected gene (*SERPING1*, encoding the C1 esterase inhibitor [C1-INH]) and uses a data retrieval scheme based on matching with the existing records, requiring the user to prioritize the most likely variant affecting function. Besides, HAEdb has not been updated to include the pathogenic classification according to the standard guidelines established by the American College of Medical

Genetics and Genomics (ACMG) [9], which would allow for standardization between laboratories and studies, while guiding global efforts for improving NGS-based diagnosis to move toward the precision medicine paradigm. More recent efforts such as that of Ponard et al [10] have aimed to fill this gap and updated the mutational spectrum in two of the known HAE genes (*SERPING1* and *F12*) based on the Leiden Open Variant Database (LOVD) v3.0, a platform-independent framework for the maintenance and curation of a web-based database of genetic variants [11]. Nevertheless, these resources are ill-adapted to current NGS technologies and to the evolving knowledge of the genetic causes of HAE.

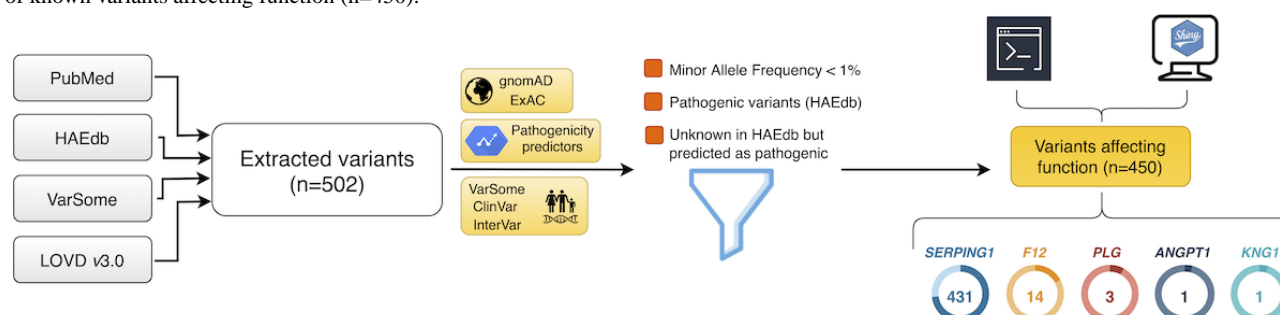
To fill this gap, we here present the Hereditary Angioedema Database Annotation (HADA) tool, a freely accessible, user-friendly, and versatile web-based interface to facilitate the identification of the genetic variants causing HAE.

## Methods

### Gene and Variant Extraction to Build the Database

We retrieved the variants found among HAE cases and relatives from HAEdb [8] (accessed June 18, 2020) and LOVD v3.09 (Figure 1). These databases contain hundreds of records for variants in the *SERPING1* or *F12* genes detected in clinical studies of families with one or more members affected by HAE. Additionally, a search was performed on January 1, 2020 in PubMed for the terms “angioedema” and “mutation” with the aim of retrieving other HAE genes from the literature. From this search, studies focusing on acquired forms of angioedema were ruled out, and an in-depth analysis of each prioritized study was performed. The VarSome database [12] was also screened to retrieve variants, as well as to update the genetic nomenclature of variant descriptors. The information described in the corresponding articles where each variant was described was manually inspected to verify that the original descriptions were accurate.

**Figure 1.** Schematic representation of the steps involved in gene and variant extraction, annotation, and database curation. All genetic variants reported in the articles studying hereditary angioedema (HAE) families in PubMed, HAEdb, VarSome, and the Leiden Open Variation Database (LOVD) were collected (n=502). ANNOVAR was used for annotation of frequencies, pathogenic predictors, and pathogenic classifications, among other information. Variants with a minor allele frequency below 1%, declared as pathogenic in HAEdb or as unknown but with a pathogenic prediction were kept as the set of known variants affecting function (n=450).



Given that a confident clinical interpretation of structural variants is challenging and that the standards for reporting those involving copy number variants have only recently been set [13], the criteria indicated above were only applied to single nucleotide variants (SNVs) and small insertion/deletions (indels). Therefore, the 45 gross mutations (ie, structural variants

in the *SERPING1* gene) from HAEdb that have been identified in HAE families were not integrated into HADA.

### Database Variant Annotation

According to their chromosomal coordinates, the selected variants were first adapted to the GRCh37/hg19 reference



genome, exonic locations, Human Genome Variation Society (HGVS) nomenclature, coding effect, and PubMed citation. ANNOVAR v18.04.16 [14] was used to annotate according to RefGene, the allele frequency in gnomAD v2.1.1 [15] and ExAC [16] (November 29, 2015 release), dbSNP build 150 information, and precalculated pathogenicity predictors (SIFT, PolyPhen2, MutationTaster, CADD, DANN, MetaSVM, LRT, and phastCons mammalian). Pathogenic probabilities according to ClinPred [17] and the ACMG pathogenic classification as determined by ClinVar [18] (March 5, 2019 release), InterVar [19] (January 18, release), and VarSome (accessed June 13, 2020) were also annotated (Figure 1).

### Database Curation

Many original descriptions of the genetic studies in HAE did not clearly declare the causality of the reported variants, thereby increasing the difficulty to identify variants with effects on the disease as opposed to variants without a disease effect [4]. To facilitate the interpretation of HADA results, we imposed the following filters for a variant to be designated as a variant affecting function (Figure 1): (1) variants described as pathogenic in HAEdb; (2) variants with unknown effects in HAEdb but predicted as pathogenic, likely pathogenic, or of uncertain significance (VUS) by VarSome, ClinVar, or InterVar (a flag will advise users in the case that contradictory classification information is reported for a variant); and (3) a reported minor allele frequency <1% in gnomAD, either on the

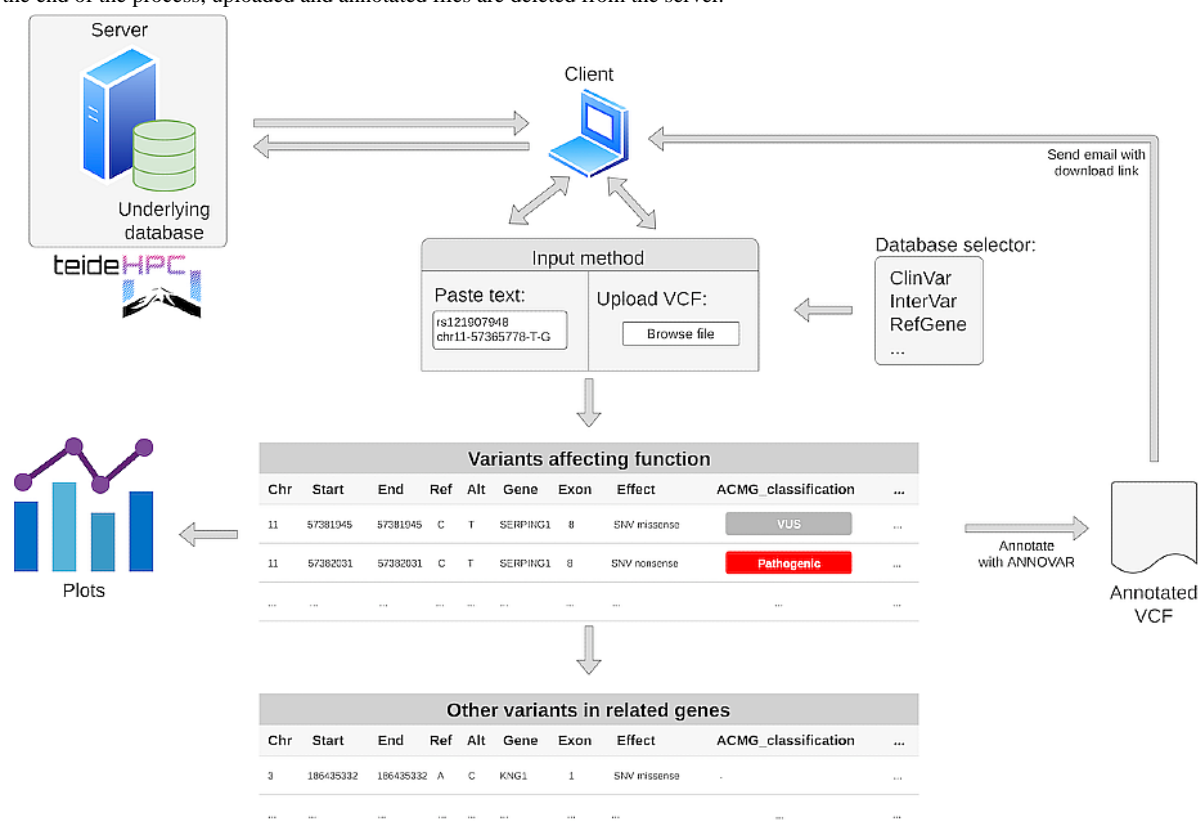
overall sample or for non-Finish Europeans (because most of the genetic studies in HAE to date have focused on European families).

Finally, to facilitate the identification of potentially novel variants in the user's uploaded data, HADA also includes all SNVs and small indels that were not classified as a variant affecting function using the criteria indicated above from dbSNP build 150, ClinVar, and InterVar located within 50-bp flanking regions of all exons from known HAE genes.

### Implementation of HADA

The database is built on MongoDB v4.2.1, a NoSQL query engine, to speed up database user queries and the variant calling file (VCF)-oriented analysis (Figure 2). HADA is built in Shiny v1.3.2, an R v3.6.1 package (R Foundation for Statistical Computing, Vienna, Austria) for building web apps. Specifically, we used ShinyJS v1.0 to run JavaScript code within the web app frontend and Plotly v4.9.0 to generate interactive plots. ANNOVAR v18.04.16 is used to provide annotations from the database in the uploads. To preserve potentially sensitive information included in the uploaded VCFs, HADA uses an encrypting code based on Cryfa [20] that automatically secures the access to the file and decrypts it once returned to the user. This provides a high level of security and transfers the data control to the user. No sensitive sample information is stored or maintained in the server.

**Figure 2.** Schematic representation of the HADA architecture and user interface. HADA is hosted on a server at TeideHPC, through a UI-based on Shiny. When the user sends a query to the server, it is encrypted and interrogates the HADA database, and returns matched variants affecting function associated with hereditary angioedema. The app also searches the data for other variants located within known hereditary angioedema genes to facilitate the identification of potentially novel variants. Once finished, an email with the link to download the annotated variant calling file (VCF) is sent to the user. At the end of the process, uploaded and annotated files are deleted from the server.





The curated database used by HADA is also available as a separate download at github [21] and it has been configured to annotate VCFs using ANNOVAR from the command line interface tool so that it can be easily incorporated into routine standalone NGS bioinformatics workflows. The HADA web interface is publicly accessible [22] hosted on TeideHPC premises [23].

### Validation in an HAE Family

To validate HADA with comprehensive NGS data, we generated whole-exome sequencing (WES) data from a patient with HAE and his two family members recruited by the Allergy Service from the Hospital Universitario Nuestra Señora de Candelaria (HUNSC), Santa Cruz de Tenerife, Spain. The study was approved by the HUNSC Ethics Committee and written informed consent was obtained from the patients. The index patient was a 31-year-old male who visited the emergency department more than 5 times due to acute angioedema attacks with facial and cutaneous symptoms, manifesting as episodes affecting the upper airways from the age of 23 years. His biochemical blood test showed normal levels of C1-INH (47 mg/dL) based on the N Antisera to Human Coagulation Factors and C1 Inhibitor kit (Siemens Healthcare Diagnostics, Marburg) and reduced C1-INH activity (10%) (Berichrom C1 inhibitor, Siemens Healthcare Diagnostics, Marburg, Germany), suggesting HAE type II. His mother (60 years old) and a sister (23 years old) reported no symptoms of HAE, and also consented to participate in the study.

Briefly, sequencing libraries were prepared from DNA extracted using a commercial column-based kit (GFX kit, GE Healthcare, Little Chalfont, UK) using Nextera DNA Exome Kit (Illumina Inc, San Francisco, CA). The TapeStation 4200 system (Agilent Technologies, Santa Clara, CA) was used for library size estimation and the concentration was determined by the Qubit dsDNA HS Assay (Thermo Fisher Scientific, Waltham, MA). Libraries were sequenced on a HiSeq 4000 Sequencing System (Illumina Inc) with paired-end 75-base reads. Libraries were sequenced along with 1% of PhiX control V3 to an average depth of 50X after removal of duplicate reads. Sequencing reads were preprocessed with bcl2fastq v2.18 and mapped to hg19/GRCh37 with the Burrows-Wheeler Aligner 0.7.15-r1140 [24]. Resulting BAM files were processed with Qualimap v2.2.1

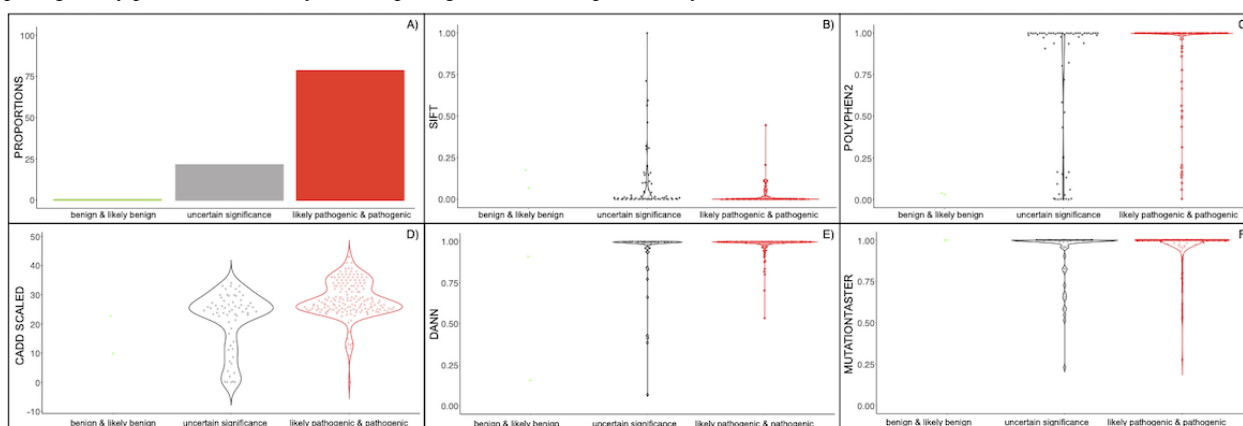
[25], SAMtools v1.3 [26], and Picard v2.10.10 [27]. Variant calling was performed using HaploTypeCaller following the Best Practices workflow recommendations for germline variant calling in GATK (v3.8) [28], obtaining one VCF for each sequenced individual. A high confidence set of variants was obtained after applying the following filters: variants with PASS, read depth $\geq$ 20, genotype quality $\geq$ 100, mapping quality $\geq$ 50. The transition-to-transversion nucleotide substitution ratio was also obtained as a quality control for the refined set of variants. Sequencing data and variant identification were obtained by the Instituto Tecnológico y de Energías Renovables (Santa Cruz de Tenerife, Spain).

## Results

### Overall Characteristics of the Database

A total of 502 variants from 5 genes causally linked to HAE were retrieved from PubMed, HAEdb, VarSome, and LOVD. A subset of 450 variants was designated as variants affecting function, including variants reported in the following 5 genes: *SERPING1* (n=431), *F12* (n=14), *PLG* (n=3), *ANGPT1* (n=1), and *KN1* (n=1) (Figure 1). Three of these genes were discovered in the last 2 years based on WES approaches in HAE patients without C1-INH defects but not carrying *F12* variants affecting function [29-31]. ClinVar offered very limited information on this set of variants affecting function, as only 34 of them (7.6%) had corresponding ACMG class assignment: 2 are reported as benign or likely benign, 6 are indicated as VUS, and 26 are classified as pathogenic and likely pathogenic. InterVar included information for half of the set (226/450, 50.2%). However, 171 (75.7%) of these were classified as VUS. VarSome was the only resource that allowed assigning ACMG classes to all retrieved variants affecting function. According to VarSome, most of the HAE variants affecting function are classified as pathogenic (183/450, 40.6%) or likely pathogenic (171/450, 38.0%) (Figure 3). Although VarSome did not classify any of the HAE variants affecting function as benign or likely benign, 96 of them (21.3%) were still reported as VUS. Besides, precalculated pathogenicity predictors were available for a mean of 243 of the variants affecting function in the database. Taken together, these results highlight the existing gap in current interpretations of variant pathogenicity [32,33].

**Figure 3.** Precalculated pathogenic scores by the American College of Medical Genetics and Genomics (ACMG) pathogenic class for variants affecting function recorded in HADA. Panel A: VarSome proportions of ACMG classes among the hereditary angioedema variants affecting function. Panels B–F: pathogenicity prediction scores by ACMG pathogenic classes as provided by VarSome.



## HADA Interface and Usage

HADA has been developed as a user-friendly graphical web-interface tool. It is designed with the objective of enabling the analysis of genetic variants regardless of the detection approach used for the screening of the genetic causes of HAE (Figure 4). As such, HADA is compatible with individual VCFs that have been obtained at any scale (gene panel, exome, or whole genome) and by any NGS technology. Alternatively, the user could opt to provide hg19 chromosome coordinates of the variant(s) of interest as the rsID number, HGVS coordinate, or amino acid change, among others. The latter is still a common standard in laboratories relying on Sanger sequencing results. All of the annotation information used to construct the underlying database can be selected by the user to extract the available information. To facilitate a prospective analysis of each query variant, references to the corresponding articles reporting the variant are also offered (in chronological order). Direct links to external tools (ie, VarSome), which allows accessing graphical information of gene transcripts, complement the possibilities of HADA. Despite the fact that sensitive patient information is not necessary for HADA to provide results, HADA integrates an automatic data encryption algorithm for the queries based on VCF data, ensuring a password-secured encryption of the input data while uploaded and of the decryption when results are downloaded. Furthermore, the information provided by the user is not stored permanently by the server.

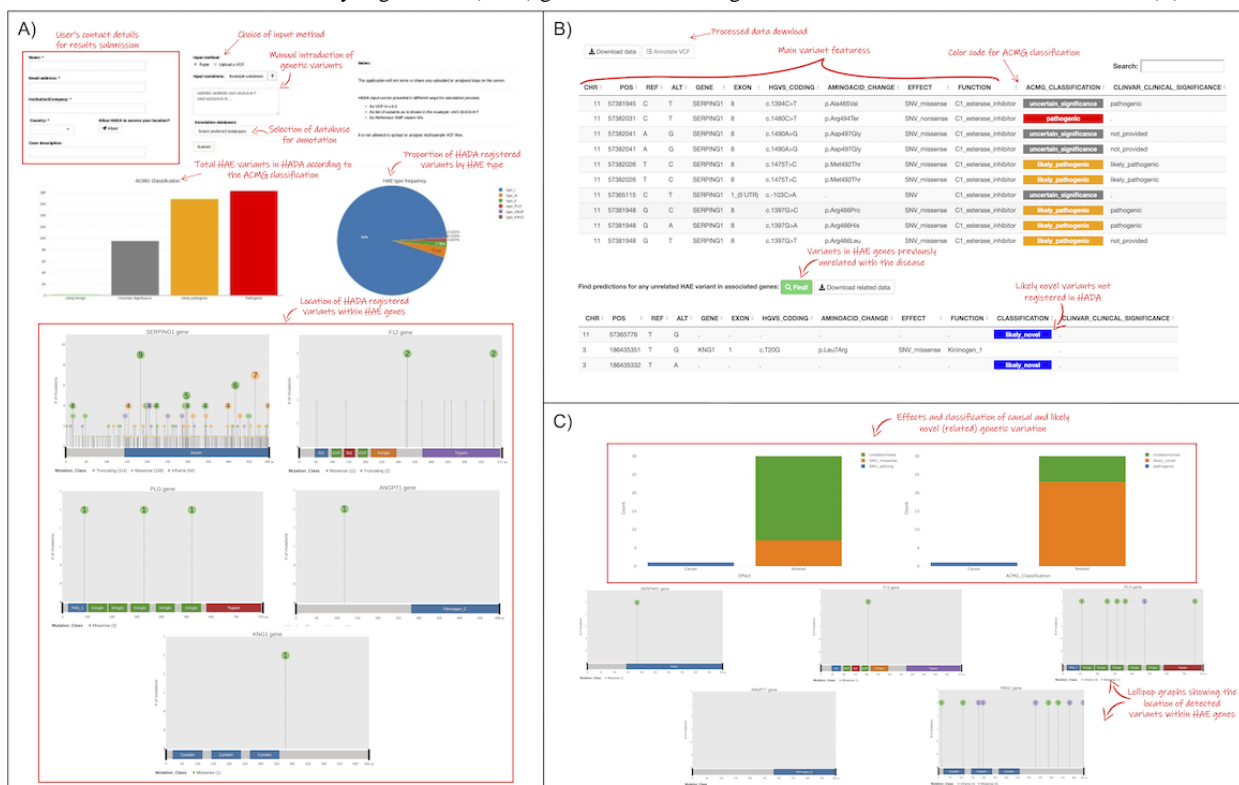
The data processing workflow is shown in Multimedia Appendix 1. Once the data are uploaded, the variants contained in the VCF

or that were provided by the user in chromosomal coordinates are matched against HADA. Automated detection and information extraction occur on the fly. Matched variants are subsequently shown in tabular format, accompanied by information previously selected by the user. HADA provides the ACMG classifications for each variant according to a color key to facilitate the identification of the HAE variants affecting function in the queries. HADA also suggests the subtype of HAE for which the variant affecting function is involved.

We anticipate that a user query might involve variants that have not yet been described among HAE genes (ie, novel HAE-related variants) or that have not been described in the scientific literature as an affecting functional variant to date. HADA integrates a genomic coordinates file that allows for the retrieval of predictions for any variant with any reference/alternative allele combination in any exonic region (plus a 50-bp flanking region on both ends of each locus) of the known HAE genes. Query results provided by HADA can be downloaded by the user in plain-text format in a comma-separated file, or as an annotated VCF in cases in which a VCF was uploaded by the user.

Finally, publication-ready plot generation and downloading are also incorporated in HADA. These functions plot graphs that summarize the information obtained from input data (ie, a VCF), as well as charts summarizing the proportions of the ACMG pathogenic classes predicted by VarSome, ClinVar, and InterVar separately, and the subtype of HAE that is associated with a variant affecting function that is identified in the query.

**Figure 4.** Selected snapshots of the HADA graphical interface with step-by-step instructions. Users can query variants of interest or alternatively upload an individual variant calling file (VCF) in the home page (A). The detected variants tab (B) shows the existence of variants affecting function in the query, as well as other variants in hereditary angioedema (HAE) genes. Plots are also generated and available for users to download (C).



### Validation of HADA in an HAE-Affected Family

To demonstrate the utility of HADA for the NGS-based identification of the variant affecting function involved in an HAE type II family, WES data (45 Mb) from the index patient, and the unaffected mother and sister were obtained. This detected a mean of 16,650 high-confidence variants per

individual. Sequencing metrics and variant calling results are shown in [Table 1](#). Resulting individual VCFs were processed with HADA, which reported a variant c.1396C>A (rs28940870) within *SERPING1* both in the index case and the sister that was heterozygous in both cases (compatible with an autosomal dominant inheritance pattern). The depth of coverage at this locus was 72X and 69X, respectively.

**Table 1.** Summary of the sequencing results in the validation study.

Feature	Index	Sister	Mother
Median insert size (bp <sup>a</sup> )	297	268	268
Total reads (millions)	130.4	100.5	148.7
Aligned reads (%)	99.5	100	100
Mean coverage (%)	49.3	44.3	54.2
Targeted fraction ≥ 30X (%)	66	55	76
Ti/Tv <sup>b</sup>	3.16	3.19	3.22
<b>Number of variants<sup>c</sup></b>	16,394	16,231	17,325
SNVs <sup>d</sup>	16,087	15,917	16,984
Indels <sup>e</sup>	307	314	341

<sup>a</sup>bp: base pairs.

<sup>b</sup>Ti/Tv: ratio of transitions to transversions.

<sup>c</sup>High confidence variants (FILTER=PASS; total depth≥20; genotype quality≥100; and mapping quality ≥ 50).

<sup>d</sup>SNVs: single nucleotide variants.

<sup>e</sup>Indels: insertion/deletion.

Most pathogenic prediction scores provided by HADA supported the variant as deleterious/damaging, and ACMG class predictors classified it either as a pathogenic/likely pathogenic variant or VUS ([Table 2](#)). This variant predicts an amino acid change (p.Arg466Ser) affecting the catalytic center of the protein, which has been previously described in independent HAE type II cases and was shown to be responsible for reduced C1-INH activity [34-36]. These observations are fully compatible with the

clinical and biochemical findings supporting a diagnosis of HAE type II in the index case. Surprisingly, the sister was asymptomatic and had never experienced angioedema attacks, which further exemplifies the incomplete penetrance of HAE or that disease onset has not yet occurred. In any case, the provided genetic information could be used to anticipate potential clinical symptoms.

**Table 2.** American College of Medical Genetics and Genomics (ACMG) class predictions and pathogenic scores provided by HADA for the c.1396C>A variant found in a hereditary angioedema-affected family.

Predictor	Estimation
<b>ACMG class prediction</b>	
VarSome	VUS <sup>a</sup>
ClinVar	Pathogenic
InterVar	Likely pathogenic
<b>Pathogenic predictors</b>	
SIFT <sup>b</sup>	Damaging
Polyphen2 <sup>c</sup>	Benign
MutationTaster	Disease-causing
CADD <sup>d</sup>	31
DANN <sup>e</sup> score	0.997
DANN rankscore	0.798
LRT <sup>f</sup>	Deleterious
MetaSVM	Damaging

<sup>a</sup>VUS: variant of uncertain significance.<sup>b</sup>SIFT: Sorting Intolerant From Tolerant.<sup>c</sup>PolyPhen2: Polymorphism Phenotyping v2.<sup>d</sup>CADD: Combined Annotation Dependent Depletion.<sup>e</sup>DANN: Deleterious Annotation of genetic variants using Neural Networks.<sup>f</sup>LRT: likelihood ratio test.<sup>g</sup>MetaSVM: meta-analytic support vector machine.

## Discussion

### Principal Findings

Here, we present HADA, a web-based analysis tool to facilitate the identification of the genetic variants causing HAE. With HADA, we aimed to provide a user-friendly tool to assist in the diagnosis of HAE that is adapted to NGS technologies and to the evolving knowledge of the causes of HAE. HADA has potential to reduce the time to interpret the detected variants in the NGS era by aggregating data from multiple sources, a process that commonly takes several hours to complete if it is handled manually [37,38]. At the moment, HADA integrates information from 450 SNVs and indels from 5 genes (*SERPING1*, *F12*, *PLG*, *ANGPT1*, and *KNG1*) that we classified as variants likely affecting function. While curating this information, we found that variant descriptions in HAE cases did not follow a standard [4], and many of the studies did not clearly declare the causality of the reported variants [35,39]. This situation helps to explain why up to 21.3% (n=96) of the simple bona fide variants affecting function continue to be reported as VUS in the best case. Finally, we demonstrated the utility of HADA for explaining HAE type II in a family that was assessed by WES in three family members, and offered conclusive information about the existence of the previously described variant c.1396C>A (p.Arg466Ser) in *SERPING1*.

As is the case for many other rare diseases, genetic testing in HAE has become an important step to reduce the diagnostic

odyssey [40], increase the diagnostic yield, and tailor treatments [4,5]. The widespread adoption of NGS technology in clinical settings has led to the emergence of a wide variety of bioinformatics tools to assist and accelerate the detection and interpretation of associated genetic variants and their impact on disease risks. In recent years, an array of variant prioritizers and interpreters have been developed to obtain optimal rankings for the variants causing the disease. However, NGS-based solutions have not been adopted in the HAE community until recently [29-31]. Part of the explanation may reside in the fact that only two causal genes were known until 2018, and therefore most clinical diagnoses could only be made based on clinical symptoms and the biochemical measurements of C4 and C1-INH levels in plasma, along with C1-INH activity [5]. Importantly, public resource updates to assist in variant interpretation lag behind the pace of genetic discoveries. This is the case of Simply-ClinVar [41] or ClinVar, which are still outdated with respect to the genetic causes of HAE, reporting information for only two of the associated genes (*SERPING1* and *F12*). Simply-ClinVar also suggests *SLC34A1* as another HAE gene, although this evidence is unsupported by the current literature. This situation is changing currently, and the benefits of using NGS technologies to assess multiple genes simultaneously for HAE diagnosis are now clearer [5]. In fact, in patients with normal levels and activity of C1-INH, genetic testing is recommended for the routine diagnosis of many HAE subtypes. Furthermore, according to Germenis et al [4] and the international consensus on genetic aspects of HAE, it is widely



recognized that C4 and C1-INH plasma measurements can generate inconclusive results even in patients with genetic defects in C1-INH. Based on this, HADA will surely help practitioners to adapt to the NGS-based genetic screens of HAE cases.

HADA has some strengths and limitations for the interpretation of genetic variants involved in the causes of HAE. Among the limitations, the tool does not currently assess VCFs from families or trios and does not allow inferring whether a variant affecting function has a de novo origin from the patient's data. Similarly, HADA uses GRCh37/hg19 coordinates, which is still considered to be the gold standard in clinical settings [42]. Despite the fact that HADA can report the existence of variants residing in introns near the key elements for splicing, as has been recently found for a type I HAE case [43,44], novel variants affecting function residing deep within intron positions will remain undetected. The main reason is the limited capacity of current algorithms to predict the pathogenic potential of deep intronic variants [45]. Similarly, a variable proportion of information remains unavailable for many of the variants affecting function collected in the underlying database. For example, allelic frequencies are currently available from gnomAD and ExAC for as few as 11 to 20 (2%-4%) of the variants affecting function in HAE. As an example, the variant detected in the family analyzed in this study is a bona-fide causal variant of HAE. However, there are no records of this variant in the gnomAD or ExAC databases. Similarly, not all variants affecting function in HAE are described in current ClinVar or InterVar versions. Collectively, this situation adds to the challenge of better predicting disease effects of many of the variants that are declared as VUS.

Structural variants have an important role in HAE due to the presence of Alu element-related reorganizations affecting function in *SERPING1*, potentially causing HAE type I [46,47]. However, structural variants will not be interpreted by HADA at the moment as they have not been collected in the database. One of the main reasons for this is that, unlike the case for SNVs and small indels, current locus-oriented databases such as HAEdb do not provide the chromosomal coordinates of recorded structural variants from HAE families, making it difficult to integrate in HADA.

Among the strengths of HADA, we highlight the versatility of the tool to allow performing simple queries or to directly upload

VCFs. This offers the possibility of analyzing a variant detected at any throughput scale, using either NGS technologies or Sanger sequencing. This feature is a great advantage compared to HAEdb or LOVD, which are not adapted for the automated analysis of NGS data (eg, from a VCF), serving only for single-variant queries, and neither of these previous tools has been subjected to a curation process or the inclusion of novel HAE genes. In addition to the imposed filters to construct the curated database of HADA, allele frequencies are also provided for all populations that are currently collected in the gnomAD and ExAC population databases. To cover the possibility of new discoveries, HADA also includes information from all other described variants in the coding regions, the 5' and 3' untranslated regions, and the 50-bp flanking regions (where most splicing variants affecting function are located) to easily flag novel variants in the uploaded data. Finally, despite the fact that its use does not require sensitive information for the analysis, HADA automatically conducts a password-secured encryption of the data and does not permanently store any data on the server.

We anticipate that HADA will be constantly updated to register new variants affecting function and associated information, changes in the pathogenic classification, and the possibility to assess structural variants, and GRCh38/hg38 coordinate and reference sequence number conversions. This will be based on third-party requests and internal updates according to novel discoveries and standards.

## Conclusions

To adapt the genetic diagnosis of HAE to the era of NGS-based genomic medicine, we have developed HADA as a free and publicly available tool for simplifying the identification of simple variants affecting function in HAE. HADA will allow users to focus on biologically relevant questions instead of having to learn to install software dependencies, variant annotation tools, and become familiar with the UNIX command line. The main advantages of HADA are that it is focused on a disease, its ease of use, the ability to display specific and curated information of HAE either from individual or VCF queries, and that it is freely available. By combining these features into a single graphical and interactive tool, we expect that variant prioritization in HAE will become easier, faster, and standardized.

## Acknowledgments

We would like to thank the patients and nurses involved in the validation study. This work was supported by the Ministerio de Ciencia e Innovación (RTC-2017-6471-1; AEI/FEDER, UE) and the Instituto de Salud Carlos III (CD19/00231), which were cofinanced by the European Regional Development Funds "A way of making Europe" from the European Union; Cabildo Insular de Tenerife (CGIEU0000219140); and by agreement OA17/008 with the Instituto Tecnológico y de Energías Renovables (ITER) to strengthen scientific and technological education, training, research, development, and innovation in genomics, personalized medicine, and biotechnology. AA was supported by a fellowship from ULL-Cajasierte. The content of this publication is solely the responsibility of the authors and does not necessarily reflect the views or policies of the funding agencies.



## Authors' Contributions

AA, AB, and CF wrote the manuscript and designed the figures. AA, LR, IR, and CF built and curated the database. ArC and JR contributed information for HAE clinical aspects and sample collection of the tested family. AIC, RM, AC, JS, and AA managed and sequenced the DNA samples of the family tested and performed the variant calling. AB, JS, and AA developed the front end. CF designed the project and obtained funding. All the authors revised and approved the final version of the manuscript.

## Conflicts of Interest

Takeda Pharmaceutical Company funded a travel grant to AA. The rest of the authors declare no conflicts of interest.

## Multimedia Appendix 1

Schematic representation of the variant calling file (VCF) variant matching against the curated database and data downloading in HADA.

[PNG File, 1417 KB - [jmir\\_v22i10e19040\\_app1.png](#)]

## References

1. Nzeako UC, Frigas E, Tremaine WJ. Hereditary angioedema: a broad review for clinicians. *Arch Intern Med* 2001 Nov 12;161(20):2417-2429. [doi: [10.1001/archinte.161.20.2417](#)] [Medline: [11700154](#)]
2. Agostoni A, Aygören-Pürsün E, Binkley KE, Blanch A, Bork K, Bouillet L, et al. Hereditary and acquired angioedema: problems and progress: proceedings of the third C1 esterase inhibitor deficiency workshop and beyond. *J Allergy Clin Immunol* 2004 Sep;114(3 Suppl):S51-131 [FREE Full text] [doi: [10.1016/j.jaci.2004.06.047](#)] [Medline: [15356535](#)]
3. Caballero T, Baeza ML, Cabañas R, Campos A, Cimbollek S, Gómez-Traseira C, Spanish Study Group on Bradykinin-Induced Angioedema (SGBA). Consensus statement on the diagnosis, management, and treatment of angioedema mediated by bradykinin. Part II. Treatment, follow-up, and special situations. *J Investig Allergol Clin Immunol* 2011;21(6):422-441; quiz 442 [FREE Full text] [Medline: [21995176](#)]
4. Germentis AE, Margaglione M, Pesquero JB, Farkas H, Cichon S, Csuka D, Hereditary Angioedema International Working Group. International Consensus on the Use of Genetics in the Management of Hereditary Angioedema. *J Allergy Clin Immunol Pract* 2020 Mar;8(3):901-911. [doi: [10.1016/j.jaip.2019.10.004](#)] [Medline: [31669336](#)]
5. Marcelino-Rodríguez I, Callero A, Mendoza-Alvarez A, Perez-Rodríguez E, Barrios-Recio J, García-Robaina JC, et al. Bradykinin-Mediated Angioedema: An Update of the Genetic Causes and the Impact of Genomics. *Front Genet* 2019;10:900. [doi: [10.3389/fgene.2019.00900](#)] [Medline: [31611908](#)]
6. Loules G, Zamanakou M, Parsopoulou F, Vatsiou S, Psarros F, Csuka D, et al. Targeted next-generation sequencing for the molecular diagnosis of hereditary angioedema due to C1-inhibitor deficiency. *Gene* 2018 Aug 15;667:76-82. [doi: [10.1016/j.gene.2018.05.029](#)] [Medline: [29753808](#)]
7. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet* 2014 Sep;30(9):418-426. [doi: [10.1016/j.tig.2014.07.001](#)] [Medline: [25108476](#)]
8. Kalmár L, Hegedüs T, Farkas H, Nagy M, Tordai A. HAEdb: a novel interactive, locus-specific mutation database for the C1 inhibitor gene. *Hum Mutat* 2005 Jan;25(1):1-5. [doi: [10.1002/humu.20112](#)] [Medline: [15580551](#)]
9. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015 May;17(5):405-424. [doi: [10.1038/gim.2015.30](#)] [Medline: [25741868](#)]
10. Ponard D, Gaboriaud C, Charignon D, Ghannam A, Wagenaar-Bos IGA, Roem D, et al. SERPING1 mutation update: Mutation spectrum and C1 Inhibitor phenotypes. *Hum Mutat* 2020 Jan;41(1):38-57. [doi: [10.1002/humu.23917](#)] [Medline: [31517426](#)]
11. Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* 2011 May;32(5):557-563. [doi: [10.1002/humu.21438](#)] [Medline: [21520333](#)]
12. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. *Bioinformatics* 2019 Jun 01;35(11):1978-1980 [FREE Full text] [doi: [10.1093/bioinformatics/bty897](#)] [Medline: [30376034](#)]
13. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med* 2020 Feb;22(2):245-257. [doi: [10.1038/s41436-019-0686-8](#)] [Medline: [31690835](#)]
14. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010 Sep;38(16):e164 [FREE Full text] [doi: [10.1093/nar/gkq603](#)] [Medline: [20601685](#)]
15. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Genome Aggregation Database Consortium, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020 May;581(7809):434-443 [FREE Full text] [doi: [10.1038/s41586-020-2308-7](#)] [Medline: [32461654](#)]

16. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016 Aug 18;536(7616):285-291 [[FREE Full text](#)] [doi: [10.1038/nature19057](https://doi.org/10.1038/nature19057)] [Medline: [27535533](#)]
17. Alirezaie N, Kernohan KD, Hartley T, Majewski J, Hocking TD. ClinPred: Prediction Tool to Identify Disease-Relevant Nonsynonymous Single-Nucleotide Variants. *Am J Hum Genet* 2018 Oct 04;103(4):474-483 [[FREE Full text](#)] [doi: [10.1016/j.ajhg.2018.08.005](https://doi.org/10.1016/j.ajhg.2018.08.005)] [Medline: [30220433](#)]
18. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014 Jan;42(Database issue):D980-D985 [[FREE Full text](#)] [doi: [10.1093/nar/gkt1113](https://doi.org/10.1093/nar/gkt1113)] [Medline: [24234437](#)]
19. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet* 2017 Feb 02;100(2):267-280 [[FREE Full text](#)] [doi: [10.1016/j.ajhg.2017.01.004](https://doi.org/10.1016/j.ajhg.2017.01.004)] [Medline: [28132688](#)]
20. Hosseini M, Pratas D, Pinho AJ. Cryfa: a secure encryption tool for genomic data. *Bioinformatics* 2019 Jan 01;35(1):146-148 [[FREE Full text](#)] [doi: [10.1093/bioinformatics/bty645](https://doi.org/10.1093/bioinformatics/bty645)] [Medline: [30020420](#)]
21. HADA download. github. 2020. URL: <https://github.com/genomicsITER/HADA> [accessed 2020-09-22]
22. HADA interface. URL: <http://hada.hpc.iter.es/> [accessed 2020-09-22]
23. Teide HPC. URL: <https://teidehpc.iter.es/> [accessed 2020-06-15]
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009 Jul 15;25(14):1754-1760 [[FREE Full text](#)] [doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324)] [Medline: [19451168](#)]
25. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 2012 Oct 15;28(20):2678-2679. [doi: [10.1093/bioinformatics/bts503](https://doi.org/10.1093/bioinformatics/bts503)] [Medline: [22914218](#)]
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009 Aug 15;25(16):2078-2079 [[FREE Full text](#)] [doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352)] [Medline: [19505943](#)]
27. Picard. URL: <http://broadinstitute.github.io/picard> [accessed 2020-09-22]
28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010 Sep;20(9):1297-1303. [doi: [10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110)] [Medline: [20644199](#)]
29. Bafunno V, Firinu D, D'Apolito M, Cordisco G, Loffredo S, Leccese A, et al. Mutation of the angiotensinogen gene (ANGPT1) associates with a new type of hereditary angioedema. *J Allergy Clin Immunol* 2018 Mar;141(3):1009-1017. [doi: [10.1016/j.jaci.2017.05.020](https://doi.org/10.1016/j.jaci.2017.05.020)] [Medline: [28601681](#)]
30. Bork K, Wulff K, Steinmüller-Magin L, Braenne I, Staubach-Renz P, Witzke G, et al. Hereditary angioedema with a mutation in the plasminogen gene. *Allergy* 2018 Feb;73(2):442-450. [doi: [10.1111/all.13270](https://doi.org/10.1111/all.13270)] [Medline: [28795768](#)]
31. Bork K, Wulff K, Rossmann H, Steinmüller-Magin L, Braenne I, Witzke G, et al. Hereditary angioedema cosegregating with a novel kininogen 1 gene mutation changing the N-terminal cleavage site of bradykinin. *Allergy* 2019 Dec;74(12):2479-2481. [doi: [10.1111/all.13869](https://doi.org/10.1111/all.13869)] [Medline: [31087670](#)]
32. Bean LJH, Hegde MR. Clinical implications and considerations for evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Med* 2017 Dec 18;9(1):111 [[FREE Full text](#)] [doi: [10.1186/s13073-017-0508-z](https://doi.org/10.1186/s13073-017-0508-z)] [Medline: [29254502](#)]
33. Niroula A, Vihinen M. How good are pathogenicity predictors in detecting benign variants? *PLoS Comput Biol* 2019 Feb;15(2):e1006481 [[FREE Full text](#)] [doi: [10.1371/journal.pcbi.1006481](https://doi.org/10.1371/journal.pcbi.1006481)] [Medline: [30742610](#)]
34. Aulak KS, Cicardi M, Harrison RA. Identification of a new P1 residue mutation (444Arg→Ser) in a dysfunctional C1 inhibitor protein contained in a type II hereditary angioedema plasma. *FEBS Lett* 1990 Jun 18;266(1-2):13-16. [doi: [10.1016/0014-5793\(90\)81494-9](https://doi.org/10.1016/0014-5793(90)81494-9)] [Medline: [2365061](#)]
35. Xu Y, Zhi Y, Yin J, Wang L, Wen L, Gu J, et al. Mutational spectrum and geno-phenotype correlation in Chinese families with hereditary angioedema. *Allergy* 2012 Nov;67(11):1430-1436. [doi: [10.1111/all.12024](https://doi.org/10.1111/all.12024)] [Medline: [22994404](#)]
36. Gösswein T, Kocot A, Emmert G, Kreuz W, Martinez-Saguer I, Aygören-Pürsün E, et al. Mutational spectrum of the C1INH (SERPING1) gene in patients with hereditary angioedema. *Cytogenet Genome Res* 2008;121(3-4):181-188. [doi: [10.1159/000138883](https://doi.org/10.1159/000138883)] [Medline: [18758157](#)]
37. Harrison SM, Dolinsky JS, Knight Johnson AE, Pesaran T, Azzariti DR, Bale S, et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genet Med* 2017 Oct;19(10):1096-1104. [doi: [10.1038/gim.2017.14](https://doi.org/10.1038/gim.2017.14)] [Medline: [28301460](#)]
38. Christensen KD, Vassy JL, Phillips KA, Blout CL, Azzariti DR, Lu CY, MedSeq Project. Short-term costs of integrating whole-genome sequencing into primary care and cardiology settings: a pilot randomized trial. *Genet Med* 2018 Dec;20(12):1544-1553. [doi: [10.1038/gim.2018.35](https://doi.org/10.1038/gim.2018.35)] [Medline: [29565423](#)]
39. Martinho A, Mendes J, Simões O, Nunes R, Gomes J, Dias Castro E, et al. Mutations analysis of C1 inhibitor coding sequence gene among Portuguese patients with hereditary angioedema. *Mol Immunol* 2013 Apr;53(4):431-434. [doi: [10.1016/j.molimm.2012.09.003](https://doi.org/10.1016/j.molimm.2012.09.003)] [Medline: [23123409](#)]

40. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet* 2017 May 04;100(5):695-705 [FREE Full text] [doi: [10.1016/j.ajhg.2017.04.003](https://doi.org/10.1016/j.ajhg.2017.04.003)] [Medline: [28475856](https://pubmed.ncbi.nlm.nih.gov/28475856/)]
41. Pérez-Palma E, Gramm M, Nürnberg P, May P, Lal D. Simple ClinVar: an interactive web server to explore and retrieve gene and disease variants aggregated in ClinVar database. *Nucleic Acids Res* 2019 Jul 02;47(W1):W99-W105 [FREE Full text] [doi: [10.1093/nar/gkz411](https://doi.org/10.1093/nar/gkz411)] [Medline: [31114901](https://pubmed.ncbi.nlm.nih.gov/31114901/)]
42. Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, Shyr Y. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics* 2017 Mar;109(2):83-90 [FREE Full text] [doi: [10.1016/j.ygeno.2017.01.005](https://doi.org/10.1016/j.ygeno.2017.01.005)] [Medline: [28131802](https://pubmed.ncbi.nlm.nih.gov/28131802/)]
43. Vatsiou S, Zamanakou M, Loules G, Psarros F, Parsopoulou F, Csuka D, et al. A novel deep intronic SERPING1 variant as a cause of hereditary angioedema due to C1-inhibitor deficiency. *Allergol Int* 2020 Jan 17:online ahead of print [FREE Full text] [doi: [10.1016/j.alit.2019.12.009](https://doi.org/10.1016/j.alit.2019.12.009)] [Medline: [31959500](https://pubmed.ncbi.nlm.nih.gov/31959500/)]
44. Hujová P, Souček P, Grodecká L, Grombířková H, Ravčuková B, Kuklínek P, et al. Deep Intronic Mutation in SERPING1 Caused Hereditary Angioedema Through Pseudoexon Activation. *J Clin Immunol* 2020 Apr;40(3):435-446. [doi: [10.1007/s10875-020-00753-2](https://doi.org/10.1007/s10875-020-00753-2)] [Medline: [31982983](https://pubmed.ncbi.nlm.nih.gov/31982983/)]
45. Lord J, Gallone G, Short PJ, McRae JF, Ironfield H, Wynn EH, Deciphering Developmental Disorders study. Pathogenicity and selective constraint on variation near splice sites. *Genome Res* 2019 Feb;29(2):159-170. [doi: [10.1101/gr.238444.118](https://doi.org/10.1101/gr.238444.118)] [Medline: [30587507](https://pubmed.ncbi.nlm.nih.gov/30587507/)]
46. López-Lera A, Garrido S, Roche O, López-Trascasa M. SERPING1 mutations in 59 families with hereditary angioedema. *Mol Immunol* 2011 Oct;49(1-2):18-27. [doi: [10.1016/j.molimm.2011.07.010](https://doi.org/10.1016/j.molimm.2011.07.010)] [Medline: [21864911](https://pubmed.ncbi.nlm.nih.gov/21864911/)]
47. Germenis AE, Speletas M. Genetics of Hereditary Angioedema Revisited. *Clin Rev Allergy Immunol* 2016 Oct;51(2):170-182. [doi: [10.1007/s12016-016-8543-x](https://doi.org/10.1007/s12016-016-8543-x)] [Medline: [27116602](https://pubmed.ncbi.nlm.nih.gov/27116602/)]

## Abbreviations

**ACMG:** American College of Medical Genetics and Genomics  
**C1-INH:** C1 esterase inhibitor  
**HADA:** Hereditary Angioedema Database Annotation  
**HAE:** hereditary angioedema  
**HGVS:** Human Genome Variation Society  
**HUNSC:** Hospital Universitario Nuestra Señora de Candelaria  
**indel:** insertion/deletion  
**LOVD:** Leiden Open Variant Database  
**NGS:** next-generation sequencing  
**SNV:** single nucleotide variant  
**VCF:** variant calling file  
**VUS:** variant of uncertain significance  
**WES:** whole-exome sequencing

*Edited by G Eysenbach; submitted 01.04.20; peer-reviewed by A Germenis, T González-Quevedo Tejerina; comments to author 13.06.20; revised version received 22.06.20; accepted 25.06.20; published 09.10.20.*

### *Please cite as:*

Mendoza-Alvarez A, Muñoz-Barrera A, Rubio-Rodríguez LA, Marcelino-Rodríguez I, Corrales A, Iñigo-Campos A, Callero A, Perez-Rodríguez E, Garcia-Robaina JC, González-Montelongo R, Lorenzo-Salazar JM, Flores C  
 Interactive Web-Based Resource for Annotation of Genetic Variants Causing Hereditary Angioedema (HADA): Database Development, Implementation, and Validation  
*J Med Internet Res* 2020;22(10):e19040  
 URL: <http://www.jmir.org/2020/10/e19040/>  
 doi:[10.2196/19040](https://doi.org/10.2196/19040)  
 PMID:[33034563](https://pubmed.ncbi.nlm.nih.gov/33034563/)

©Alejandro Mendoza-Alvarez, Adrián Muñoz-Barrera, Luis Alberto Rubio-Rodríguez, Itahisa Marcelino-Rodríguez, Almudena Corrales, Antonio Iñigo-Campos, Ariel Callero, Eva Perez-Rodríguez, Jose Carlos Garcia-Robaina, Rafaela González-Montelongo, Jose Miguel Lorenzo-Salazar, Carlos Flores. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 09.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium,

provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Letter to the Editor

# Is a Ratio Scale Assumption for Physician Ratings Justified? Comment on “What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data”

Uwe Konerding<sup>1,2</sup>, DPhil

<sup>1</sup>Trimberg Research Academy, University of Bamberg, Bamberg, Germany

<sup>2</sup>Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany

**Corresponding Author:**

Uwe Konerding, DPhil

Trimberg Research Academy

University of Bamberg

An der Weberei 5

Bamberg, D-96045

Germany

Phone: 49 951 863 3098

Email: [uwe.konerding@uni-bamberg.de](mailto:uwe.konerding@uni-bamberg.de)

**Related Articles:**

Comment on: <https://www.jmir.org/2020/2/e13830/>

Comment in: <https://www.jmir.org/2020/10/e21057/>

(*J Med Internet Res* 2020;22(10):e18289) doi:[10.2196/18289](https://doi.org/10.2196/18289)

**KEYWORDS**

patient satisfaction; modeling; method; scale level; measurement theory

In a paper published in the *Journal of Medical Internet Research*, Bidmon et al [1] investigated the relationships between 24 attributes of the service provided by physicians and 4 features of patient satisfaction with these physicians. The service attributes were assessed via 4-category rating scales coded from 1 for “strongly disagree” to 4 for “strongly agree” and the satisfaction features via 5-category rating scales from 1 for “bad” to 5 for “excellent.” The latter measures were rescaled to range from 1 to 4.

The authors stated that their measures “have to be assumed at ratio scale level.” Ratio scale level is a concept defined in the representational theory of measurement [2]. According to this theory, measurement means assigning numbers to objects so that the numerical relations reflect the empirical relations between objects. Examples of empirical relations are the relations that emerge when a double pan balance is used. When two objects *A* and *B* are compared, the relations “*A* is heavier than *B*,” “*B* is heavier than *A*,” or “*A* is as heavy as *B*” might emerge. When, for all objects, “*A* is heavier than *B*” and “*B* is heavier than *C*” imply that “*A* is heavier than *C*,” then these relations can be represented numerically by always assigning

the higher number to the heavier object. This is the prototypical example of an ordinal scale. When two objects *A* and *B* are placed together on one pan of the scale and an individual object *C* on the other, empirical relations between pairs of objects and individual objects emerge that are analogous to those for comparing individual objects. When, for these relations, the same regularities hold as for the addition of numbers, these relations can be represented by assigning numbers so that  $\text{number}(A) + \text{number}(B) = \text{number}(C)$ , if and only if *A* and *B* together are as heavy as *C*. Such an assignment is empirically determined except for the arbitrary choice of the measurement unit. The position of the zero-point is empirically determined. This is, by definition, a ratio scale.

In the measurements presented by Bidmon et al [1], the basic objects are statements regarding attributes or, respectively, physician practices. The relevant empirical relations are results of comparative judgments regarding these objects. Presently, there is no evidence that, for such judgments, empirical structures exist that permit the position of the zero-point to be determined. Hence, the zero-points for the rating scales analyzed by Bidmon et al [1] cannot be determined empirically. In fact,



these zero-points are determined by arbitrary settings established by convention. Therefore, these scales are definitely not ratio scales.

In the absence of empirically determined zero-points, the analyses performed by Bidmon et al [1] are questionable. The authors estimated parameters for the model

$$\ln(Y) = b_0 + b_1 \ln(X) \quad (1)$$

where  $Y$  represents satisfaction with a specific feature and  $X$  is an agreement with a statement regarding a service attribute. The estimations for these parameters change when the locations of the zero-points are changed. Consequently, what these parameters tell us about the actual relationships between satisfaction and service attributes is unclear.

## Conflicts of Interest

None declared.

## References

1. Bidmon S, Elshiewy O, Terlutter R, Boztug Y. What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data. *J Med Internet Res* 2020 Feb 3;22(2):e13830. [doi: [10.2196/13830](https://doi.org/10.2196/13830)] [Medline: [32012063](https://pubmed.ncbi.nlm.nih.gov/32012063/)]
2. Krantz DH, Luce RD, Suppes P, Tversky A. Foundations of Measurement: Additive and Polynomial Representations, Vol 1. New York: Academic Press; 1971.

*Edited by T Derrick, G Eysenbach; submitted 17.02.20; peer-reviewed by S Bidmon, J Doll, A Samaranayaka, B Breyer; comments to author 08.04.20; revised version received 14.05.20; accepted 10.08.20; published 26.10.20.*

*Please cite as:*

Konerding U

*Is a Ratio Scale Assumption for Physician Ratings Justified? Comment on "What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data"*

*J Med Internet Res* 2020;22(10):e18289

URL: <http://www.jmir.org/2020/10/e18289/>

doi: [10.2196/18289](https://doi.org/10.2196/18289)

PMID: [33104009](https://pubmed.ncbi.nlm.nih.gov/33104009/)

©Uwe Konerding. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Letter to the Editor

# Authors' Reply to: Is a Ratio Scale Assumption for Physician Ratings Justified? Comment on "What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data"

Sonja Bidmon<sup>1</sup>, PhD; Ossama Elshiewy<sup>2</sup>, PhD; Ralf Terlutter<sup>1</sup>, PhD; Yasemin Boztug<sup>2</sup>, PhD

<sup>1</sup>Department of Marketing and International Management, Alpen-Adria-Universitaet Klagenfurt, Klagenfurt am Woerthersee, Austria

<sup>2</sup>Department of Business Administration, Marketing and Consumer Behavior, University of Goettingen, Goettingen, Germany

**Corresponding Author:**

Sonja Bidmon, PhD

Department of Marketing and International Management

Alpen-Adria-Universitaet Klagenfurt

Universitaetsstrasse 65-67

Klagenfurt am Woerthersee, 9020

Austria

Phone: 43 463 2700 4048

Email: [sonja.bidmon@aau.at](mailto:sonja.bidmon@aau.at)

**Related Articles:**

Comment on: <https://www.jmir.org/2020/10/e18289/>

Comment on: <https://www.jmir.org/2020/2/e13830/>

(*J Med Internet Res* 2020;22(10):e21057) doi:[10.2196/21057](https://doi.org/10.2196/21057)

**KEYWORDS**

online physician ratings; patient satisfaction; multiattribute models; health care management

We appreciate the comments made by Konerding [1] and are thankful for the opportunity to take part in this research dialogue. Konerding [1] states that "the zero-points for the rating scales analyzed by Bidmon et al [2] cannot be determined empirically" and "what these parameters tell us about the actual relationships between satisfaction and service attributes is unclear." It is obvious that the arguments of Konerding [1] rely solely on the paradigm of representational theory of measurement. However, the academic literature acknowledges three theories in this domain: the representational, the operational, and the classical (we refer to Michell [3] for an excellent discussion on this topic, which has already been cited in our paper [2]). Consequently, we understand the point of view conveyed by Konerding [1], but perceive it as too narrow in the spirit of empirical research. We are pleased to clarify as follows.

First, our assumption for the scale level of the satisfaction ratings draws from a well-known conceptual framework: the 3-factor model of customer satisfaction [4-6]. This model includes the concept of linear and nonlinear relationships between satisfaction ratings of a service attribute and the overall

satisfaction rating with the service (describing diminishing, constant, or increasing returns).

Second, the methodological assumptions to empirically identify the 3-factor model of customer satisfaction are clearly spelled out in our methods section (see our paper [2], subsection *Statistical Analysis*), and the approach using the log-log regression model to estimate elasticities is well established in the literature of econometric models of demand [7]. Elasticities allow the interpretation of slope coefficients in terms of diminishing, constant, and increasing returns.

Third, in our work [2], we emphasize the range and conditions where our results can be interpreted: the empirical meaning of our nonlinear slope coefficients and the prediction of changes in patient satisfaction apply to the range of publicly available ratings on the physician rating website. Explicitly assuming ratio-scale level within our specific range of satisfaction ratings is a helpful and valid assumption to empirically identify the 3-factor model of customer satisfaction (which has only been implicitly assumed in previous research). Our choice of numerical coding for the ratings (which follows common

practice in empirical research) ensures that the diminishing, constant, and increasing returns are correctly identified within the relevant range of our satisfaction ratings.

Fourth, we provide a robustness check in our paper [2] to show that our empirical findings do not rely on the log-log regression model or the elasticities (this robustness check was explicitly suggested by the author of the comments [1] during the review process of our paper [2]). The most important finding from applying the alternative approach is that it leads to results that are identical to those from our main approach for the classification of the service attributes to the 3-factor model of customer satisfaction. We would like to take the opportunity to

emphasize that this alternative approach (leading to the same results) should not be preferred to our main approach because it violates the principle of sparse parametrization and lacks straightforward hypothesis testing.

Taken together, we are thankful for the fruitful discussion concerning our main approach, which has led us to clearly spell out the assumptions under which our results are valid, and to emphasize the robustness check that supports the validity of our main approach. We advise researchers to bear these assumptions in mind when interpreting our results and especially when adopting our main approach to empirically identify nonlinear slopes from rating data.

## Conflicts of Interest

None declared.

## References

1. Konerding U. Is a Ratio Scale Assumption for Physician Ratings Justified? Comment on "What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data". J Med Internet Res 2020;22(10):e18289. [doi: [10.2196/18289](https://doi.org/10.2196/18289)]
2. Bidmon S, Elshiewy O, Terlutter R, Boztug Y. What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data. J Med Internet Res 2020 Feb 03;22(2):e13830 [FREE Full text] [doi: [10.2196/13830](https://doi.org/10.2196/13830)] [Medline: [32012063](https://pubmed.ncbi.nlm.nih.gov/32012063/)]
3. Michell J. Measurement scales and statistics: A clash of paradigms. Psychological Bulletin 1986;100(3):398-407. [doi: [10.1037/0033-2909.100.3.398](https://doi.org/10.1037/0033-2909.100.3.398)]
4. Arbore A, Busacca B. Customer satisfaction and dissatisfaction in retail banking: Exploring the asymmetric impact of attribute performances. Journal of Retailing and Consumer Services 2009 Jul;16(4):271-280. [doi: [10.1016/j.jretconser.2009.02.002](https://doi.org/10.1016/j.jretconser.2009.02.002)]
5. Matzler K, Sauerwein E. The factor structure of customer satisfaction: An empirical test of the importance-grid and the penalty-reward-contrast analysis. Int J of Service Industry Mgmt 2002 Oct;13(4):314-332. [doi: [10.1108/09564230210445078](https://doi.org/10.1108/09564230210445078)]
6. Matzler K, Bailom F, Hinterhuber H, Renzl B, Pichler J. The asymmetric relationship between attribute-level performance and overall customer satisfaction: a reconsideration of the importance-performance analysis. Industrial Marketing Management 2004 May;33(4):271-277. [doi: [10.1016/S0019-8501\(03\)00055-5](https://doi.org/10.1016/S0019-8501(03)00055-5)]
7. Hanssens DM, Parsons LJ, Schultz RL. Market Response Models: Econometric and Time Series Analysis. Boston, MA: Kluwer Academic Publishers; 2001.

*Edited by T Derrick, G Eysenbach; submitted 04.06.20; peer-reviewed by U Konerding; accepted 01.10.20; published 26.10.20.*

*Please cite as:*

*Bidmon S, Elshiewy O, Terlutter R, Boztug Y*

*Authors' Reply to: Is a Ratio Scale Assumption for Physician Ratings Justified? Comment on "What Patients Value in Physicians: Analyzing Drivers of Patient Satisfaction Using Physician-Rating Website Data"*

*J Med Internet Res 2020;22(10):e21057*

*URL: <https://www.jmir.org/2020/10/e21057>*

*doi: [10.2196/21057](https://doi.org/10.2196/21057)*

*PMID: [33104006](https://pubmed.ncbi.nlm.nih.gov/33104006/)*

©Sonja Bidmon, Ossama Elshiewy, Ralf Terlutter, Yasemin Boztug. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Letter to the Editor

# Comment on “Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial”

Yusuke Saishoji<sup>1</sup>, MD; Akihiro Shiroshita<sup>2</sup>, MD; Yasushi Tsujimoto<sup>3,4</sup>, MD, MPH

<sup>1</sup>Department of General Internal Medicine, National Hospital Organization Nagasaki Medical Center, Omura, Japan

<sup>2</sup>Department of Pulmonology, Kameda Medical Center, Kamogawa, Japan

<sup>3</sup>Department of Nephrology and Dialysis, Kyoritsu Hospital, Kawanishi, Japan

<sup>4</sup>Healthcare Epidemiology, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan

**Corresponding Author:**

Yusuke Saishoji, MD

Department of General Internal Medicine

National Hospital Organization Nagasaki Medical Center

Kubara 1001-1

Omura, 856-8562

Japan

Phone: 81 957 52 3121

Email: [yusukesaisioji@gmail.com](mailto:yusukesaisioji@gmail.com)

**Related Articles:**

Comment on: <https://www.jmir.org/2020/6/e16772/>

Comment in: <https://www.jmir.org/2020/10/e24084>

(*J Med Internet Res* 2020;22(10):e21505) doi:[10.2196/21505](https://doi.org/10.2196/21505)

**KEYWORDS**

internet-based intervention; education; spin bias

We read the article “Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial” by Chan et al [1] with great interest. The idea that face to face education is difficult and education via the internet or social networking systems is necessary is intriguing; this is an important perspective in the midst of the COVID-19 pandemic. In the article, the editor wrote, “readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness”; hence, we would like to discuss some perspectives.

First, the primary outcome was measured using the differences in the scores of the pre- and postintervention knowledge assessments, which comprised the 25-item Dementia Knowledge Assessment Scale (DKAS) and a formative evaluation of 20 multiple-choice questions. However, the authors’ conclusion is

focused on the outcome of improving participants’ knowledge concerning a single subscale in DKAS. This interpretation might be a spin that could warp the interpretation of results and mislead readers [2].

Second, the article has issues with multiple testing. When tests are divided into subscales, some of them may have significant differences. To show a significant difference in the effect, corrections to the multiple tests are required [3].

Finally, this study is a pre and post study; therefore, a paired *t* test should be used instead of a two-sample *t* test. The two-sample *t* test estimates the treatment effect using only the responses at follow-up, and it does not use any information at baseline, which may be useful for increasing efficiency if the baseline and follow-up outcomes are correlated [4].

**Conflicts of Interest**

None declared.

## References

1. Chan WS, Leung AY. Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial. *J Med Internet Res* 2020 Jun 02;22(6):e16772 [FREE Full text] [doi: [10.2196/16772](https://doi.org/10.2196/16772)] [Medline: [32484441](https://pubmed.ncbi.nlm.nih.gov/32484441/)]
2. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010 May 26;303(20):2058-2064. [doi: [10.1001/jama.2010.651](https://doi.org/10.1001/jama.2010.651)] [Medline: [20501928](https://pubmed.ncbi.nlm.nih.gov/20501928/)]
3. Li G, Taljaard M, Van den Heuvel ER, Levine MA, Cook DJ, Wells GA, et al. An introduction to multiplicity issues in clinical trials: the what, why, when and how. *Int J Epidemiol* 2017 Apr 01;46(2):746-755. [doi: [10.1093/ije/dyw320](https://doi.org/10.1093/ije/dyw320)] [Medline: [28025257](https://pubmed.ncbi.nlm.nih.gov/28025257/)]
4. Yang L, Tsiatis AA. Efficiency Study of Estimators for a Treatment Effect in a Pretest–Posttest Trial. *The American Statistician* 2001 Nov;55(4):314-321. [doi: [10.1198/000313001753272466](https://doi.org/10.1198/000313001753272466)]

## Abbreviations

**DKAS:** Dementia Knowledge Assessment Scale

*Edited by T Derrick, G Eysenbach; submitted 30.06.20; peer-reviewed by WSY Chan; accepted 01.10.20; published 30.10.20.*

*Please cite as:*

Saishoji Y, Shiroshita A, Tsujimoto Y

Comment on “Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial”  
*J Med Internet Res* 2020;22(10):e21505

URL: <http://www.jmir.org/2020/10/e21505/>

doi: [10.2196/21505](https://doi.org/10.2196/21505)

PMID: [33124985](https://pubmed.ncbi.nlm.nih.gov/33124985/)

©Yusuke Saishoji, Akihiro Shiroshita, Yasushi Tsujimoto. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org>, as well as this copyright and license information must be included.



Letter to the Editor

# Authors' Reply to: Comment on "Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial"

Windy SY Chan<sup>1,2\*</sup>, BPharm, MClintPharm, DHSc; Angela YM Leung<sup>3\*</sup>, BN, MHA, PhD

<sup>1</sup>School of Health Sciences, Caritas Institute of Higher Education, New Territories, Hong Kong

<sup>2</sup>Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Kowloon, Hong Kong

<sup>3</sup>Centre for Gerontological Nursing, School of Nursing, The Hong Kong Polytechnic University, Kowloon, Hong Kong

\* all authors contributed equally

**Corresponding Author:**

Windy SY Chan, BPharm, MClintPharm, DHSc

School of Health Sciences

Caritas Institute of Higher Education

2 Chui Ling Lane, Tseung Kwan O

New Territories

Hong Kong

Phone: 852 36536664

Email: [sychan@cihe.edu.hk](mailto:sychan@cihe.edu.hk)

**Related Articles:**

Comment on: <http://www.jmir.org/2020/10/e21505/>

Comment on: <https://www.jmir.org/2020/6/e16772/>

(*J Med Internet Res* 2020;22(10):e24084) doi:[10.2196/24084](https://doi.org/10.2196/24084)

**KEYWORDS**

dementia care; social network sites; Facebook

We would like to thank Dr Saishoji and colleagues [1] for the opportunity to respond to the issues raised in their letter and to clarify aspects of our study design in relation to these concerns. We would also like to thank Dr Saishoji and colleagues for their interest in our paper [2] and for taking the time to express their concerns.

Our study, a pilot randomized controlled trial, adopted mixed-research methods (quantitative and qualitative) to evaluate the feasibility and acceptability of utilizing Facebook to deliver a continuous professional education (CPE) program to health care professionals. We believe that it is important to explore different effects of the intervention on participants' own provision of care. Therefore, we focused on measuring the knowledge relevant for the practical care of people with dementia rather than a high level of scientific knowledge related to dementia [3].

In addition to the primary outcome (knowledge about dementia, via the Dementia Knowledge Assessment Scale), we also evaluated participants' compliance, participants' engagement in the intervention, participants' satisfaction, and participants'

attitudes toward using Facebook for professional education from a quantitative perspective. Although significant differences between the intervention group and the control group were not observed in the primary outcomes, we found that the Facebook intervention did well in offering gains to participants' knowledge and enhanced their engagement and compliance. For that reason, we tried to elaborate on all the outcomes to bring readers a comprehensive picture of the actual application of Facebook for future CPE programs.

We agree that there may be an increased risk of type 1 error when applying multiple statistical tests, and in some cases, correction, such as the Bonferroni correction, can be used. However, we understand that adopting this kind of correction is not a must; it depends on the rationale of the study [4]. Our study was exploratory, involving a few posthoc comparisons, which were regarded as hypotheses for further investigation. Thus, we believe that no correction would be needed in our pilot study.

We appreciate the comment on how our statistical analysis was reported. Hence, we would like to take this opportunity to

provide additional details pertaining to our analysis. In measuring the primary outcomes, an independent sample *t* test was used to compare the changes in the mean knowledge gain scores between the two groups at the postintervention assessments.

We agree with Dr Saishoji and colleagues [1] that online education is intriguing and important during the current COVID-19 pandemic. We would be glad to keep communicating with investigators on research in this area.

## Conflicts of Interest

None declared.

## References

1. Saishoji Y, Shiroshita A, Tsujimoto Y. Comment on "Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial". J Med Internet Res 2020 Oct 30;22(10):e21505. [doi: [10.2196/21505](https://doi.org/10.2196/21505)]
2. Chan WS, Leung AY. Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial. J Med Internet Res 2020 Jun 02;22(6):e16772 [FREE Full text] [doi: [10.2196/16772](https://doi.org/10.2196/16772)] [Medline: [32484441](https://pubmed.ncbi.nlm.nih.gov/32484441/)]
3. Tierney L, Mason R, Doherty K, Winbolt M, Long M, Robinson A. Workshops on diagnosis and management of dementia for general practitioners: a pre-post intervention study of dementia knowledge. BMJ Open 2019 Apr 09;9(4):e027804. [doi: [10.1136/bmjopen-2018-027804](https://doi.org/10.1136/bmjopen-2018-027804)] [Medline: [30967411](https://pubmed.ncbi.nlm.nih.gov/30967411/)]
4. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014 Sep;34(5):502-508. [doi: [10.1111/opo.12131](https://doi.org/10.1111/opo.12131)] [Medline: [24697967](https://pubmed.ncbi.nlm.nih.gov/24697967/)]

## Abbreviations

**CPE:** continuous professional education

*Edited by T Derrick, G Eysenbach; submitted 03.09.20; peer-reviewed by Y Saishoji; accepted 01.10.20; published 30.10.20.*

*Please cite as:*

Chan WSY, Leung AYM

Authors' Reply to: Comment on "Facebook as a Novel Tool for Continuous Professional Education on Dementia: Pilot Randomized Controlled Trial"

J Med Internet Res 2020;22(10):e24084

URL: <https://www.jmir.org/2020/10/e24084>

doi: [10.2196/24084](https://doi.org/10.2196/24084)

PMID: [33124996](https://pubmed.ncbi.nlm.nih.gov/33124996/)

©Windy SY Chan, Angela YM Leung. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Corrigenda and Addenda

# Correction: Building a Digital Tool for the Adoption of the World Health Organization's Antenatal Care Recommendations: Methodological Intersection of Evidence, Clinical Logic, and Digital Technology

Samira M Haddad<sup>1,2</sup>, MSc, MD, PhD; Renato T Souza<sup>1,2</sup>, MSc, MD, PhD; Jose Guilherme Cecatti<sup>1,2</sup>, MSc, MD, PhD; Maria Barreix<sup>3</sup>, MHS; Tigest Tamrat<sup>3</sup>, MPH; Carolyn Footitt<sup>4</sup>, MSPH; Garrett L Mehl<sup>3</sup>, PhD; Inraini F Syah<sup>5</sup>, MPH; Anuraj H Shankar<sup>5,6,7</sup>, PhD; Özge Tunçalp<sup>3</sup>, MD, PhD

<sup>1</sup>Department of Obstetrics and Gynecology, School of Medical Sciences, University of Campinas, Campinas, Brazil

<sup>2</sup>Center for Research in Reproductive Health of Campinas (CEMICAMP), Campinas, Brazil

<sup>3</sup>UNDP–UNFPA–UNICEF–WHO–World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Reproductive Health and Research, World Health Organization, Geneva, Switzerland

<sup>4</sup>Ona Systems Inc, Nairobi, Kenya

<sup>5</sup>Summit Institute of Development, Mataram, Indonesia

<sup>6</sup>Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

<sup>7</sup>Eijkman-Oxford Clinical Research Unit, Eijkman Institute for Molecular Biology, Jakarta, Indonesia

**Corresponding Author:**

Jose Guilherme Cecatti, MSc, MD, PhD  
Department of Obstetrics and Gynecology  
School of Medical Sciences  
University of Campinas  
101 Alexander Fleming Street  
University of Campinas  
Campinas, 13083-891  
Brazil  
Phone: 55 1935219482  
Email: [cecatti@unicamp.br](mailto:cecatti@unicamp.br)

**Related Article:**

Correction of: <https://www.jmir.org/2020/10/e16355>

(*J Med Internet Res* 2020;22(10):e24891) doi:[10.2196/24891](https://doi.org/10.2196/24891)

In “Building a Digital Tool for the Adoption of the World Health Organization's Antenatal Care Recommendations: Methodological Intersection of Evidence, Clinical Logic, and Digital Technology” (*J Med Internet Res* 2020;22(10):e16355) the authors noted three errors.

The name of author Tigest Tamrat was incorrectly listed as “Tigist Tamrat”. This has now been changed to the correct spelling.

Under the section “Structured Documentation”, one instance of the phrase “business process mapping (BPM)” has been corrected to “business process mapping notation (BPMN)”.

Figure 6 has been replaced with a new version of subsection B (seen below), as the authors did not have permission to use images displayed in the originally published version of subsection B. Figure 6 subsections A, C, and D have not been changed from the originally published version.

**Figure 6.** Screenshots of the World Health Organization Digital Antenatal Care (WHO digital ANC) module. A: List of patients; B: Individual patient record summary; C: Home screen; D: Patient contact summary.



The correction will appear in the online version of the paper on the JMIR Publications website on October 13, 2020, together with the publication of this correction notice. Because this was

made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

*Submitted 08.10.20; this is a non-peer-reviewed article; accepted 08.10.20; published 13.10.20.*

*Please cite as:*

*Haddad SM, Souza RT, Cecatti JG, Barreix M, Tamrat T, Footitt C, Mehl GL, Syah IF, Shankar AH, Tunçalp Ö*

*Correction: Building a Digital Tool for the Adoption of the World Health Organization's Antenatal Care Recommendations: Methodological Intersection of Evidence, Clinical Logic, and Digital Technology*

*J Med Internet Res 2020;22(10):e24891*

*URL: <http://www.jmir.org/2020/10/e24891/>*

*doi: [10.2196/24891](https://doi.org/10.2196/24891)*

*PMID: [33048822](https://pubmed.ncbi.nlm.nih.gov/33048822/)*

©Samira M Haddad, Renato T Souza, Jose Guilherme Cecatti, Maria Barreix, Tigest Tamrat, Carolyn Footitt, Garrett L Mehl, Inraini F Syah, Anuraj H Shankar, Özge Tunçalp. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 13.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Corrigenda and Addenda

# Correction: Twelve-Month Follow-Up to a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention for Rural Adults With Depression Symptoms: Single-Arm Longitudinal Study

Mark Schure<sup>1</sup>, BSc, MSc, PhD; Bernadette McCrory<sup>2</sup>, MPH, PhD; Kathryn Tuchscherer Franklin<sup>1</sup>, MEd, PhD; John Greist<sup>3</sup>, MD; Ruth Striegel Weissman<sup>3</sup>, PhD

<sup>1</sup>Department of Health & Human Development, Montana State University, Bozeman, MT, United States

<sup>2</sup>Department of Mechanical & Industrial Engineering, Montana State University, Bozeman, MT, United States

<sup>3</sup>Center for Mental Health Research and Recovery, Montana State University, Bozeman, MT, United States

**Corresponding Author:**

Mark Schure, BSc, MSc, PhD

Department of Health & Human Development

Montana State University

305 Herrick Hall

Bozeman, MT, 59717

United States

Phone: 1 406 994 3248

Fax: 1 406 994 2013

Email: [mark.schure@montana.edu](mailto:mark.schure@montana.edu)

**Related Article:**

Correction of: <https://www.jmir.org/2020/10/e21336>

(*J Med Internet Res* 2020;22(10):e25146) doi:[10.2196/25146](https://doi.org/10.2196/25146)

In “Twelve-Month Follow-Up to a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention for Rural Adults With Depression Symptoms: Single-Arm Longitudinal Study” (*J Med Internet Res* 2020;22(10):e21336) the authors noted one error.

The affiliation for authors Mark Schure and Kathryn Tuchscherer Franklin was incorrectly listed as:

*Department of Mechanical & Industrial Engineering,  
Montana State University, Bozeman, MT, United  
States*

The correct affiliation for these authors is:

*Department of Health & Human Development,  
Montana State University, Bozeman, MT, United  
States*

This affiliation is affiliation 1 in the corrected manuscript. Accordingly, affiliations 1 and 2 in the originally published manuscript have been renumbered to affiliations 2 and 3 in the corrected manuscript.

The Corresponding Author address for Mark Schure has also been corrected from:

*Mark Schure, BSc, MSc, PhD*

*Department of Mechanical & Industrial Engineering*

*Montana State University*

*Bozeman, MT*

*United States*

to:

*Mark Schure, BSc, MSc, PhD*

*Department of Health & Human Development*

*Montana State University*

*305 Herrick Hall*

*Bozeman, MT, 59717*

*United States*

The correction will appear in the online version of the paper on the JMIR Publications website on October 23, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.



*Submitted 19.10.20; this is a non-peer-reviewed article; accepted 19.10.20; published 23.10.20.*

Please cite as:

*Schure M, McCrory B, Tuchscherer Franklin K, Geist J, Weissman RS*

*Correction: Twelve-Month Follow-Up to a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention for Rural Adults With Depression Symptoms: Single-Arm Longitudinal Study*

*J Med Internet Res 2020;22(10):e25146*

URL: <http://www.jmir.org/2020/10/e25146/>

doi: [10.2196/25146](https://doi.org/10.2196/25146)

PMID: [33095741](https://pubmed.ncbi.nlm.nih.gov/33095741/)

©Mark Schure, Bernadette McCrory, Kathryn Tuchscherer Franklin, John Geist, Ruth Striegel Weissman. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Viewpoint

# Adapting an Outpatient Psychiatric Clinic to Telehealth During the COVID-19 Pandemic: A Practice Perspective

Farzan Sasangohar<sup>1</sup>, BA, BCS, MASc, SM, PhD; Major R Bradshaw<sup>2</sup>, BS, PhD; Marianne Millen Carlson<sup>2</sup>, BS, MA, PsyD; James N Flack<sup>2</sup>, BA, MD; James C Fowler<sup>2</sup>, BA, PhD; Diana Freeland<sup>2</sup>, BA, MA; John Head<sup>2</sup>, BM, MM, MT-BC, NMT; Kate Marder<sup>2</sup>, BA, MA; William Orme<sup>2</sup>, BS, MA, PhD; Benjamin Weinstein<sup>2</sup>, BS, MD; Jacob M Kolman<sup>1</sup>, BA, MA; Bitu Kash<sup>1</sup>, BS, MBA, PhD; Alok Madan<sup>2</sup>, BA, MA, MPH, PhD

<sup>1</sup>Center for Outcomes Research, Houston Methodist Hospital, Houston, TX, United States

<sup>2</sup>Behavioral Health Clinic, Houston Methodist Hospital, Houston, TX, United States

**Corresponding Author:**

Farzan Sasangohar, BA, BCS, MASc, SM, PhD

Center for Outcomes Research

Houston Methodist Hospital

6565 Fannin St

Houston, TX

United States

Phone: 1 979 458 2337

Email: [sasangohar@tamu.edu](mailto:sasangohar@tamu.edu)

## Abstract

As the demand for telepsychiatry increases during the COVID-19 pandemic, the strengths and challenges of telepsychiatry implementation must be articulated to improve clinical practices in the long term. Currently, observations within US contexts are lacking; therefore, we report on the rapid implementation of telepsychiatry and workflow experiences in a psychiatric practice based within a large health care system in southeast Texas with a national catchment area. We discuss the logistics of the implementation, including modes of communication, scheduling, coordination, and capacity; the psychological effects of web-based services, including both the loss of the physical therapeutic environment and the unique interpersonal dynamics experienced in the virtual environment; and postadoption patterns of engagement with our services and with other clinical functions affected by the rapid adaptation to telemedicine. Our art therapy group programming serves as an applied case study, demonstrating the value of a well-managed web-based program (eg, patients were receptive and well-engaged, and they appreciated the continuity of accessible service) as well as the challenges (eg, the need for backup plans and technological fallbacks, managing interruptions and telecommunication learning curves, and working around the difference in resources for art and music therapy between a well-stocked clinical setting versus clients' home spaces). We conclude from our experience that the overall strengths of telepsychiatry include receptive and well-engaged responses from patients as well as the expansion of boundaries, which provides a directly contextualized view into patients' home lives. Challenges and corresponding recommendations include the need for more careful safety planning for high-risk patients; maintaining professional boundaries in the newly informal virtual setting; designing the physical space to both frame the patient encounter and maintain work-life balance for the therapist; allowing for delays and interruptions (including an initial acclimation session); and preserving interprofessional care team collaboration when the physical locations that normally facilitate such encounters are not accessible. We believe that careful observations of the strengths and challenges of telepsychiatry during this pandemic will better inform practices that are considering telepsychiatry adoption both within pandemic contexts and more broadly thereafter.

(*J Med Internet Res* 2020;22(10):e22523) doi:[10.2196/22523](https://doi.org/10.2196/22523)

**KEYWORDS**

telemedicine; psychiatry; preventive psychiatry; SARS virus; pandemic; prevention; COVID-19; telehealth; perspective

## Introduction

The spread of COVID-19 and wide-scale self-quarantine and shelter-in-place orders have led many nonemergency practices

to adopt telehealth solutions to continue serving their patients [1], and psychiatry is no exception. While some cases will still warrant inpatient psychiatric admissions, including coordinated care for psychiatric inpatients who test positive for COVID-19

[2], telepsychiatry is a viable option for handling outpatient scenarios. Best practices are already in place [3], and there is a supportive evidence base for telepsychiatry [4-6], particularly for web-based therapy for depression [7] and posttraumatic stress disorder [8]. However, the forced adoption of and mass transition to telehealth during the COVID-19 pandemic has resulted in significant challenges for the implementation of telehealth for psychiatric services [9,10]. While guidelines exist for postpandemic telepsychiatry [11,12], given the unique characteristics of this COVID-19 pandemic, lessons learned from these transitions as well as implementation-specific guidelines would be highly applicable to other practices. To date, clinical implementations and lessons learned (and with largely positive results) have been reported largely from settings outside the United States, such as China [13,14], Europe [15-23], Turkey [24], and Australia [10,25-28], with only some US coverage (notably [9], [29], and [30]). In this paper, we document our experience with telehealth adoption at a psychiatry practice embedded in a large health system in the United States and summarize several important successful improvisations and challenges to support broader and more diverse adoption efforts.

Established in January 2018, the psychiatric outpatient clinic at Houston Methodist Hospital offers programs of care to address gaps in the current mental health treatment continuum. The therapeutic outpatient assessment services span approximately 10 business days (2 to 4 hours per day) and are tailored to patients with complex and persistent psychiatric conditions, including depression, anxiety-related disorders, trauma, psychosis, interpersonal dysfunction, chronic pain, sleep disorders, emotional regulation problems, and suicidality. The team-based, multidisciplinary assessment is designed to optimize insight, produce maximal diagnostic clarity, and create a roadmap for future treatment.

The clinic offers access to mental health care through a >5-week Functional Rehabilitation Program that includes morning group therapy sessions and individual sessions in the afternoon. The program proceeds on an outpatient basis, enabling patients to complete treatment while remaining engaged with their lives. The approach to treating patients incorporates an understanding of broader contexts, such as medical comorbidities and social determinants of health. The clinic also offers services to meet the varied and dynamic needs of our patients, including an intensive outpatient program, a modified functional rehabilitation program, individual psychotherapy, couples counseling, psychotropic medication management, and art and music therapies.

In some ways, our clinic had a “head start” on the COVID-19 pandemic compared to most other clinics, as it was already transitioning to digital platforms before the pandemic [31]. The system-wide support for innovative technologies in our health system had resulted in early adoption of and significant investments in telehealth, including secure and integrated technology to facilitate virtual visits embedded within our electronic health record (EHR), a virtual intensive care unit [32], and adoption of telemedicine postoperative follow-up [33]. We had already vested and integrated a platform called CareSense (MedTrak, Inc) for previsit and postvisit care coordination and patient-reported outcomes (see [31] and Fowler

et al, forthcoming). Hence, frequently cited sources of provider hesitation to adopt telemedicine, including workflow integration and infrastructure-related logistics [4,34-37], did not apply to our practice or to our overall culture. However, the COVID-19 pandemic still necessitated rapid adoption of changes to go live before full testing.

## Implementation and Experiences

On March 18, 2020, our outpatient clinic transitioned to a 100% telehealth platform, and it is continuing with this approach as of this writing. Our complete transition and continued use of telehealth technologies contrasts with many other outpatient practices, where clinics have followed suit with local businesses and have opened to provide at least some in-person treatment. We chose a more conservative approach to the ongoing use of telehealth technologies for many reasons. Hospital-wide policies require face masks and six feet of social distancing for all outpatient visits for both patients and providers. These safety requirements reduce the risk of COVID-19 transmission but come at a significant cost to the daily practice of outpatient psychiatric care. Our assessments and interventions rely heavily on nonverbal and verbal communication. Masks cover well over half of patients' faces, limiting providers' assessment of patient affect. Additionally, masks muffle voices. This facet of mask-wearing has the potential to negatively affect the content of verbal output and subsequently affect comprehension for patients and providers alike. Verbal communication from at least six feet away only worsens the potential negative consequences of muffled voices.

### Diversifying Modes of Communication: One Size Does Not Fit All

In our practice, we found it essential to leverage multiple platforms and modalities (eg, Cisco Webex, Microsoft Teams, email, telephone calls, the EHR, patient portal communications) to facilitate the initial transition. While such an inclusive strategy may increase the logistical complexity of care for the providers, our clients found the backup options and redundant systems to be essential. For instance, one client became anxious and frustrated by initial difficulties downloading Webex; therefore, we pivoted to using FaceTime (Apple Inc), which did not require an additional tool and was familiar to the client. Another client, who was feeling lonely and disconnected but struggled to find time for sessions because of childcare demands, was able to feel connected by occasionally texting her therapist. The texts included YouTube links to humorous videos on shared pandemic struggles. In addition to serving a wide range of client preferences, this flexibility was key to help us maintain the availability of service even when one modality lagged or became unresponsive. In addition, having different levels of fidelity on hand (eg, from high-tech Webex to low-tech telephone calls) created a scaffold for backup options and platforms that can be used in tandem. Of course, a list of both COVID-19-provisional approved applications and technology that is more generally compliant with the Health Insurance Portability and Accountability Act (HIPAA) should be consulted to ensure digital security and privacy [38].

Conversely, diversifying the modes of communication also carries efficiency costs (both within the care team and in team-client interactions). These include logging in to platforms and waiting for others to do the same; pausing for deliberate transition between speakers; and reporting the same information in different forms and in multiple places (eg, rapid switches between sessions on Microsoft Teams and formal documentation of the same in the EHR).

### Special Logistical Concerns

Handoff procedures were particularly transformed in the new process. Using the gaps between sessions for quick handoffs results in moving from meeting to meeting without a break and then ending the day with more in-depth documentation. We experienced an increased need for communication between team members; casual hallway or “water-cooler” handoffs frequently occur in physical clinic settings, but these could not occur when the care team shifted to telecommuting. Some team members adapted to the loss of this natural and implicit means of coordination by more explicit efforts, while others “went silent.” We learned that staff should consciously expect and prepare for these easily underestimated differences in communication dynamics.

Management of schedules and appointment times was also more stressful and demanding, as tardiness or absence from appointments became ambiguous. An office no-show can occur for many reasons but at least is always counted as an absence. Web-based absence can result from an otherwise ready client experiencing technical difficulties, a broken or misdirected link, or an adjacent appointment; other no-show reasons include forgetfulness or distractions due to other turmoil (quarantine-related or otherwise). In addition, team members’ and clients’ free time has been transformed, limited or expanded, and distributed differently due to lockdown. Factors such as childcare, other caregiver needs, or the presence or absence of other adults in the home have become factors that affect schedules.

Finally, we found that team members and clients had variable capacity to formulate their communication behaviors and subsequently adapt them to the virtual leap. Technological comfort, literacy, and fluency also played roles. When finally online, users experienced differences in the conversational flow in the web-based environment and had to adjust their interviewing styles, such as avoiding talking over the patient. Video exists in a liminal space between in-person interaction and nonvisual interaction. When face-to-face, interlocutors experience familiar nonverbal cues to signal when a speaker is finished or when a listener is still attentive or becoming unsettled, which informs the turn order of speaking as needed; in radio communications, there is a clear expectation that these cues will be absent (hence, conventions are in place such as saying “over” to replace the cues). Web-based video lacks the fidelity to retain the cues but “teases” users with at least some visual context; therefore, we may not properly reset our expectations (nods are missed in choppy video feeds, pitch cues are missed in the audio transmission, eye contact is disrupted by the offset camera eye, etc.). Delays in communication as speakers are confused by the turn order affect the length and

efficiency of the session, and more time is required to build rapport and trust with new patients.

### Psychological Effects of Web-Based Audiovisual Chat

Our clinic emphasizes differentiation through “time and team” to produce a therapeutic effect. With telehealth, we need to be more mindful of the “time” aspect, which as noted above has been transformed. Web-based chat fatigue (also referred to as “Zoom fatigue” during the COVID-19 pandemic) of patients and team members has sometimes motivated us to limit the duration and frequency of sessions to spread out the mental effort required. Patients still receive the time that our clinic promises; however, they may need to be more flexible with the duration of our programs. As a team, we do not have as much time to talk about patients together because we are confined to team meetings and have strict cutoff times and schedules before the next Webex meeting begins. We also miss the in-person consultations between staff, which have social and not purely logistical benefits. For both our clients and health care workers, a well-established period of reflective *time* can be as therapeutic and rejuvenating as any sacred or holy *space* [39]; therefore, the fragmentation of that time into asynchronous processes such as emails is felt. We have used other technology formats and call or text providers as needed to communicate. The needed information is communicated; however, these consultations are quick and detailed information is lost.

Time and space coalesce to form a therapeutic environment for mental health. Space is now virtual, with sessions taking place in the patient’s home or home office in a manner contiguous with the rest of their day. Thus, they can lose the effect of relocating to a separate and unique space, such as the psychiatrist’s office. In this new space, team members and clients have needed to work harder to generate and sustain interpersonal connection. Social signals and efforts must be amplified to overcome the loss of intimacy and belonging derived from physical proximity. This is difficult when new relationships begin in the virtual environment, in contrast to existing clients, with whom rapport was already established before the transition to virtual sessions. Conversely, video chat garners some closeness because of the virtual invitation into normally private spaces (eg, the therapist’s and client’s home office or kitchen table). Viewers obtain unintended but inevitable glimpses into each other’s home lives (children, pets, spouses roaming in the visual or audio background) [40], and even dress tends to be less formal.

### Services and Service Lines Under Telehealth

We carefully scrutinized which services can transition without an unacceptable loss of integrity and which cannot. Individual therapy is an easy and established service for telehealth [5,7]; group therapy has been received well by providers and clients alike, perhaps because the program is small. Art and music therapy programs had precedents; however, the success of transition was uncertain, and exploratory efforts were required to establish workability, as in our case study below. Neuropsychological services had to be limited to simple screening. These choices were determined by individual workplaces but should not arise in a vacuum.



Some programs may thrive more during quarantine than they otherwise would via telemedicine or other modes. Namely, we suspect that people are currently happy to connect in group programs because they are not seeing anyone else due to quarantine. Although there are still some technological challenges, patients are glad to have a social space. Once clients re-engage with their normal social supports and are not quarantined, participation in group sessions may decrease.

Specific functions of our psychiatric practice have also been transformed. Some psychological testing can now be completed through vendor-supplied web-based portals. Other tests require a physical presence, such as medical tests, genetic testing, and magnetic resonance imaging (MRI), and therefore present a challenge. More time is needed to order these tests and to receive results. We no longer have the luxury of ordering and coordinating care within our system. We must rely on external facilities, and relevant information is not always readily available in our EHR. Releases of information and other paper forms must be completed by hand, scanned by patients, and then emailed to us. Fillable forms (eg, smart PDFs) are more efficient and easier for patients who may lack access to technology or suitable home office supplies.

### Case Study: Art/Music Therapy Group

Art and music therapies are integral components of our uniquely structured Functional Rehabilitation Program; all patients participate in these forms of therapy in both individual and group settings. While more common forms of psychotherapy are also features of the core clinical programming (eg, acceptance and commitment therapy, dialectical behavioral therapy skills training, process-oriented group therapy), we chose to highlight art and music therapy because of the additional logistical challenges these therapies required in transitioning from in-person sessions to a telehealth platform—namely, the need for therapy-specific equipment, such as painting supplies and musical instruments.

The first tele-art therapy group was held by telephone rather than video due to technical difficulties. While it was difficult to conduct an art therapy group session without being able to see anything, the planned session included a writing prompt, and patients were able to share what they had written along with a description of what they had started to create (demonstrating the *importance of backup plans and backup technologies* noted above). The timing was a challenge as well because the session took a long time to get started due to technical issues; however, the patients were open to continuing to work on the art pieces on their own, and they brought their pieces to the next session (*anticipating in-session time spent on technology; managing interruptions*). By the second session, the video was functional, and the patients were able to show their images. The idea of using the group session as the time to start a piece and then continuing to work on the piece or create additional images during the week was well received, with mixed participation (*patients as receptive and well-engaged*). Availability of art materials is another challenge; some people have plentiful supplies at home, while others have more limited supplies. Ordering items for delivery now faces delays in shipping for

many items and low stock due to the increase in the numbers of people who are quarantined or sheltering in place who are ordering art supplies (*setting the in-home therapeutic space*).

Our providers observed that the patients seemed grateful to have the groups as a way to create structure in their days and to connect with others. Pre-existing rapport noticeably helped in this case: while similar online offerings are available, there is a sense of community for existing patients that they would not find by selecting a different offering, such as an online support group or art class. At the same time, our staff served as an anchor that enabled the patients to reach out to other support groups and engage in more opportunities during this time than they might normally have done because they knew they could check in with us about the other things they were trying.

Despite positive outcomes, the applied and physical nature of art and music therapy poses unique challenges, largely revolving around the resources necessary for traditional session plans. The therapist and patient do not have equal and shared access to instruments. If a patient wants to use music education in their therapy, such as learning guitar skills, they do not have access to the music therapist's instruments to achieve this. Time lags in internet video conferencing also make joint music-making difficult. This removes real-time improvisation and some re-creative approaches from the therapist's repertoire until technological adjustments or advancements are made; for example, music-making interventions such as group drumming are removed from the music therapist's toolbox.

In this type of therapy, there is also an increased burden on the patient to learn technological skills to participate in telemedicine. Even if the therapist has the resources to provide high-quality video and audio, that quality can be lost if the patients' technological resources are not equivalent. Higher-quality technology resources are needed to ensure quality experiences more than in conversation-based interventions (which, as we've seen, can default to mechanisms as simple as the telephone). USB microphones and high-speed home internet are not ubiquitous even among fairly privileged patients, raising additional concerns of potentially exacerbating disparities for lower-income patients and their care access to telemedicine programs of this nature. For patients who can obtain the necessary resources, technical literacy may still present a barrier; therefore, the therapist serves a dual role as the patient's information technology support in troubleshooting not only the teleconferencing method in general but also the audiovisual equipment required.

Although not ideal for re-creative or improvisational models, the internet and its associated technologies are rich in opportunities for receptive and compositional methods of music therapy. Most technology platforms (which are free but also require a certain amount of internet bandwidth) enable screen sharing and high-quality video for receptive music therapy interventions. This includes screen sharing for collaboration on lyrics or looking up chord charts on the internet. These web-based techniques take time to learn; therefore, the therapist should budget extra time accordingly for self-orientation and patient orientation. It is important for the therapist to test their new technology skills before entering sessions. The therapist



should practice their new skills with a fellow therapist or family member. Free and easy-to-learn software programs are available for music production and composition, such as GarageBand and Audacity. Granted, “easy-to-learn” is relative to the more complex software available as well as the patient’s or therapist’s baseline technological skills, although learning and improving one’s self-efficacy can have benefits and rewards as well.

## **Summary of Strengths, Challenges, and Recommendations**

Our clinical team identified several strengths and key challenges when moving forward with our rapidly implemented telemedicine program; these may be instructive both to other psychiatric practices that are thinking of adopting telemedicine and to practices seeking to improve currently implemented programs.

### **Strengths**

#### ***Patients Are Receptive and Well-Engaged***

All our prepandemic established patients continued to engage in some degree of programming when we transitioned from an in-person platform to telehealth. We were mindful of individual differences in the extent to which patients were willing to transition to telehealth programming and in their preferred platforms. Patients and providers engaged in a shared decision-making process, and we tailored our programming and mode of service delivery based on patient preferences. Postpandemic patients generally chose a reduced intensity of treatment from the beginning of their treatment with us. Overall, most patients seemed relieved that in the midst of the pandemic, therapeutic structures and support were still accessible. Rapport is enhanced by a sense that both providers and patients are going through a crisis together; this humanizes the providers to the patients while also giving providers an opportunity to model intentional coping, balanced responses to the crisis (on most occasions), and active anxiety regulation in real time. Our experience is corroborated by other US sites that implemented telemedicine rapidly [9] as well as by generally positive reviews showing patient satisfaction with telepsychiatry [5,6].

#### ***Virtual Groups Have Been Well-Attended, and Engagement in Some Respects Has Increased***

The relative isolation that group members are feeling has spurred them out of their comfort zones to engage more actively and depend more on their peers. For some group members, especially those with more avoidance and social anxiety, the current circumstances are leading them by necessity to experience their peers as a new and positive source of support.

#### ***Telepsychiatry Expands the Boundaries of Psychological Intervention Into the Real World***

By necessity, telehealth visits take providers out of the one-on-one environment of traditional brick-and-mortar practice and into their patients’ everyday lives, living rooms, and backyards. In one session, we were actually able to see one patient relate with her son in real time instead of just talking with her about their interaction and interpreting it as filtered

through her perception. This direct and contextualized interaction increases the ecological validity and relevance of therapy and opens opportunities for more flexible service delivery models in future practice that are more effective and helpful for patients and less restrained by antiquated “red tape.”

### **Challenges and Recommendations**

#### ***High-Risk Patients***

High-risk patients require more active safety planning, case management, and tolerance of anxiety and uncertainty related to unknown aspects due to lack of in-person visits. For our established chronically suicidal patients, we made emergency contact details readily available prior to scheduled sessions, expressed the ability and willingness to engage emergency medical services if necessary, and provided direct access to our inpatient psychiatric unit should the need arise. Fortunately, we did not need to mobilize any of these contingencies for prepandemic established patients. However, for patients who were admitted to our practice after we transitioned to an exclusively virtual platform, risk management created significant logistical challenges. We did facilitate inpatient psychiatric admissions with significant assistance from family members, who needed to bring their loved ones to the hospital emergency department, were not allowed to accompany them past that point, and were required to communicate with the inpatient care team almost exclusively by telephone. Additionally, we supplemented our outpatient programming with home health aides out of medical necessity. Management of risk in a virtual setting requires considerably more time, effort, and creative problem-solving than in-person clinical care [20]. To date, we have managed risk without adverse events. In this area, further research on the process, patient and provider experience, and outcomes is warranted.

#### ***Establish a Provider Culture of Telemedicine Adoption***

One observed strength was not a strength of our telemedicine program per se but of provider willingness to adopt telemedicine as a solution; 100% of providers transitioned to telemedicine platforms with varying degrees of fluency and need for ongoing logistical support. Studies have reported provider hesitation to adopt telemedicine based on concerns about workload [34], usability [35,36], EHR and systems integration [36], other logistics related to time and staff [37], and incentives, reimbursements, and regulations [35,36], especially in the case of telepsychiatry, which deserves special consideration when COVID-19–based federal policy changes and leniencies expire [4,41]. Hence, the success of telemedicine adoption is greatly facilitated by investing in the cultural, logistical, and infrastructural factors to enable success (see also Kalin et al [29], who likewise credit their successful COVID-19–related telepsychiatry adoption to organizational culture and prior technological readiness). However, “from-scratch” implementation is also possible, as in the case reported by Yellowlees et al [9], who impressively planned a complete program to adapt in-person sessions to telepsychiatry in one day and had fully implemented it for all patients within three working days. Other attempts were more encumbered by technical and training difficulties; however, it was found, as in

our study, that use of multiple platforms had a mitigating effect [30].

### ***Set the Therapeutic Frame With Intentionality to Manage Risk***

The transition to a virtual frame risks blurring normal therapeutic boundaries by decreasing the formality of the encounter. For example, there is a need to demarcate a telephone-based therapy appointment from a telephone call with a friend or a fellow citizen in the midst of a pandemic. In an office, this frame is set through physical elements, including the office, furniture, and waiting room. In virtual space, this frame must be instituted verbally, though not rigidly, through maintaining intentionality about how the conversation relates to the patient's overall goals. Mindfulness and management of the professional practitioner-patient relationship remains as important as ever. It is also worth noting that most web platforms allow for anonymous telephone calls or web-calling using a web-based profile, which allows the therapist to avoid disclosing their own personal account or telephone number when checking in with the patient.

### ***Setting the Physical and Visual Backdrop Requires Forethought and Design***

Setting the frame requires more than verbal and behavioral monitoring. Although it can be humanizing for the patient to see the therapist in their natural habitat, it may be necessary to hide unattractive objects in the therapist's home. A foldable standing blind can be useful for hiding items such as laundry baskets, beds, and the kitchen counter. The blind also provides a consistent, professional, and aesthetically pleasing therapeutic environment. Lighting is also a consideration; for webcasting, it is better to be lit from the front than from the back. To avoid additional glare from the ceiling fan, the laptop computer or camera should be positioned so that the ceiling fan is not in view.

### ***Maintain Work-Life Balance by Separation of Space and Time***

The blinds and other physical rearrangements needed for the therapeutic space can also benefit the therapist as a transitory demarcation of work versus home spaces. Home is a refuge from work, and working from home blurs the roles of the home and work environments. Using a blind to "set up" the office changes the environment from home to work and vice versa at the end of the day. This transition helps to "give back" the space to the therapist and acts as an affordance to return to normal use of the space. If the therapist's floor plan allows for a permanent home office space, this can also demarcate a work zone from a home zone; however, the lack of physical actions and motions for setup and breakdown in this case should be replaced by disciplined scheduling to signal work time and free time in some other way. It is worth noting that health care workers, including psychiatrists and mental health workers, can also benefit from receiving telepsychiatry services [24].

### ***Introduce the Client to the Virtual Space and Expect to Take Up Initial Session Time***

When beginning initial sessions, it is important to allow time for orientation to the technology [9]. In the same way a new patient may need help to find their way into the building and park their car, they may need our help to orient to the technology. It is important for the therapist to take a "walking with" stance with the patient and express that the therapist and the patient are "in this together." Allowing time for technology and normalizing it as part of the orientation to therapy can help assuage a new patient's anxiety about their first session. The therapist should aside a generous 10 or 15 minutes simply to help the patient with the technology. This also has the effect of building trust between the patient and the therapist. Provider-guided orientation may thus be uniquely warranted in telepsychiatry in contrast to other forms of telemedicine, in which the provider and patient may both prefer to save time by using pre-session tutorials or medical assistant-led guidance to use the technology instead.

### ***Interruptions May Disrupt Session Plans and Will Require Management***

Potential interruptions naturally arise as the patient's family or pets enter the picture, particularly if home childcare is an issue for either the patient or the therapist. These interruptions are inevitable and must be managed. In group sessions, this may cause patients to miss key discussions. Interruption recovery techniques include video replay (if sessions are recorded) or history logs (if time-stamped real-time notes are taken) [42]. Privacy issues can be mitigated through sound localization. Headphones are fairly essential, as they prevent family members from overhearing confidential session details or comments by therapy group members. Headphones also signal to others that the patient is absorbed in a task. If a meditative or focused exercise is planned, such as guided imagery or progressive muscle relaxation, it is important to let patients know this ahead of time so that they can take steps to help their family understand that they cannot be disturbed (and establish a backup plan in case of internet connectivity disruptions). Therapists and patients alike can also look into any affordable or employer-provided childcare options they feel safe with during the pandemic to mitigate childcare-related interruptions while remaining consistent with their household's social distancing plan.

### ***Shelter-in-Place Restrictions Challenge Active Interprofessional Collaboration***

Clinical isolation highlights the value of in-person informal consultation and colocation of a multidisciplinary team, both to preserve benefits to the patient derived from such team support [20,24] and for the morale of the team itself. Having short conversations in the hall, dropping by a colleague's office for a minute, and eating lunch with colleagues keep providers apprised of shared clinical work. Without these channels, more active time and participation are needed. We have implemented additional scheduled consultation meetings to account for the loss of informal consultation.

## Conclusion

The circumstances of the COVID-19 pandemic have created unprecedented challenges that are taxing the physical and mental

health of the populace; however, the pandemic has also created unprecedented opportunities to learn, teach, innovate, and evaluate telepsychiatry strategies as necessity spurs their adoption.

## Conflicts of Interest

None declared.

## References

- Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. *JMIR Public Health Surveill* 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: [10.2196/18810](https://doi.org/10.2196/18810)] [Medline: [32238336](https://pubmed.ncbi.nlm.nih.gov/32238336/)]
- Percudani M, Corradin M, Moreno M, Indelicato A, Vita A. Mental Health Services in Lombardy during COVID-19 outbreak. *Psychiatry Res* 2020 Jun;288:112980 [FREE Full text] [doi: [10.1016/j.psychres.2020.112980](https://doi.org/10.1016/j.psychres.2020.112980)] [Medline: [32315881](https://pubmed.ncbi.nlm.nih.gov/32315881/)]
- Best Practices in Videoconferencing-Based Telemental Health. American Psychiatric Association and American Telemedicine Association. 2018 Apr. URL: <https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/Telepsychiatry/APA-ATA-Best-Practices-in-Videoconferencing-Based-Telemental-Health.pdf> [accessed 2020-09-28]
- Chen JA, Chung W, Young SK, Tuttle MC, Collins MB, Darghouth SL, et al. COVID-19 and telepsychiatry: Early outpatient experiences and implications for the future. *Gen Hosp Psychiatry* 2020;66:89-95 [FREE Full text] [doi: [10.1016/j.genhosppsych.2020.07.002](https://doi.org/10.1016/j.genhosppsych.2020.07.002)] [Medline: [32750604](https://pubmed.ncbi.nlm.nih.gov/32750604/)]
- Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. *World J Psychiatry* 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: [10.5498/wjp.v6.i2.269](https://doi.org/10.5498/wjp.v6.i2.269)] [Medline: [27354970](https://pubmed.ncbi.nlm.nih.gov/27354970/)]
- Chakrabarti S. Usefulness of telepsychiatry: A critical evaluation of videoconferencing-based approaches. *World J Psychiatry* 2015 Sep 22;5(3):286-304 [FREE Full text] [doi: [10.5498/wjp.v5.i3.286](https://doi.org/10.5498/wjp.v5.i3.286)] [Medline: [26425443](https://pubmed.ncbi.nlm.nih.gov/26425443/)]
- Wright JH, Caudill R. Remote Treatment Delivery in Response to the COVID-19 Pandemic. *Psychother Psychosom* 2020;89(3):130-132 [FREE Full text] [doi: [10.1159/000507376](https://doi.org/10.1159/000507376)] [Medline: [32213775](https://pubmed.ncbi.nlm.nih.gov/32213775/)]
- Acierno R, Gros DF, Ruggiero KJ, Hernandez-Tejada BMA, Knapp RG, Lejuez CW, et al. Behavioral activation and therapeutic exposure for posttraumatic stress disorder: a noninferiority trial of treatment delivered in person versus home-based telehealth. *Depress Anxiety* 2016 May;33(5):415-423. [doi: [10.1002/da.22476](https://doi.org/10.1002/da.22476)] [Medline: [26864655](https://pubmed.ncbi.nlm.nih.gov/26864655/)]
- Yellowlees P, Nakagawa K, Pakyurek M, Hanson A, Elder J, Kales HC. Rapid Conversion of an Outpatient Psychiatric Clinic to a 100% Virtual Telepsychiatry Clinic in Response to COVID-19. *Psychiatr Serv* 2020 Jul 01;71(7):749-752. [doi: [10.1176/appi.ps.202000230](https://doi.org/10.1176/appi.ps.202000230)] [Medline: [32460683](https://pubmed.ncbi.nlm.nih.gov/32460683/)]
- Zulfic Z, Liu D, Lloyd C, Rowan J, Schubert KO. Is telepsychiatry care a realistic option for community mental health services during the COVID-19 pandemic? *Aust N Z J Psychiatry* 2020 Jun 22;48:67420937788. [doi: [10.1177/0004867420937788](https://doi.org/10.1177/0004867420937788)] [Medline: [32571079](https://pubmed.ncbi.nlm.nih.gov/32571079/)]
- Augusterfer EF, Mollica RF, Lavelle J. A review of telemental health in international and post-disaster settings. *Int Rev Psychiatry* 2015;27(6):540-546. [doi: [10.3109/09540261.2015.1082985](https://doi.org/10.3109/09540261.2015.1082985)] [Medline: [26576720](https://pubmed.ncbi.nlm.nih.gov/26576720/)]
- Smith K, Ostinelli E, Macdonald O, Cipriani A. COVID-19 and Telepsychiatry: Development of Evidence-Based Guidance for Clinicians. *JMIR Ment Health* 2020 Aug 28;7(8):e21108 [FREE Full text] [doi: [10.2196/21108](https://doi.org/10.2196/21108)] [Medline: [32658857](https://pubmed.ncbi.nlm.nih.gov/32658857/)]
- Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. *Lancet Psychiatry* 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: [10.1016/S2215-0366\(20\)30077-8](https://doi.org/10.1016/S2215-0366(20)30077-8)] [Medline: [32085841](https://pubmed.ncbi.nlm.nih.gov/32085841/)]
- Zhou J, Liu L, Xue P, Yang X, Tang X. Mental Health Response to the COVID-19 Outbreak in China. *Am J Psychiatry* 2020 Jul 01;177(7):574-575. [doi: [10.1176/appi.ajp.2020.20030304](https://doi.org/10.1176/appi.ajp.2020.20030304)] [Medline: [32375540](https://pubmed.ncbi.nlm.nih.gov/32375540/)]
- Corruble E. A Viewpoint From Paris on the COVID-19 Pandemic: A Necessary Turn to Telepsychiatry. *J Clin Psychiatry* 2020 Mar 31;81(3):online [FREE Full text] [doi: [10.4088/JCP.20com13361](https://doi.org/10.4088/JCP.20com13361)] [Medline: [32237302](https://pubmed.ncbi.nlm.nih.gov/32237302/)]
- Čosić K, Popović S, Šarlija M, Kesedžić I. Impact of Human Disasters and COVID-19 Pandemic on Mental Health: Potential of Digital Psychiatry. *Psychiatr Danub* 2020;32(1):25-31 [FREE Full text] [doi: [10.24869/psyd.2020.25](https://doi.org/10.24869/psyd.2020.25)] [Medline: [32303026](https://pubmed.ncbi.nlm.nih.gov/32303026/)]
- Fagiolini A, Cuomo A, Frank E. COVID-19 Diary From a Psychiatry Department in Italy. *J Clin Psychiatry* 2020 Mar 31;81(3):online [FREE Full text] [doi: [10.4088/JCP.20com13357](https://doi.org/10.4088/JCP.20com13357)] [Medline: [32237301](https://pubmed.ncbi.nlm.nih.gov/32237301/)]
- Prisco V, Prisco L, Donnarumma B. Telepsychiatry in adults and adolescents: a useful tool against CoViD-19. Article in Italian. *Recenti Prog Med* 2020;111(7):411-414. [doi: [10.1701/3407.33923](https://doi.org/10.1701/3407.33923)] [Medline: [32658880](https://pubmed.ncbi.nlm.nih.gov/32658880/)]
- Schuh Teixeira AL, Spadini AV, Pereira-Sanchez V, Ojeahere MI, Morimoto K, Chang A, et al. The urge to implement and expand telepsychiatry during the COVID-19 crisis: Early career psychiatrists' perspective. *Rev Psiquiatria Salud Ment* 2020;13(3):174-175 [FREE Full text] [doi: [10.1016/j.rpsm.2020.06.001](https://doi.org/10.1016/j.rpsm.2020.06.001)] [Medline: [32653319](https://pubmed.ncbi.nlm.nih.gov/32653319/)]
- O'Brien M, McNicholas F. The use of telepsychiatry during COVID-19 and beyond. *Ir J Psychol Med* 2020 May 21:1-6 [FREE Full text] [doi: [10.1017/ipm.2020.54](https://doi.org/10.1017/ipm.2020.54)] [Medline: [32434596](https://pubmed.ncbi.nlm.nih.gov/32434596/)]



21. Olwill C, Mc Nally D, Douglas L. Psychiatrist experience of remote consultations by telephone in an outpatient psychiatric department during the COVID-19 pandemic. *Ir J Psychol Med* 2020 May 22;1-8 [FREE Full text] [doi: [10.1017/ipm.2020.51](https://doi.org/10.1017/ipm.2020.51)] [Medline: [32438945](https://pubmed.ncbi.nlm.nih.gov/32438945/)]
22. Haxhihamza K, Arsova S, Bajraktarov S, Kalpak G, Stefanovski B, Novotni A, et al. Patient Satisfaction with Use of Telemedicine in University Clinic of Psychiatry: Skopje, North Macedonia During COVID-19 Pandemic. *Telemed J E Health* 2020 Aug 17;online ahead of print. [doi: [10.1089/tmj.2020.0256](https://doi.org/10.1089/tmj.2020.0256)] [Medline: [32809916](https://pubmed.ncbi.nlm.nih.gov/32809916/)]
23. Usman M, Fahy S. Coping with the COVID-19 crisis: an overview of service adaptation and challenges encountered by a rural Psychiatry of Later Life (POLL) team. *Ir J Psychol Med* 2020 Jul 02;1-5 [FREE Full text] [doi: [10.1017/ipm.2020.86](https://doi.org/10.1017/ipm.2020.86)] [Medline: [32611473](https://pubmed.ncbi.nlm.nih.gov/32611473/)]
24. Dursun OB, Turan B, Pakyürek M, Tekin A. Integrating Telepsychiatric Services into the Conventional Systems for Psychiatric Support to Health Care Workers and Their Children During COVID-19 Pandemics: Results from A National Experience. *Telemed J E Health* 2020 Aug 18;online ahead of print. [doi: [10.1089/tmj.2020.0237](https://doi.org/10.1089/tmj.2020.0237)] [Medline: [32821025](https://pubmed.ncbi.nlm.nih.gov/32821025/)]
25. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The Role of Telehealth in Reducing the Mental Health Burden from COVID-19. *Telemed J E Health* 2020 Apr;26(4):377-379. [doi: [10.1089/tmj.2020.0068](https://doi.org/10.1089/tmj.2020.0068)] [Medline: [32202977](https://pubmed.ncbi.nlm.nih.gov/32202977/)]
26. Looi JC, Pring W. Private metropolitan telepsychiatry in Australia during Covid-19: current practice and future developments. *Australas Psychiatry* 2020 Jun 02;1039856220930675. [doi: [10.1177/1039856220930675](https://doi.org/10.1177/1039856220930675)] [Medline: [32484737](https://pubmed.ncbi.nlm.nih.gov/32484737/)]
27. Reay RE, Looi JC, Keightley P. Telehealth mental health services during COVID-19: summary of evidence and clinical practice. *Australas Psychiatry* 2020 Jul 28;1039856220943032. [doi: [10.1177/1039856220943032](https://doi.org/10.1177/1039856220943032)] [Medline: [32722963](https://pubmed.ncbi.nlm.nih.gov/32722963/)]
28. Kavoor AR, Chakravarthy K, John T. Remote consultations in the era of COVID-19 pandemic: Preliminary experience in a regional Australian public acute mental health care setting. *Asian J Psychiatr* 2020 Jun;51:102074 [FREE Full text] [doi: [10.1016/j.ajp.2020.102074](https://doi.org/10.1016/j.ajp.2020.102074)] [Medline: [32294583](https://pubmed.ncbi.nlm.nih.gov/32294583/)]
29. Kalin ML, Garlow SJ, Thertus K, Peterson MJ. Rapid Implementation of Telehealth in Hospital Psychiatry in Response to COVID-19. *Am J Psychiatry* 2020 Jul 01;177(7):636-637. [doi: [10.1176/appi.ajp.2020.20040372](https://doi.org/10.1176/appi.ajp.2020.20040372)] [Medline: [32605442](https://pubmed.ncbi.nlm.nih.gov/32605442/)]
30. Sharma A, Sasser T, Schoenfelder Gonzalez E, Vander Stoep A, Myers K. Implementation of Home-Based Telemental Health in a Large Child Psychiatry Department During the COVID-19 Crisis. *J Child Adolesc Psychopharmacol* 2020 Sep;30(7):404-413. [doi: [10.1089/cap.2020.0062](https://doi.org/10.1089/cap.2020.0062)] [Medline: [32639849](https://pubmed.ncbi.nlm.nih.gov/32639849/)]
31. Fowler JC, Madan A, Frueh BC, Bradshaw M, Flack J, Weinstein B. Lessons learned while integrating patient-reported outcomes in a psychiatric hospital. *Psychotherapy (Chic)* 2019 Mar;56(1):91-99. [doi: [10.1037/pst0000174](https://doi.org/10.1037/pst0000174)] [Medline: [30431293](https://pubmed.ncbi.nlm.nih.gov/30431293/)]
32. Sasangohar F, Jones S, Masud F, Vahidy F, Kash B. Provider Burnout and Fatigue During the COVID-19 Pandemic: Lessons Learned From a High-Volume Intensive Care Unit. *Anesth Analg* 2020 Jul;131(1):106-111 [FREE Full text] [doi: [10.1213/ANE.0000000000004866](https://doi.org/10.1213/ANE.0000000000004866)] [Medline: [32282389](https://pubmed.ncbi.nlm.nih.gov/32282389/)]
33. Larsen EP, Bonet SA, Sultana I, Zheng F, Sasangohar F, Kash BA. COR Project Report: Adoption of telemedicine for post-surgical follow-up visits. Center for Outcomes Research, Houston Methodist Research Institute. 2019 Sep. URL: <https://1vv82dtluny8l2py3ufo882r-wpengine.netdna-ssl.com/wp-content/uploads/sites/23/2020/06/Telemed-Adoption.pdf> [accessed 2020-09-28]
34. Fish A, George S, Terrien E, Eccles A, Baker R, Ogunyemi O. Workflow concerns and workarounds of readers in an urban safety net teleretinal screening study. *AMIA Annu Symp Proc* 2011;2011:417-426 [FREE Full text] [Medline: [22195095](https://pubmed.ncbi.nlm.nih.gov/22195095/)]
35. Uscher-Pines L, Kahn JM. Barriers and facilitators to pediatric emergency telemedicine in the United States. *Telemed J E Health* 2014 Nov;20(11):990-996 [FREE Full text] [doi: [10.1089/tmj.2014.0015](https://doi.org/10.1089/tmj.2014.0015)] [Medline: [25238565](https://pubmed.ncbi.nlm.nih.gov/25238565/)]
36. Koopman RJ, Wakefield BJ, Johannning JL, Keplinger LE, Kruse RL, Bomar M, et al. Implementing home blood glucose and blood pressure telemonitoring in primary care practices for patients with diabetes: lessons learned. *Telemed J E Health* 2014 Mar;20(3):253-260 [FREE Full text] [doi: [10.1089/tmj.2013.0188](https://doi.org/10.1089/tmj.2013.0188)] [Medline: [24350806](https://pubmed.ncbi.nlm.nih.gov/24350806/)]
37. Shaw RJ, Kaufman MA, Bosworth HB, Weiner BJ, Zullig LL, Lee SD, et al. Organizational factors associated with readiness to implement and translate a primary care based telemedicine behavioral program to improve blood pressure control: the HTN-IMPROVE study. *Implement Sci* 2013 Sep 08;8:106 [FREE Full text] [doi: [10.1186/1748-5908-8-106](https://doi.org/10.1186/1748-5908-8-106)] [Medline: [24010683](https://pubmed.ncbi.nlm.nih.gov/24010683/)]
38. Cowan A, Johnson R, Close H. Telepsychiatry in Psychotherapy Practice. *Innov Clin Neurosci* 2020 Apr 01;17(4-6):23-26 [FREE Full text] [Medline: [32802589](https://pubmed.ncbi.nlm.nih.gov/32802589/)]
39. Heschel A. *The Sabbath*. New York, NY: Farrar Straus Giroux; 2005.
40. Kahn MW. Pandemic and Persona. *N Engl J Med* 2020 Jul 02;383(1):e1. [doi: [10.1056/NEJMp2010377](https://doi.org/10.1056/NEJMp2010377)] [Medline: [32374955](https://pubmed.ncbi.nlm.nih.gov/32374955/)]
41. Sousa AD, Karia S. Telepsychiatry during COVID-19: Some clinical, public health, and ethical dilemmas. *Indian J Public Health* 2020 Jun;64(Supplement):S245-S246 [FREE Full text] [doi: [10.4103/ijph.IJPH\\_511\\_20](https://doi.org/10.4103/ijph.IJPH_511_20)] [Medline: [32496267](https://pubmed.ncbi.nlm.nih.gov/32496267/)]
42. Sasangohar F, Scott SD, Donmez B. Interruption Management and Recovery in Time-critical Supervisory-level Tasks. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2013 Sep 30 Presented at: International Human Factors and Ergonomics Society Annual Meeting; September 30-October 4, 2013; San Diego, CA p. 1745-1749. [doi: [10.1177/1541931213571389](https://doi.org/10.1177/1541931213571389)]

## Abbreviations

**EHR:** electronic health record

**HIPAA:** Health Insurance Portability and Accountability Act

**MRI:** magnetic resonance imaging

*Edited by G Eysenbach; submitted 14.07.20; peer-reviewed by P Yellowlees, B Johnson, T Aslanidis; comments to author 25.08.20; revised version received 04.09.20; accepted 14.09.20; published 01.10.20.*

*Please cite as:*

Sasangohar F, Bradshaw MR, Carlson MM, Flack JN, Fowler JC, Freeland D, Head J, Marder K, Orme W, Weinstein B, Kolman JM, Kash B, Madan A

*Adapting an Outpatient Psychiatric Clinic to Telehealth During the COVID-19 Pandemic: A Practice Perspective*

*J Med Internet Res* 2020;22(10):e22523

URL: <https://www.jmir.org/2020/10/e22523>

doi: [10.2196/22523](https://doi.org/10.2196/22523)

PMID: [32936768](https://pubmed.ncbi.nlm.nih.gov/32936768/)

©Farzan Sasangohar, Major R Bradshaw, Marianne Millen Carlson, James N Flack, James C Fowler, Diana Freeland, John Head, Kate Marder, William Orme, Benjamin Weinstein, Jacob M Kolman, Bitu Kash, Alok Madan. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 01.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Original Paper

# CoV-Seq, a New Tool for SARS-CoV-2 Genome Analysis and Visualization: Development and Usability Study

Boxiang Liu<sup>1\*</sup>, PhD; Kaibo Liu<sup>1\*</sup>, MSc; He Zhang<sup>1</sup>, MSc; Liang Zhang<sup>1,2</sup>, MSc; Yuchen Bian<sup>1</sup>, PhD; Liang Huang<sup>1,2</sup>, PhD

<sup>1</sup>Baidu Research, Sunnyvale, CA, United States

<sup>2</sup>School of Electrical Engineering & Computer Science, Oregon State University, Corvallis, OR, United States

\*these authors contributed equally

**Corresponding Author:**

Boxiang Liu, PhD  
Baidu Research  
1195 Bordeaux Drive  
Sunnyvale, CA, 94089  
United States  
Phone: 1 3093108265  
Email: [jollier.liu@gmail.com](mailto:jollier.liu@gmail.com)

## Abstract

**Background:** COVID-19 became a global pandemic not long after its identification in late 2019. The genomes of SARS-CoV-2 are being rapidly sequenced and shared on public repositories. To keep up with these updates, scientists need to frequently refresh and reclean data sets, which is an ad hoc and labor-intensive process. Further, scientists with limited bioinformatics or programming knowledge may find it difficult to analyze SARS-CoV-2 genomes.

**Objective:** To address these challenges, we developed CoV-Seq, an integrated web server that enables simple and rapid analysis of SARS-CoV-2 genomes.

**Methods:** CoV-Seq is implemented in Python and JavaScript. The web server and source code URLs are provided in this article.

**Results:** Given a new sequence, CoV-Seq automatically predicts gene boundaries and identifies genetic variants, which are displayed in an interactive genome visualizer and are downloadable for further analysis. A command-line interface is available for high-throughput processing. In addition, we aggregated all publicly available SARS-CoV-2 sequences from the Global Initiative on Sharing Avian Influenza Data (GISAID), National Center for Biotechnology Information (NCBI), European Nucleotide Archive (ENA), and China National GeneBank (CNCB), and extracted genetic variants from these sequences for download and downstream analysis. The CoV-Seq database is updated weekly.

**Conclusions:** We have developed CoV-Seq, an integrated web service for fast and easy analysis of custom SARS-CoV-2 sequences. The web server provides an interactive module for the analysis of custom sequences and a weekly updated database of genetic variants of all publicly accessible SARS-CoV-2 sequences. We believe CoV-Seq will help improve our understanding of the genetic underpinnings of COVID-19.

(*J Med Internet Res* 2020;22(10):e22299) doi:[10.2196/22299](https://doi.org/10.2196/22299)

**KEYWORDS**

COVID-19; SARS-CoV-2; bioinformatics; genetics; genome; virus; sequence; data sets; programming; web server

## Introduction

Since its identification in late 2019, the novel coronavirus SARS-CoV-2 has caused an outbreak of viral pneumonia and has become a global pandemic. Despite efforts to contain its spread, as of late September 2020, SARS-CoV-2 had infected nearly 33 million patients and caused nearly 1 million deaths worldwide [1]. To understand its evolution and genetics,

scientists have sequenced SARS-CoV-2 genomes from patients across different age groups, genders, ethnicities, locations, and disease stages [2]. These genomic sequences are being shared on public repositories at a rapid pace, with thousands of new sequences every week [3-5]. To keep up with the latest developments, scientists need to frequently download and clean new data sets, which is an ad hoc and time-consuming process. Furthermore, scientists with limited knowledge of bioinformatics

or programming may experience difficulty in analyzing SARS-CoV-2 genomes.

We developed the CoV-Seq toolkit to address these challenges. CoV-Seq consists of several components: a data analysis pipeline that takes FASTA sequences and generates variant callsets in variant call format (VCF) and open reading frame (ORF) predictions. The pipeline automatically filters low-quality sequences, removes duplicate sequences, performs sequence alignment, and identifies and annotates genetic variants. We provide a web server [6] to enable the rapid analysis of custom sequences without any programming. The web interface includes an interactive genome visualizer and tabulated displays of genetic variants and ORF predictions. All results can be downloaded for downstream analysis. Further, we provide a command-line interface to allow high-throughput processing in local environments. To facilitate data sharing, we aggregate SARS-CoV-2 sequences from the Global Initiative on Sharing Avian Influenza Data (GISAID) [3], National Center for Biotechnology Information (NCBI) [4], European Nucleotide Archive (ENA) [5], and China National GeneBank (CNGB).

## Methods

### Data Collection

The majority of publicly available SARS-CoV-2 genomic sequences are deposited into the following databases: GISAID (Multimedia Appendix 2 lists GISAID contributors), NCBI, ENA, and CNGB. All databases provide the option of downloading data in a batch. We used Selenium [7] to automate the data download process.

### Data Preprocessing

We aggregated SARS-CoV-2 sequences from GISAID, NCBI, ENA, and CNGB. Many sequences represented incomplete genomes, sometimes containing only a single gene. We filtered these genomes using a lenient cutoff of 25,000 nucleotides because doing so removed distinctly incomplete genomes while retaining complete genomes (Figure 1 in Multimedia Appendix 1). Both NCBI and ENA are part of the International Nucleotide Sequence Database Collaboration (INSDC) and therefore contain duplicate submissions, which we removed by comparing the accession IDs. Further, dual submissions can appear in both GISAID and INSDC under different accession IDs. We considered two submissions as suspect duplications if they had identical genomic sequences. These suspect duplications were marked in the metadata but not removed because a strain can infect multiple patients.

### Pairwise Sequence Alignment

We performed pairwise alignment against the reference sequence (NCBI accession ID: NC\_045512.2) using Multiple Alignment using Fast Fourier Transform (MAFFT) [8] with default options.

### Variant Calling

We used a custom Python script for variant calling, in which we considered single nucleotide polymorphisms (SNPs), insertions, and deletions. We left-normalized each variant with bcftools [9] and removed samples with too many variants, indicative of sequencing error. We used a lenient cutoff of 350

variants because doing so removed samples with extremely large numbers of variants while keeping most samples (Figure 2 in Multimedia Appendix 1). During postprocessing, we removed multiallelic sites because these sites were more likely to occur in regions prone to sequencing error, such as the two ends of the genome (Figure 3 in Multimedia Appendix 1). Further, we removed variants within the poly-A tail. The filtered variant callset was annotated with snpEff [10].

### ORF Boundary Detection

To detect ORF boundaries, we employed a method similar to Viral Annotation Pipeline and iDentification (VAPiD) [11]. Since NCBI Genbank provides genetic annotations for the SARS-CoV-2 reference genomes, we translated the coordinates of ORF boundaries from the reference genome to the query genome using their pairwise alignment. For multisegment ORFs (ie, due to ribosomal slippage), we annotated each segment independently and combined them afterward.

### Interactive Visualization

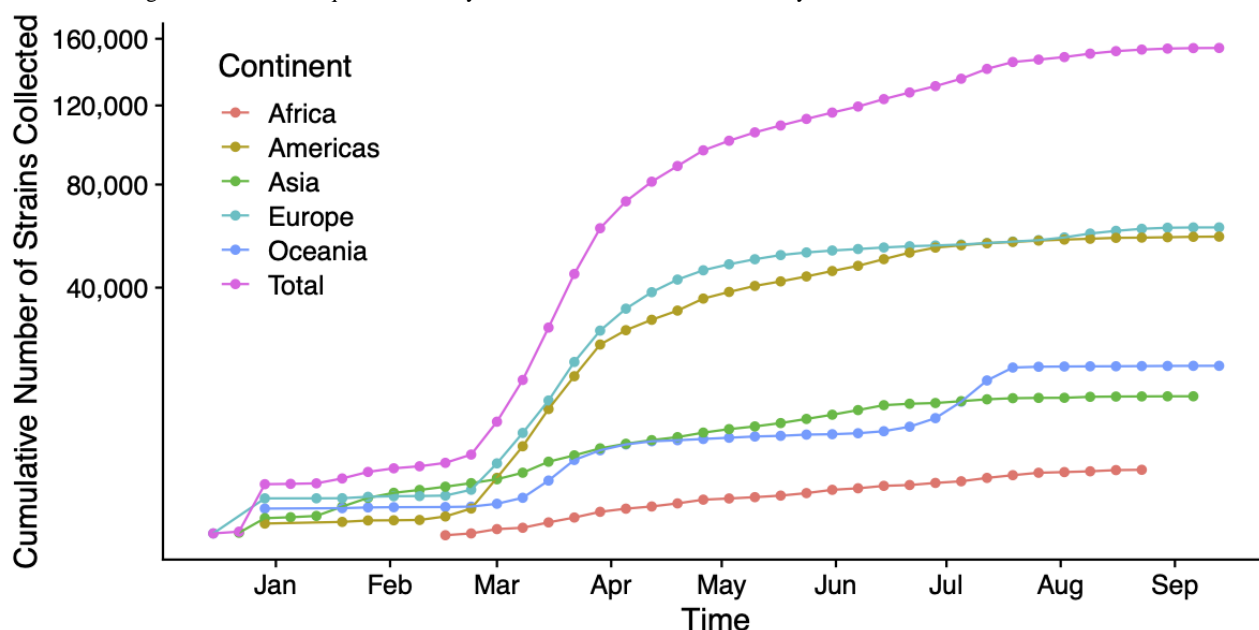
The CoV-Seq web server is hosted on Amazon Web Services (AWS) Elastic Beanstalk with load balancer enabled, running with the Flask framework [12], Jinja template engine [13], and Werkzeug [14] Web Server Gateway Interface (WSGI) toolkit. After the user submits data through either a text box (for a single sequence) or file upload (for an arbitrary number of sequences), the back-end program will perform pairwise sequence alignment, variant calling, and ORF boundary detection to generate results in VCF and JSON formats. The front-end templates will then render genome sequences on a new page using JSON data, the ECharts [15] library, and dedicated designed JavaScript functions. The result page highlights mutations on submitted sequences against the reference sequence and shows details upon cursor hover. Users can also zoom in to check the details of the genomic sequence. Selecting a specific ORF will expand the display to show the mutation table, the ORF table, and the gene sequence for the selected ORF.

## Results

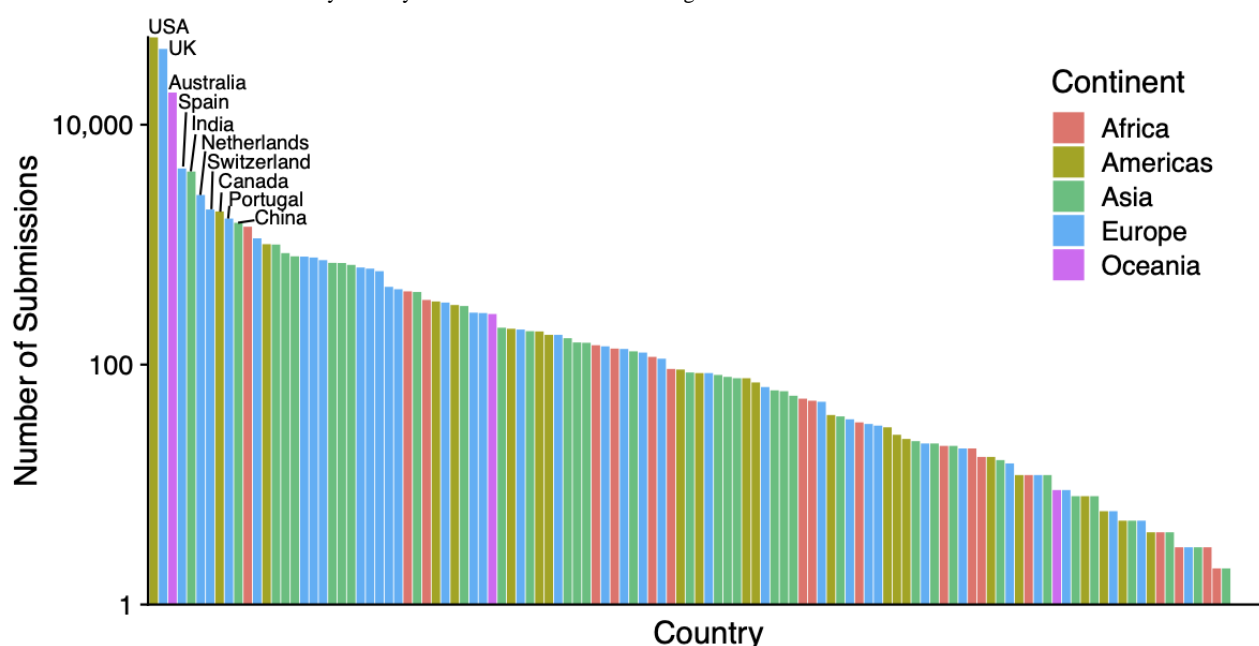
### Statistics on Collected Sequences

The number of SARS-CoV-2 sequences submitted to GISAID, NCBI, ENA, and CNGB has increased over time. Figure 1 shows the cumulative number of sequences by collection date. Sequences from Asia increased steadily since January, whereas sequences from other continents initially grew slowly but saw a dramatic increase in pace in March. The submission volume reflects the global spread of the virus, starting with the initial outbreak within China in January and followed by the subsequent global outbreak in March. Note that although the number of submissions correlates with the number of cases overall, the number of submissions does not reflect the actual case numbers (Figure 2). The ten countries with the highest number of submissions include five European countries (United Kingdom, Spain, Portugal, the Netherlands, and Switzerland), two Asian countries (India and China), two North American countries (United States and Canada), and one Oceanian country (Australia).

**Figure 1.** The cumulative number of sequences hosted by public databases has increased over time. Sequences from Asia increased steadily since January, whereas sequences from other continents saw a dramatic increase in March. Each data point represents a week. Note that the x-axis shows collection dates, which can precede submission dates by several weeks (eg, a sequence collected in July may be submitted in August). Therefore, lines that do not reach August indicate that sequences recently collected have not been submitted yet.



**Figure 2.** The number of submissions by country. The ten countries with the highest numbers of submissions are marked.

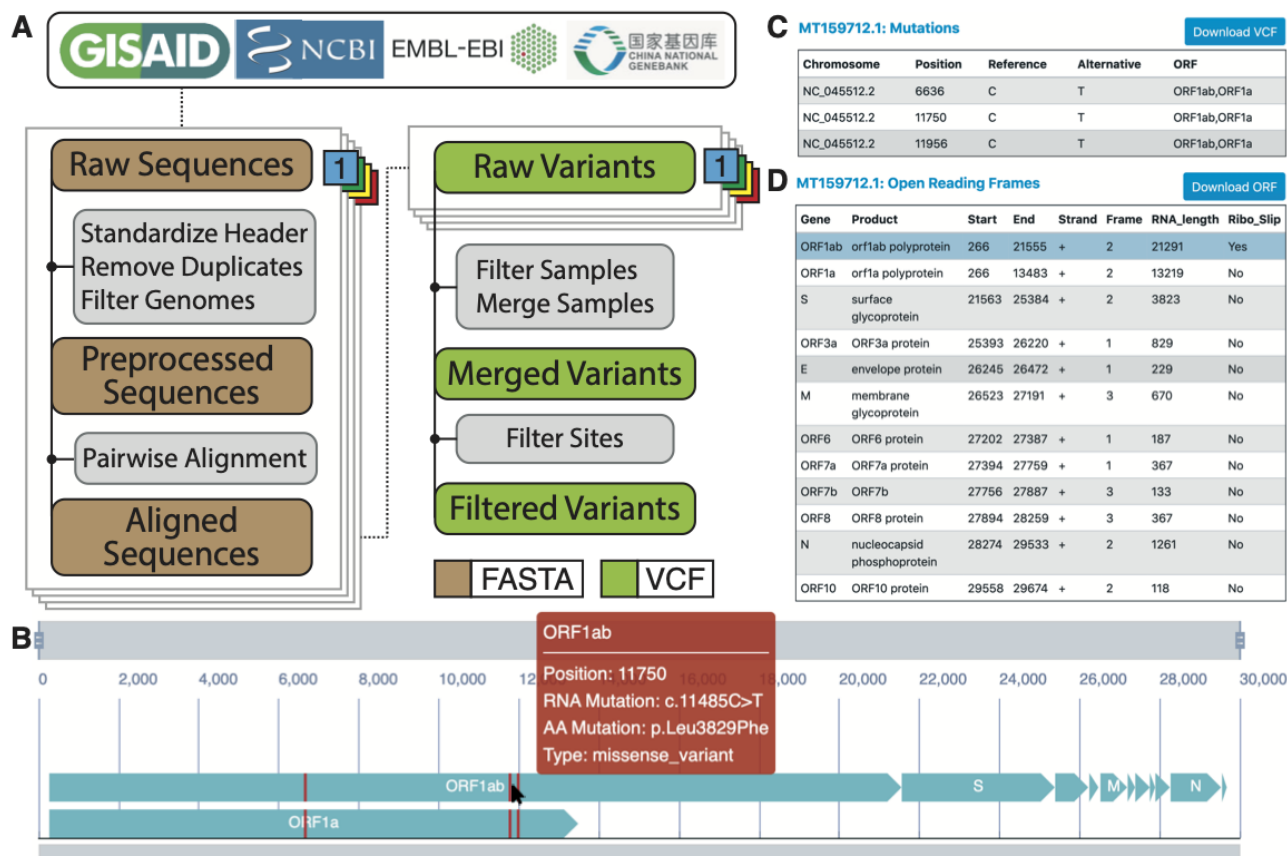


### Interactive Visualization With Custom Sequences

CoV-Seq provides an intuitive web interface (Figure 3A) for analyzing and visualizing SARS-CoV-2 variants and ORFs. Upon receiving custom sequences, CoV-Seq identifies ORF boundaries and genetic variants and displays them alongside a full-length genome (Figure 3B). Users interact with the genome by dragging the zoom bar to adjust the magnification and the position bar to pan along the genome. When the genome window contains less than 150 nucleotides, letters will appear to indicate both the nucleotide bases and amino acid residues. Hovering

over ORF bodies or variants will trigger pop-up windows for relevant information (Figure 3B, red box). Clicking on an ORF brings up three tables. The first table shows all variants and their positions, alleles, and ORFs in which they belong (Figure 3C). A second table shows ORF annotations obtained by aligning input sequences against the reference sequence and transferring annotations from Genbank (Figure 3D). A third table shows nucleotide and protein sequences for the selected ORF (not shown). All tables can be downloaded for further analysis.

**Figure 3.** The CoV-Seq pipeline and web interface. (A) Genomic sequences are collected from GISAID, NCBI, ENA, and CNGB. We remove incomplete genomes (length <25,000 nucleotides) and duplicate genomes before alignment with MAFFT against the reference genome NC\_045512.2. We use a custom Python script to generate raw variant calls and remove samples with too many mutations, indicative of sequencing error. After merging VCFs, we remove multiallelic sites and variants with the poly-A tail for a filtered set of variants. (B) The interactive genome visualizer shows ORFs (turquoise) and mutations (red). Users can zoom with the top bar and pan with the bottom bar. Hovering over ORF bodies and mutations will trigger pop-up windows for relevant information. (C) The mutation table shows positions, alleles, and intersecting ORFs. (D) The ORF table shows predicted ORF boundaries and supporting information. CNGB: China National GeneBank; ENA: European Nucleotide Archive; GISAID: Global Initiative on Sharing Avian Influenza Data; MAFFT: Multiple Alignment using Fast Fourier Transform; NCBI: National Center for Biotechnology Information; ORF: open reading frame; VCF: variant call format.



## CoV-Seq Command-Line Interface

Due to the rapid accumulation of SARS-CoV-2 genomic sequences, a point-and-click web interface is time-consuming for large collections. Therefore, we provide a command-line interface (CLI) for high-throughput processing of sequence batches [16]. The CLI allows the user to identify variants and ORF boundaries from multiple sequences with a single command. Further, the CLI allows the user to identify variants from multiple sequences from a FASTA file with one command.

## Downloading Analysis-Ready Data

To facilitate downstream analysis with publicly available data, we aggregated SARS-CoV-2 genomic sequences from GISAID, NCBI, ENA, and CNGB; we also identified and annotated genetic variants. In addition, we aggregated metadata with key information such as the location and collection date for each sequence (see Methods). All aggregated information can be downloaded from the CoV-Seq web server [17]. Based on this set of information, we provide statistics on the geographical and chronological distributions of sequence submissions and encourage further analysis by other scientists.

## Discussion

### Principal Results

In this paper, we have described CoV-Seq, a web server that enables the rapid analysis of SARS-CoV-2 genomic sequences. CoV-Seq consists of several components. The interactive visualization module accepts custom sequences as input and displays the genetic variants and ORF boundaries on an interactive genome browser. For batch processing needs, CoV-Seq provides a CLI interface that processes many sequences at once. To encourage downstream analysis with publicly available data, CoV-Seq provides downloadable analysis results (updated weekly) with sequence metadata and genetic mutations.

### Limitations

CoV-Seq is currently limited to SARS-CoV-2 sequences. The web server does not allow custom reference sequences other than SARS-CoV-2. We chose to focus on this virus because it has constituted the majority of processing requests during the COVID-19 pandemic. We plan to provide additional functionality to accept custom reference sequences in a future release.



## Comparison With Prior Work

The existing software packages VAPiD [11] and Viral Genome ORF Reader (VIGOR) [18] focus on gene annotations. To our knowledge, a software package that identifies, annotates, and visualizes genetic variants of SARS-CoV-2 has not previously been created.

## Conclusions

We developed the CoV-Seq web server for fast and easy analysis of SARS-CoV-2 sequences. We hope CoV-Seq will help improve our understanding of the genetics of COVID-19. In the future, we plan to expand the scope of CoV-Seq to include other viruses.

## Acknowledgments

We thank the Chinese Center for Disease Control and Prevention for codeveloping the idea behind this study and Xing Li for helpful discussions. We have no funding source to disclose.

## Authors' Contributions

BL and KL designed the web server. BL, KL, HZ, and LZ performed analyses. BL and LH wrote the paper.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Supplementary figures.

[PDF File (Adobe PDF File), 63 KB - [jmir\\_v22i10e22299\\_app1.pdf](#)]

### Multimedia Appendix 2

Acknowledgment table for GISAID contributors.

[XLSX File (Microsoft Excel File), 1882 KB - [jmir\\_v22i10e22299\\_app2.xlsx](#)]

## References

- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020 May;20(5):533-534 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30120-1](#)] [Medline: [32087114](#)]
- Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018 Dec 01;34(23):4121-4123 [FREE Full text] [doi: [10.1093/bioinformatics/bty407](#)] [Medline: [29790939](#)]
- Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 2017 Jan;1(1):33-46 [FREE Full text] [doi: [10.1002/gch2.1018](#)] [Medline: [31565258](#)]
- Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. *Nucleic Acids Res* 2015 Jan;43(Database issue):D571-D577 [FREE Full text] [doi: [10.1093/nar/gku1207](#)] [Medline: [25428358](#)]
- Kanz C, Aldebert P, Althorpe N, Baker W, Baldwin A, Bates K, et al. The EMBL Nucleotide Sequence Database. *Nucleic Acids Res* 2005 Jan 01;33(Database issue):D29-D33 [FREE Full text] [doi: [10.1093/nar/gki098](#)] [Medline: [15608199](#)]
- CoV-Seq web server. URL: <http://covseq.baidu.com/> [accessed 2020-09-24]
- Muthukadan B. Selenium with Python. URL: <https://selenium-python.readthedocs.io/> [accessed 2020-09-23]
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013 Apr;30(4):772-780. [doi: [10.1093/molbev/mst010](#)] [Medline: [23329690](#)]
- Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011 Nov 01;27(21):2987-2993 [FREE Full text] [doi: [10.1093/bioinformatics/btr509](#)] [Medline: [21903627](#)]
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6(2):80-92 [FREE Full text] [doi: [10.4161/fly.19695](#)] [Medline: [22728672](#)]
- Shean RC, Makhous N, Stoddard GD, Lin MJ, Greninger AL. VAPiD: a lightweight cross-platform viral annotation pipeline and identification tool to facilitate virus genome submissions to NCBI GenBank. *BMC Bioinformatics* 2019 Jan 23;20(1):48 [FREE Full text] [doi: [10.1186/s12859-019-2606-y](#)] [Medline: [30674273](#)]
- Mönnich A, Ronacher A, Lord D, Unterwaditzer M. Flask. URL: <https://flask.palletsprojects.com/en/1.1.x/> [accessed 2020-09-23]
- Mönnich A, Ronacher A, Lord D, Unterwaditzer M. Jinja. URL: <https://palletsprojects.com/p/jinja/> [accessed 2020-09-23]
- Mönnich A, Ronacher A, Lord D, Unterwaditzer M. Werkzeug. URL: <https://werkzeug.palletsprojects.com/en/1.0.x/> [accessed 2020-09-23]



15. Echarts. URL: <https://echarts.apache.org/en/index.html> [accessed 2020-09-23]
16. CoV-Seq GitHub. URL: <https://github.com/boxiangliu/covseq> [accessed 2020-09-24]
17. CoV-Seq aggregated data. URL: <http://covseq.baidu.com/browse> [accessed 2020-09-24]
18. Wang S, Sundaram JP, Spiro D. VIGOR, an annotation program for small viral genomes. BMC Bioinformatics 2010 Sep 07;11:451 [FREE Full text] [doi: [10.1186/1471-2105-11-451](https://doi.org/10.1186/1471-2105-11-451)] [Medline: [20822531](https://pubmed.ncbi.nlm.nih.gov/20822531/)]

## Abbreviations

**CLI:** command-line interface  
**GISAID:** Global Initiative on Sharing Avian Influenza Data  
**NCBI:** National Center for Biotechnology Information  
**ENA:** European Nucleotide Archive  
**CNGB:** China National GeneBank  
**ORF:** open reading frame  
**VCF:** variant call format

*Edited by G Eysenbach; submitted 08.07.20; peer-reviewed by V Duvvuri, M Pradhan; comments to author 17.08.20; revised version received 19.08.20; accepted 03.09.20; published 02.10.20.*

*Please cite as:*

Liu B, Liu K, Zhang H, Zhang L, Bian Y, Huang L

CoV-Seq, a New Tool for SARS-CoV-2 Genome Analysis and Visualization: Development and Usability Study

J Med Internet Res 2020;22(10):e22299

URL: <https://www.jmir.org/2020/10/e22299>

doi: [10.2196/22299](https://doi.org/10.2196/22299)

PMID: [32931441](https://pubmed.ncbi.nlm.nih.gov/32931441/)

©Boxiang Liu, Kaibo Liu, He Zhang, Liang Zhang, Yuchen Bian, Liang Huang. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 02.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Characteristics and Symptoms of App Users Seeking COVID-19–Related Digital Health Information and Remote Services: Retrospective Cohort Study

Amichai Perlman<sup>1\*</sup>, PharmD, PhD; Alina Vodonos Zilberg<sup>1\*</sup>, PhD; Peter Bak<sup>1</sup>, PhD; Michael Dreyfuss<sup>1</sup>, MD, PhD; Maya Leventer-Roberts<sup>1,2</sup>, MD, MPH; Yael Vurembrand<sup>1,3</sup>, MD; Howard E Jeffries<sup>1,4</sup>, MD, MPH; Eyal Fisher<sup>1,5</sup>, MSc; Yael Steurman<sup>1</sup>, PhD; Yinat Namir<sup>1</sup>, MSc; Yaara Goldschmidt<sup>1</sup>, PhD; Daniel Souroujon<sup>1</sup>, MD

<sup>1</sup>K Health Inc, New York, NY, United States

<sup>2</sup>Icahn School of Medicine at Mount Sinai, New York, NY, United States

<sup>3</sup>Center for Corona Treatment, Maccabi Health Services, Tel Aviv, Israel

<sup>4</sup>Department of Pediatrics, Seattle Children's Hospital, University of Washington School of Medicine, Seattle, WA, United States

<sup>5</sup>Li Ka Shing Centre, Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom

\*these authors contributed equally

**Corresponding Author:**

Daniel Souroujon, MD

K Health Inc

298 Fifth Avenue, 7th Floor

New York, NY, 10001

United States

Phone: 1 616 304 4654

Email: [daniel@khealth.ai](mailto:daniel@khealth.ai)

## Abstract

**Background:** Patient-facing digital health tools have been promoted to help patients manage concerns related to COVID-19 and to enable remote care and self-care during the COVID-19 pandemic. It has also been suggested that these tools can help further our understanding of the clinical characteristics of this new disease. However, there is limited information on the characteristics and use patterns of these tools in practice.

**Objective:** The aims of this study are to describe the characteristics of people who use digital health tools to address COVID-19–related concerns; explore their self-reported symptoms and characterize the association of these symptoms with COVID-19; and characterize the recommendations provided by digital health tools.

**Methods:** This study used data from three digital health tools on the K Health app: a protocol-based COVID-19 self-assessment, an artificial intelligence (AI)–driven symptom checker, and communication with remote physicians. Deidentified data were extracted on the demographic and clinical characteristics of adults seeking COVID-19–related health information between April 8 and June 20, 2020. Analyses included exploring features associated with COVID-19 positivity and features associated with the choice to communicate with a remote physician.

**Results:** During the period assessed, 71,619 individuals completed the COVID-19 self-assessment, 41,425 also used the AI-driven symptom checker, and 2523 consulted with remote physicians. Individuals who used the COVID-19 self-assessment were predominantly female (51,845/71,619, 72.4%), with a mean age of 34.5 years (SD 13.9). Testing for COVID-19 was reported by 2901 users, of whom 433 (14.9%) reported testing positive. Users who tested positive for COVID-19 were more likely to have reported loss of smell or taste (relative rate [RR] 6.66, 95% CI 5.53–7.94) and other established COVID-19 symptoms as well as ocular symptoms. Users communicating with a remote physician were more likely to have been recommended by the self-assessment to undergo immediate medical evaluation due to the presence of severe symptoms (RR 1.19, 95% CI 1.02–1.32). Most consultations with remote physicians (1940/2523, 76.9%) were resolved without need for referral to an in-person visit or to the emergency department.

**Conclusions:** Our results suggest that digital health tools can help support remote care and self-management of COVID-19 and that self-reported symptoms from digital interactions can extend our understanding of the symptoms associated with COVID-19.

**KEYWORDS**

digital health; remote care; symptom checker; telemedicine; COVID-19; symptom; cohort study; self-reported; online tool

## Introduction

Contemporary health care systems have limited capacity for managing epidemics; they are structured on the model of in-person interactions between patients and clinicians, which results in the congregation of patients in emergency departments and waiting areas during crises [1]. During the COVID-19 pandemic, it has been suggested that digital technologies can be used to deliver rapid urgent assessments, aid management of nonurgent conditions, and reduce the risk of iatrogenic COVID-19 exposure [1-3]. Indeed, several medical associations have encouraged the use of these tools during the COVID-19 epidemic, including the American Medical Association and the American Academy of Family Physicians [4,5]. Moreover, it has been suggested that these technologies can provide novel insights regarding the epidemiology and clinical characteristics of this new disease [6], especially in early stages and in community settings, as most published studies focus on hospitalized patients [7].

Digital tools have been promoted and deployed during previous infectious disease outbreaks, such as severe acute respiratory syndrome (SARS) and Ebola virus [8,9]. These tools focused primarily on surveillance, contact tracing, case management, and management of laboratory results [10]. In recent years, a wide range of patient-facing digital health modalities have emerged. These include protocol-based triage tools, artificial intelligence (AI)-driven tools for diagnosis and self-assessment, use of at-home remote monitoring devices, and virtual consultations by remote physicians. These tools are rapidly being deployed to support the medical management of the current epidemic [2]. However, despite the rapid growth in patient-facing digital and AI-driven technologies, other than reports of user satisfaction [11-13], there are limited data on the usage patterns and characteristics of these technologies in the management of COVID-19 and on their utility for advancing clinical research of COVID-19. Examining the characteristics and symptoms of digital technology users can help identify novel features associated with positivity or severity of COVID-19 and advance our understanding of the disease. Analysis of use patterns of digital tools can provide insight regarding their utility in the management of COVID-19 and can also highlight opportunities for further development.

During the pandemic, K Health, a novel AI-driven digital health platform deployed in the United States [14], offered three tools: a protocol-based COVID-19 self-assessment tool, an AI-driven symptom checker, and text-based telemedicine visits. This study aims to further our understanding of the use of digital health technologies in COVID-19 management and research. Specifically, the study describes the demographic and clinical characteristics of people using digital health tools for COVID-19-related concerns, explores self-reported symptom patterns among digital health users with COVID-19 compared to users who do not have the disease, and characterizes the

information and management recommendations provided by different digital health tools.

## Methods

### Population and Settings

K Health is a novel AI-driven digital health app; at the time this paper was written, it was available for download in the United States, Mexico, Indonesia, and Israel [14]. The app was built using a data-driven approach in collaboration with Maccabi Health Services (MHS). MHS is the second largest health maintenance organization in Israel, with over 2 million members [15,16]. Data from the MHS health records were used to develop an AI-driven symptom-checker, whose methodology has been described elsewhere [14]. The K Health app was launched in the United States in 2018 and has been used by over 4 million adults [14]. The app is free to download, and the symptom checker is available for free to adults over 18 years of age; thus, the app provides the public with reliable and personalized information on diagnoses related to their symptoms and medical history.

In June 2019, K Health began offering a service that enables users to consult directly with a remote physician. The platform provides the users with the option to receive a diagnosis, prescriptions, lab referrals, a referral to remote management of mental health concerns, or a referral to primary or emergency care. Consulting with a remote physician involves a fee or enrollment in a relevant health plan.

Early in April 2020, a protocol-based COVID-19 self-assessment was added to the app to provide users with up-to-date guidance for managing suspected cases of COVID-19. The self-assessment was developed by a team of board-certified physicians and was based on guidance issued by the World Health Organization and the US Centers for Disease Control and Prevention [17,18]. The user was asked to report the presence and severity of COVID-19-related symptoms as well as the presence of concomitant conditions known to increase risk of severe disease. Following the self-assessment, users were provided with one of four recommended actions based on their risk profile: social distancing, quarantine, isolation, or seeking immediate medical evaluation. Users were also informed if they were at increased risk for COVID-19 complications, and users with risk factors and symptoms were encouraged to consult a physician. For further details, including the full questionnaire, see [Multimedia Appendix 1](#).

### Study Design

This study describes use of three digital tools, namely a protocol-based COVID-19 self-assessment tool, an AI-driven symptom checker, and communication with a remote physician, by adults (>18 years of age) seeking COVID-19-related health information and services during the 10-week period between the launch date of April 8 to the date of data extraction on June

20, 2020. Seeking COVID-19–related health information was defined as use of the COVID-19 self-assessment tool. This self-assessment tool can be activated by the user, and it is also activated automatically when users begin a symptom check with a complaint of cough, dyspnea, or fever. The study did not include repeated self-assessments.

After users perform the COVID-19 self-assessment, they are prompted to use the AI-driven symptom checker to receive additional information about other conditions that may be related to their symptoms. After receiving the results and information regarding these conditions, the users can choose to communicate with a remote physician for definitive medical management.

This study analyzed deidentified data. The protocol of the study was reviewed by the Western Institutional Review Board and qualified for exemption in accordance with 45 CFR § 46.104(d)(4). K Health is compliant with the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR). Encrypted transportation and storage were used at all stages of data management.

## Variables

Deidentified data on the characteristics of the digital interactions in the app were collected, including user-inputted demographic and clinical characteristics; COVID-19 exposure reporting; reports of COVID-19 testing, including results; and symptoms. We also collected the outputs provided by the three digital modalities: risk categories according to the protocol-based COVID-19 self-assessment; the most common conditions for similar people and symptoms (“potential diagnoses”) provided by the AI-driven symptom checker; and diagnoses, management, and disposition by remote physicians.

User characteristics included age, sex, risk factors, and comorbidities. Comorbidities included hypertension, smoking, obesity, diabetes (type 1 and type 2), cardiovascular disease, chronic lung disease (asthma, chronic obstructive pulmonary disorder, and interstitial lung disease), chronic renal and liver disease, cancer, and immune suppression.

Symptoms were obtained from two types of interactions on the platform: first, as part of the structured protocol-based COVID-19 risk assessment, and second, as part of the dynamic AI-driven health dialog.

The COVID-19 self-assessment included questions regarding symptoms known to be associated with COVID-19 (eg, cough, dyspnea, and fever) as well as symptoms hypothesized to be related to COVID-19 (eg, eye symptoms), which were inquired about for further research and development of the COVID-19 protocol. The symptom checker dialog allows the user to spontaneously discuss a wider array of 331 symptoms; it also queries the user regarding additional symptoms and symptom-specific attributes, such as the severity, timing, duration, and quality of a symptom (eg, dry or productive cough), as previously described [6].

## Analyses

The analyses assessed the user characteristics, reported symptoms, and output of the aforementioned digital services, namely the self-assessment protocol, AI- and data-driven

symptom checker, and remote physician consultation. This assessment was performed using descriptive statistics and bivariate analyses.

Descriptive univariate summary statistics were developed to assess the user characteristics, reported symptoms, diagnoses, and disposition. Symptoms were assessed for both the COVID-19 self-assessment and the symptom checker. Bivariate analyses were conducted for between-group comparisons of characteristics and symptoms. These included two primary analyses.

The first analysis explored the potential utility of self-reported data from digital health tools to identify symptoms associated with COVID-19. This analysis compared the self-reported symptoms by digital health users who reported testing positive for COVID-19 with those of users who reported testing negative. To ensure that the symptoms of cough, fever, and dyspnea were optimally captured, the presence of these symptoms was collected from both the self-assessment and the symptom checker dialog. More details are provided in [Multimedia Appendix 1](#).

The second analysis aimed to identify predictors related to the choice to consult with a remote physician. This analysis compared data provided by individuals who opted to consult with a remote physician with those provided by the entire cohort. Specifically, this comparison aimed to evaluate whether consultation with a physician was related to risk of COVID-19, symptom severity, or comorbidity.

Appropriate statistical tests were used to assess the significance of between-group differences: chi-square tests for differences in proportions and *t* tests for differences in continuous measures. Test assumptions were assessed analytically and graphically and were judged to be adequately met. We considered *P* values of .05 or less to be significant and did not correct for multiple comparisons. Analyses were conducted using Python version 3.6.9.

## Results

During the assessed timeframe, 71,619 individuals completed the COVID-19 self-assessment. The self-assessment included questions on COVID-19 exposure, testing, comorbid conditions, and the presence of COVID-19–related symptoms. The self-assessment output provided users with protocol-based information and recommendations on their COVID-19 risk. Of the 71,619 individuals who completed the self-assessment, 41,425 (57.8%) also completed a more detailed AI-driven symptom checker. The symptom checker provides the option to evaluate a wide range of 331 symptoms and to receive information about non-COVID-19 conditions that may be related to these symptoms. Following use of the symptom checker, a subset of 2523 users proceeded to consult with remote physicians, who provided guidance on the users’ disposition and health management ([Figure 1](#)).

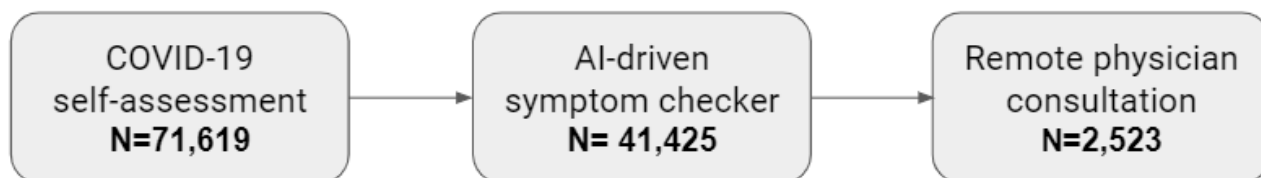
Individuals using the COVID-19 self-assessment were predominantly female (51,845/71,619, 72.4%), with a mean age of 34.5 years (SD 13.9). The most commonly reported comorbidities were chronic lung disease, primarily asthma

(16,080/71,619, 22.5%), and hypertension (13,952/71,619, 19.5%) (Table 1).

The self-assessment results provided users with information and recommendations related to their risk of COVID-19, with

the most common recommendations involving social distancing (20,984/71,619, 29.3%) and isolation (23,706/71,619, 33.1%) (Table 2).

**Figure 1.** Flowchart describing the cohort of US individuals who used the COVID-19 self-assessment between April 8 and June 20, 2020. AI: artificial intelligence.



**Table 1.** Characteristics of the cohort of COVID-19 self-assessment users (N=71,619).

Characteristic	Value
<b>Demographics</b>	
Age, mean (SD)	34.5 (13.9)
Female, n (%)	51,845 (72.4)
<b>Chronic conditions, n (%)</b>	
Hypertension	13,952 (19.5)
Morbid obesity	7676 (10.7)
Smoking	13,505 (18.9)
Stroke	897 (1.3)
Cancer or immunosuppression	1868 (2.6)
Chronic kidney disease	1313 (1.8)
Chronic lung disease	16,080 (22.5)
Cardiovascular disease	3835 (5.4)
Diabetes	5958 (8.3)
<b>Tested for COVID-19 (n=2901), n (%)</b>	
Tested positive	433 (14.9)
Tested negative	2468 (85.0)



**Table 2.** Outputs provided by the digital health tools (N=71,619), n (%).

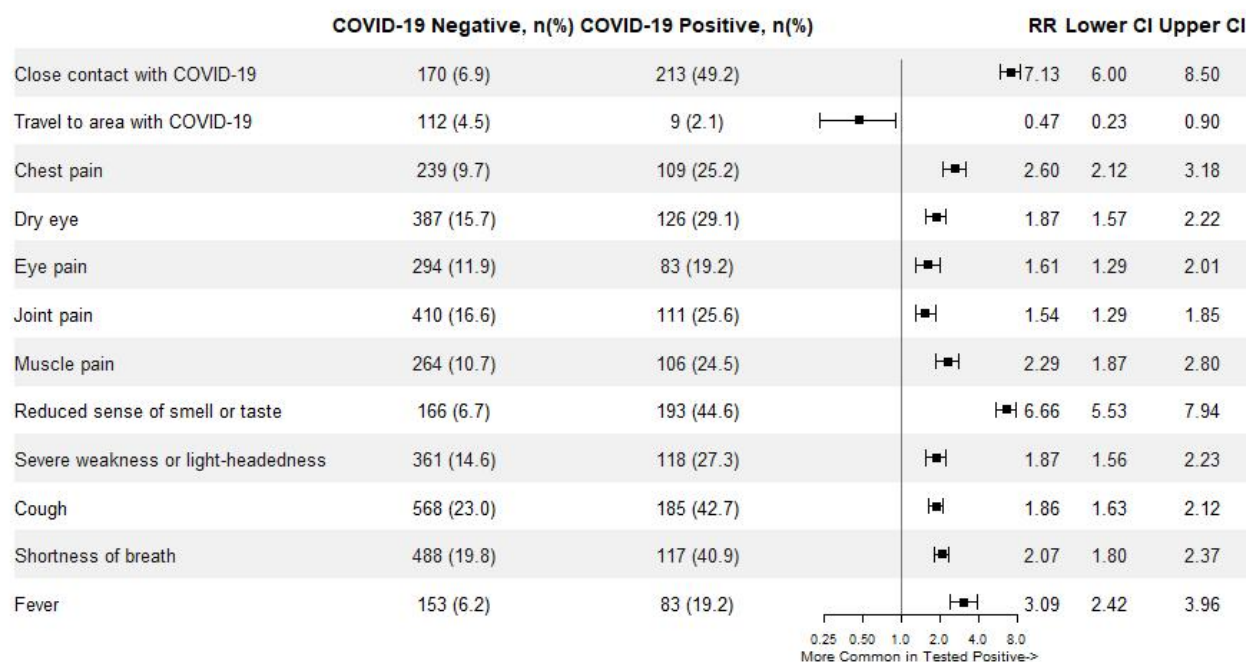
Output	Value
<b>COVID-19 assessment recommendation</b>	
Practice social distancing	20,984 (29.3)
Isolate yourself	23,706 (33.1)
Quarantine yourself	7735 (10.8)
Seek immediate evaluation	19,194 (26.8)
<b>Most common symptom checker potential diagnoses (n=41,425)</b>	
Upper respiratory infection	15,232 (36.7)
Anxiety disorder	2457 (5.9)
Gastroesophageal reflux disease	2075 (5.0)
Dehydration	1512 (3.6)
Tension-type headache	1152 (2.8)
Allergic rhinitis	957 (2.3)
Depressive mood disorder	833 (2.0)
Pulmonary embolism	783 (1.9)
Pneumonia	721 (1.7)
Acute food poisoning	708 (1.7)
<b>Remote physician management and disposition<sup>a</sup> (n=2523)</b>	
Medication or laboratory tests	1072 (42.5)
Information and reassurance	431 (17.1)
Other	363 (14.4)
Referral to primary care or a specialist	325 (12.9)
Referral to the emergency department	257 (10.2)
Referral to COVID-19 testing	169 (6.7)
Referral to a remote behavioral health service	73 (2.9)

<sup>a</sup>Dispositions are not mutually exclusive.

Testing for COVID-19 was reported by 2901 users, of whom 433 (14.9%) reported testing positive. Users who were tested for COVID-19 were predominantly female (2022/2901, 69.7%), with a mean age of 38.1 years (SD 13.8). Of users who were tested, those who reported testing positive were of similar age (mean 37.5 years, SD 13.8, vs mean 38.2 years, SD 13.8,  $P=.32$ ), less frequently female (62.4% vs 69.9%), and much more likely to have reported a close contact with a known case of COVID-19 (relative rate [RR] 7.13, 95% CI 6.00-8.49), compared to those who tested negative. The symptoms reported by users as part

of the COVID-19 self-assessment differed between those who reported positive and negative test results. Most notably, the relative prevalence of reporting loss of smell or taste was significantly higher among individuals who tested positive for COVID-19 (RR 6.65, 95% CI 5.53-7.94), as was the prevalence of fever (RR 2.58, 95% CI 2.04-3.26) and chest pain (RR 2.59, 95% CI 2.12-3.18). Cough, difficulty eating or drinking, dry eye, eye pain, feeling weak and lightheaded, and chest pain were also significantly associated with positive COVID-19 tests (Figure 2).

**Figure 2.** Forest plot presenting the RRs and 95% CIs of each item in the self-assessment questionnaire among users who reported testing positive for COVID-19 compared to those who reported testing negative. Muscle pain was not present in early versions of the self-assessment. Proportions were calculated for self-assessments where the symptom was reportable. RR: relative rate.



Individuals using the COVID-19 self-assessment were prompted to use the AI-driven symptom checker to receive additional information about other conditions potentially related to their symptoms. This tool enables users to input a wide range of symptoms that are not systematically collected as part of the COVID-19 self-assessment; thus, it recorded a wide range of additional symptoms that were not already included in the protocol-based self-assessment. Among these, the most common additional symptoms reported by individuals who tested positive for COVID-19 were headache, sore throat, fatigue, diarrhea, chest pain, nausea, nasal congestion, runny nose, and sweating (Supplementary Table 1, Multimedia Appendix 1). The AI-driven algorithm provided users with information about the most common conditions people with symptoms and personal characteristics similar to theirs were diagnosed with (“potential diagnoses”). The most common non-COVID-19 conditions classified as relevant for the 41,425 users were upper respiratory infection (15,232, 36.7%), anxiety disorder (2457, 5.9%),

gastroesophageal reflux disease (GERD) (2075, 5.0%), and dehydration (1512, 3.6%) (Table 2).

Among individuals initially evaluating their COVID-19 risk, a small subset (2523/41,425, 3.5%) chose to communicate with a remote physician. Users who chose to communicate with a remote physician were older (mean age 38.6 years, SD 13.2, vs mean age 34.3 years,  $P<.001$ ) and more likely to be male (37.9% vs 27.2%,  $P<.001$ ) than users who did not; however, they were less likely to have comorbidities (RR 0.88, 95% CI 0.82-0.95). These individuals were more likely to report severe symptoms such as severe chest pain (RR 1.13, 95% CI 1.03-1.25) and to be encouraged by the self-assessment to seek immediate evaluation (RR 1.19, 95% CI 1.02-1.32). Additionally, features and symptoms supporting higher suspicion of COVID-19 were more common among users who communicated with a remote physician, including exposure to a close contact with COVID-19 (RR 1.21, 95% CI 1.1-1.35), fever (RR 1.56, 95% CI 1.43-1.72), shortness of breath (RR 1.25, 95% CI 1.16-1.33), and loss of smell and taste (RR 1.22, 95% CI 1.09-1.37) (Table 3).

**Table 3.** User characteristics and self-assessment protocol disposition associated with consulting a remote physician (N=71,619).

Item or disposition	Did not consult a remote physician (n=69,096), n (%)	Consulted a remote physician (n=2523), n (%)	RR <sup>a</sup> (95% CI)
<b>Self-assessment items</b>			
Close contact	7431 (10.8)	331 (13.1)	1.21 (1.10-1.35)
Travel	1908 (2.8)	59 (2.3)	0.82 (0.66-1.09)
Eye pain	6257 (9.1)	218 (8.6)	0.95 (0.84-1.09)
Dry eye	7242 (10.5)	269 (10.6)	1.02 (0.93-1.17)
Joint pain	61,741 (89.4)	2271 (90.0)	1.01 (0.83-1.06)
Loss of smell or taste	6051 (8.8)	270 (10.7)	1.22 (1.09-1.37)
Feeling weak/light-headed	11,101 (16.1)	462 (18.3)	1.14 (1.05-1.24)
Chest pain	8279 (12.0)	343 (13.6)	1.13 (1.03-1.26)
Muscle pain	7643 (11.1)	289 (11.5)	1.04 (0.93-1.16)
Cough <sup>b</sup>	21,832 (31.60)	859 (34.05)	1.07 (1.02-1.14)
Shortness of breath <sup>b</sup>	14,424 (20.8)	656 (26.0)	1.25 (1.16-1.33)
Fever <sup>b</sup>	7147 (10.34)	409 (16.21)	1.57 (1.43-1.72)
Testing for COVID-19	660 (1.0)	34 (1.3)	1.30 (1.00-2.00)
<b>Disposition of self-assessment</b>			
Practice social distancing	17,827 (25.8)	777 (30.8)	1.19 (1.15-1.13)
Isolate yourself	24,118 (34.9)	781 (30.9)	0.89 (0.67-0.96)
Quarantine yourself	9016 (13.1)	176 (6.9)	0.53 (0.34-0.65)
Seek immediate evaluation	18,135 (26.3)	789 (31.3)	1.19 (1.02-1.32)

<sup>a</sup>RR: relative rate.<sup>b</sup>Users could report these symptoms at the end of the self-assessment or as part of the symptom-checker dialog. See [Multimedia Appendix 1](#) for details.

The remote physicians provided a wide range of evaluation and counseling services, including assessing severe cases and referring them to the emergency department (257/2523, 10.2%) or to ambulatory care in the community (325/2523, 12.9%), advising on and prescribing medications (1072/2523, 42.5%), providing remote behavioral health services to individuals with mild to moderate anxiety or depression (73/2523, 2.9%), providing additional information and reassurance (431/2523, 17.1%), and referring users to COVID-19 testing (169/2523, 6.7%) ([Table 2](#)).

## Discussion

### Principal Findings and Interpretation

This study describes the characteristics and management of >71,000 individuals who used digital tools to seek health information and services related to COVID-19. These individuals were relatively young and predominantly female. Users received information regarding COVID-19 risk and management, received AI-driven information about other relevant diagnoses, and consulted with remote physicians. Users who chose to communicate with a remote physician were more likely to have been classified as requiring immediate medical evaluation by the COVID-19 self-assessment. Correspondingly, these users were also older, more likely to report severe

symptoms, and more likely to report characteristics cognate with risk of COVID-19 (such as known exposure to COVID-19 and loss of sense of smell or taste). The majority of consultations with a remote physician (1940/2523, 76.9%) were resolved without need for referral to an in-person health visit or to the emergency department.

Taken together, the differential communication with remote physicians according to symptom severity and the high resolution rates without need for referral to in-person visits suggest that the digital tools provided information and advice that assisted users in making health decisions. These tools may therefore reduce the burden on health care systems during times when resources are limited and may help minimize unnecessary physical interactions that could lead to iatrogenic COVID-19 exposure. However, research on health care use and health outcomes following digital health tool use is needed to conclusively demonstrate this potential.

In addition, individuals who chose to communicate with a remote physician tended to have lower rates of comorbidity. This suggests that these individuals were more likely to have an established relationship with a health care provider whom they consulted regularly and were more likely to require an in-person evaluation and more complex care when presenting with significant symptoms.

This study also highlighted differences in self-reported symptoms between users who reported testing positive for COVID-19 and those who reported testing negative. The symptom most strongly associated with positive testing for COVID-19 was loss of sense of taste or smell (RR 6.66, 95% CI 5.53-7.93). Surprisingly, travel to an area with COVID-19 appeared to be associated with a lower rate of positive testing (RR 0.47, 95% CI 0.23-0.90); however, this estimate is based on a relatively small number of subjects and should therefore be interpreted with caution. As the decision to test or be tested is driven by the presence of risk factors for COVID-19, travel to an area with COVID-19 may appear protective because individuals who were tested due to this risk factor may be less likely to have other stronger predictors of positivity, such as exposure to an individual with COVID-19 or the presence of COVID-19 symptoms. Additional features associated with COVID-19 included other widely recognized COVID-19 symptoms, such as fever and cough. These results emphasize the potential utility of taste and smell as strong signals of COVID-19 positivity in the community setting, and they are similar to results reported in recent studies of self-reported symptoms among individuals who tested positive for COVID-19 [19,20].

Finally, in our study, we found that dry eye (RR 1.87, 95% CI 1.57-2.22) and eye pain (RR 1.61, 95% CI 1.29-2.01) were more common among individuals who reported testing positive for COVID-19. These symptoms were added to the self-assessment to explore the possible link between COVID-19 and eye symptoms as well as the utility of these symptoms in evaluating suspicion of COVID-19. Although some early hospital-based studies reported low rates of ocular symptoms in patients with COVID-19 [20], several other studies have suggested that ocular symptoms are common among individuals with COVID-19 [21-23]. Ocular symptoms have been documented in some cases as the first [24] and even only [25] symptomatic manifestation of COVID-19, and studies have documented positive reverse transcriptase-polymerase chain reaction COVID-19 test results from ocular secretions [23,26]. This study adds to the body of evidence suggesting that the manifestations of COVID-19 include ocular symptoms and that they may be more common symptoms of COVID-19 than generally recognized.

Seen from a wider perspective, these results demonstrate that self-reported symptoms on a digital app can replicate symptoms known to be associated with COVID-19, that they can help distinguish between individuals who test positive and negative for COVID-19, and that they may add to our understanding of symptoms associated with COVID-19. This exemplifies the potential of data generated from digital tools to improve our understanding of the clinical manifestations of COVID-19 and of patient-reported experiences in general.

### Comparison With Prior Work

The potential benefits of digital tools during the COVID-19 epidemic have been noted in multiple health policy commentaries [1-3,19,27]. This stance has been adopted by several leading medical associations, including the American Medical Association and the American Academy of Family Physicians [4,5]. However, there has been limited research to

date on the actual usage patterns and impact of these tools during the COVID-19 epidemic. The body of literature on digital health for COVID-19 primarily features perspectives and opinion pieces, guidance papers, and a few studies. These studies tend to be small, focus on a single digital tool, and/or report primarily on survey results of user satisfaction. Two larger studies have provided some additional insight. The first study reported the results of a satisfaction survey of 6194 people following wide-scale deployment of digital tools for COVID-19 education, self-assessment, and symptom monitoring in the Netherlands; high satisfaction rates were reported [11]. The second study reported on the treatment and high satisfaction rates of 4589 patients in China using web-based physician consultation for COVID-19 concerns [13].

The demographic characteristics of digital health users described in this study match previous reports on digital health users. While the use of digital health technologies among seniors has been reported to be increasing [28], users of digital health applications are still predominantly younger adults [29]. The current low level of adoption of digital health by older adults is unfortunate, as digital health technologies have the potential to improve communication and collaboration and promote healthy and independent ageing. Indeed, recent research suggests that digital solutions tailored for older adults improve health management [30]. Barriers to the adoption of digital health by older adults include visual impairment, limitations in dexterity, and lack of self-confidence when using technology. Several solutions are being developed to circumvent these barriers, including the development of voice-based applications and unobtrusive sensors and trackers [31].

### Strengths and Limitations

This study has a number of strengths. First, it uses a large sample of over 71,000 individuals to provide timely information on the use of a number of different digital health tools for managing COVID-19-related concerns. Second, the study reported on the differential use of these tools by these individuals as well as on both self-reported variables and physician-reported disposition and management. Third, the study was able to provide insight into self-reported symptoms of individuals who were tested for COVID-19; the results highlighted the strong link between COVID-19 and loss of sense of smell and taste and added to the body of evidence that ocular symptoms may be a more common feature of COVID-19 than is widely recognized.

The study has a number of limitations as well. The population of this study is not representative of the entire population, and the study uses data from digital tools developed by a single provider. However, the characteristics of the study population and the COVID-19 symptoms reported correspond with those in previous reports on digital health tools [11,20-22]. In addition, data on COVID-19 test status were based on self-reporting and may not accurately capture test results. Furthermore, the data are limited to individuals who used a tool for initial COVID-19 assessment. Tools enabling longitudinal logging and monitoring of symptoms can both provide additional utility to users and improve our understanding of disease progression and the time course of symptoms. Enabling users to provide unsolicited data on their experience may provide important insight as well, as

the self-assessment tool described focused on a predefined list of questions. Lastly, while we had data on the disposition of users who chose to consult with a remote physician, the disposition of individuals who did not consult with a remote physician following use of the automated tools is unknown. Research on health care use and health outcomes of digital health users is needed to conclusively demonstrate the utility of these services in assisting individuals with health decision-making and reducing the burden on the health system.

## Conclusion

This study describes the integration of three digital health tools for the direct management of COVID-19–related concerns. The results suggest that automated, data-driven digital solutions as well as remote care provided by a human physician can help provide health information and guidance during an epidemic. In addition, interactions across digital services can provide insight regarding the characteristics of new diseases. The integration of these tools can be an important resource for health care providers and policy makers.

## Acknowledgments

This study was funded by K Health Inc.

## Authors' Contributions

AP and AVZ contributed equally to the study. AVZ had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. AP, AVZ, MD, YV, EF, YS, YN, and DS conceived and designed the study. AP, AVZ, PB, MD, MLV, YV, HEJ, EF, YS, YN, YG, and DS acquired, analyzed, or interpreted the data. AP and AVZ drafted the manuscript. PB, MD, MLR, YV, HEJ, EF, YS, YS, YG, and DS critically revised the manuscript for important intellectual content. AVZ, PB, and YN performed the statistical analysis. YN, TG, and DS supervised the work.

## Conflicts of Interest

All authors are salaried employees or consultants at K Health Inc.

Multimedia Appendix 1

Supplementary information.

[DOCX File, 9 KB - [jmir\\_v22i10e23197\\_app1.docx](#)]

## References

1. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. *N Engl J Med* 2020 Jun 04;382(23):e82. [doi: [10.1056/nejmp2005835](#)]
2. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. *N Engl J Med* 2020 Apr 30;382(18):1679-1681. [doi: [10.1056/NEJMp2003539](#)] [Medline: [32160451](#)]
3. Schwamm LH, Estrada J, Erskine A, Licurse A. Virtual care: new models of caring for our patients and workforce. *Lancet Digit Health* 2020 Jun;2(6):e282-e285 [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30104-7](#)] [Medline: [32382724](#)]
4. AMA Telehealth quick guide. American Medical Association. URL: <https://www.ama-assn.org/practice-management/digital/ama-telehealth-quick-guide> [accessed 2020-07-29]
5. Using Telehealth to Care for Patients During the COVID-19 Pandemic. American Academy of Family Physicians. URL: <https://www.aafp.org/patient-care/emergency/2019-coronavirus/telehealth.html> [accessed 2020-07-29]
6. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: [10.1126/science.abc0473](#)] [Medline: [32371477](#)]
7. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, et al. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PLoS One* 2020 Jun 23;15(6):e0234765 [FREE Full text] [doi: [10.1371/journal.pone.0234765](#)] [Medline: [32574165](#)]
8. Bempong NE, Ruiz De Castañeda R, Schütte S, Bolon I, Keiser O, Escher G, et al. Precision Global Health - The case of Ebola: a scoping review. *J Glob Health* 2019 Jun;9(1):010404 [FREE Full text] [doi: [10.7189/jogh.09.010404](#)] [Medline: [30701068](#)]
9. Eysenbach G. SARS and population health technology. *J Med Internet Res* 2003 Jun 30;5(2):e14 [FREE Full text] [doi: [10.2196/jmir.5.2.e14](#)] [Medline: [12857670](#)]
10. Tom-Aba D, Nguku PM, Arinze CC, Krause G. Assessing the Concepts and Designs of 58 Mobile Apps for the Management of the 2014-2015 West Africa Ebola Outbreak: Systematic Review. *JMIR Public Health Surveill* 2018 Oct 29;4(4):e68 [FREE Full text] [doi: [10.2196/publichealth.9015](#)] [Medline: [30373727](#)]



11. Timmers T, Janssen L, Stohr J, Murk JL, Berrevoets MAH. Using eHealth to Support COVID-19 Education, Self-Assessment, and Symptom Monitoring in the Netherlands: Observational Study. *JMIR Mhealth Uhealth* 2020 Jun 23;8(6):e19822 [FREE Full text] [doi: [10.2196/19822](https://doi.org/10.2196/19822)] [Medline: [32516750](https://pubmed.ncbi.nlm.nih.gov/32516750/)]
12. Saleem JJ, Read JM, Loehr BM, Frisbee KL, Wilck NR, Murphy JJ, et al. Veterans' response to an automated text messaging protocol during the COVID-19 pandemic. *J Am Med Inform Assoc* 2020 Aug 01;27(8):1300-1305 [FREE Full text] [doi: [10.1093/jamia/ocaa122](https://doi.org/10.1093/jamia/ocaa122)] [Medline: [32470974](https://pubmed.ncbi.nlm.nih.gov/32470974/)]
13. Liu L, Gu J, Shao F, Liang X, Yue L, Cheng Q, et al. Application and Preliminary Outcomes of Remote Diagnosis and Treatment During the COVID-19 Outbreak: Retrospective Cohort Study. *JMIR mHealth uHealth* 2020 Jul 03;8(7):e19417 [FREE Full text] [doi: [10.2196/19417](https://doi.org/10.2196/19417)] [Medline: [32568722](https://pubmed.ncbi.nlm.nih.gov/32568722/)]
14. Koren G, Souroujon D, Shaul R, Bloch A, Leventhal A, Lockett J, et al. "A patient like me" – An algorithm-based program to inform patients on the likely conditions people with symptoms like theirs have. *Medicine* 2019;98(42):e17596. [doi: [10.1097/md.00000000000017596](https://doi.org/10.1097/md.00000000000017596)]
15. Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. *Nat Med* 2019 Jul;25(7):1143-1152 [FREE Full text] [doi: [10.1038/s41591-019-0503-6](https://doi.org/10.1038/s41591-019-0503-6)] [Medline: [31273328](https://pubmed.ncbi.nlm.nih.gov/31273328/)]
16. Eisenberg V, Weil C, Chodick G, Shalev V. Epidemiology of endometriosis: a large population-based database study from a healthcare provider with 2 million members. *BJOG* 2018 Jan;125(1):55-62 [FREE Full text] [doi: [10.1111/1471-0528.14711](https://doi.org/10.1111/1471-0528.14711)] [Medline: [28444957](https://pubmed.ncbi.nlm.nih.gov/28444957/)]
17. Coronavirus Disease (COVID-19). US Centers for Disease Control and Prevention. 2020 Feb 11. URL: <https://www.cdc.gov/coronavirus/2019-ncov/index.html> [accessed 2020-06-09]
18. World Health Organization. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance> [accessed 2020-06-09]
19. Menni C, Sudre CH, Steves CJ, Ourselin S, Spector TD. Quantifying additional COVID-19 symptoms will save lives. *Lancet* 2020 Jun;395(10241):e107-e108. [doi: [10.1016/s0140-6736\(20\)31281-2](https://doi.org/10.1016/s0140-6736(20)31281-2)]
20. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: [10.1056/NEJMoa2002032](https://doi.org/10.1056/NEJMoa2002032)] [Medline: [32109013](https://pubmed.ncbi.nlm.nih.gov/32109013/)]
21. Gangaputra SS, Patel SN. Ocular Symptoms among Nonhospitalized Patients Who Underwent COVID-19 Testing. *Ophthalmology* 2020 Oct;127(10):1425-1427 [FREE Full text] [doi: [10.1016/j.ophtha.2020.06.037](https://doi.org/10.1016/j.ophtha.2020.06.037)] [Medline: [32585259](https://pubmed.ncbi.nlm.nih.gov/32585259/)]
22. Wu P, Duan F, Luo C, Liu Q, Qu X, Liang L, et al. Characteristics of Ocular Findings of Patients With Coronavirus Disease 2019 (COVID-19) in Hubei Province, China. *JAMA Ophthalmol* 2020 May 01;138(5):575-578 [FREE Full text] [doi: [10.1001/jamaophthalmol.2020.1291](https://doi.org/10.1001/jamaophthalmol.2020.1291)] [Medline: [32232433](https://pubmed.ncbi.nlm.nih.gov/32232433/)]
23. Aiello F, Gallo Afflitto G, Mancino R, Li JO, Cesareo M, Giannini C, et al. Coronavirus disease 2019 (SARS-CoV-2) and colonization of ocular tissues and secretions: a systematic review. *Eye (Lond)* 2020 Jul;34(7):1206-1211 [FREE Full text] [doi: [10.1038/s41433-020-0926-9](https://doi.org/10.1038/s41433-020-0926-9)] [Medline: [32424327](https://pubmed.ncbi.nlm.nih.gov/32424327/)]
24. Daruich A, Martin D, Bremond-Gignac D. Unilateral conjunctivitis as first presentation of Coronavirus Disease 2019 (COVID-19): A telemedicine diagnosis. *J Fr Ophtalmol* 2020 May;43(5):e167-e168 [FREE Full text] [doi: [10.1016/j.jfo.2020.04.001](https://doi.org/10.1016/j.jfo.2020.04.001)] [Medline: [32327279](https://pubmed.ncbi.nlm.nih.gov/32327279/)]
25. Scalinci SZ, Trovato Battagliola E. Conjunctivitis can be the only presenting sign and symptom of COVID-19. *IDCases* 2020;20:e00774 [FREE Full text] [doi: [10.1016/j.idcr.2020.e00774](https://doi.org/10.1016/j.idcr.2020.e00774)] [Medline: [32373467](https://pubmed.ncbi.nlm.nih.gov/32373467/)]
26. Xia J, Tong J, Liu M, Shen Y, Guo D. Evaluation of coronavirus in tears and conjunctival secretions of patients with SARS-CoV-2 infection. *J Med Virol* 2020 Jun 12;92(6):589-594 [FREE Full text] [doi: [10.1002/jmv.25725](https://doi.org/10.1002/jmv.25725)] [Medline: [32100876](https://pubmed.ncbi.nlm.nih.gov/32100876/)]
27. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. *Lancet Digit Health* 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: [10.1016/s2589-7500\(20\)30142-4](https://doi.org/10.1016/s2589-7500(20)30142-4)]
28. Levine DM, Lipsitz SR, Linder JA. Trends in Seniors' Use of Digital Health Technology in the United States, 2011-2014. *JAMA* 2016 Aug 02;316(5):538-540. [doi: [10.1001/jama.2016.9124](https://doi.org/10.1001/jama.2016.9124)] [Medline: [27483069](https://pubmed.ncbi.nlm.nih.gov/27483069/)]
29. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. *BMJ Open* 2019 Aug 01;9(8):e027743. [doi: [10.1136/bmjopen-2018-027743](https://doi.org/10.1136/bmjopen-2018-027743)] [Medline: [31375610](https://pubmed.ncbi.nlm.nih.gov/31375610/)]
30. Richard E, Moll van Charante EP, Hoevenaer-Blom MP, Coley N, Barbera M, van der Groep A, et al. Healthy ageing through internet counselling in the elderly (HATICE): a multinational, randomised controlled trial. *Lancet Digit Health* 2019 Dec;1(8):e424-e434. [doi: [10.1016/s2589-7500\(19\)30153-0](https://doi.org/10.1016/s2589-7500(19)30153-0)]
31. Evangelista L, Steinhubl SR, Topol EJ. Digital health care for older adults. *Lancet* 2019 Apr;393(10180):1493. [doi: [10.1016/s0140-6736\(19\)30800-1](https://doi.org/10.1016/s0140-6736(19)30800-1)]

## Abbreviations

**AI:** artificial intelligence

**GDPR:** General Data Protection Regulation

**HIPAA:** Health Insurance Portability and Accountability Act

**MHS:** Maccabi Health Services

**RR:** relative rate

**SARS:** severe acute respiratory syndrome

*Edited by G Fagherazzi; submitted 04.08.20; peer-reviewed by H Mehdizadeh, KC Wong; comments to author 03.09.20; revised version received 13.09.20; accepted 19.09.20; published 20.10.20.*

*Please cite as:*

Perlman A, Vodonos Zilberg A, Bak P, Dreyfuss M, Leventer-Roberts M, Vurembrand Y, Jeffries HE, Fisher E, Steuerman Y, Namir Y, Goldschmidt Y, Souroujon D

Characteristics and Symptoms of App Users Seeking COVID-19–Related Digital Health Information and Remote Services: Retrospective Cohort Study

*J Med Internet Res* 2020;22(10):e23197

URL: <http://www.jmir.org/2020/10/e23197/>

doi: [10.2196/23197](https://doi.org/10.2196/23197)

PMID: [32961527](https://pubmed.ncbi.nlm.nih.gov/32961527/)

©Amichai Perlman, Alina Vodonos Zilberg, Peter Bak, Michael Dreyfuss, Maya Leventer-Roberts, Yael Vurembrand, Howard E Jeffries, Eyal Fisher, Yael Steuerman, Yinat Namir, Yaara Goldschmidt, Daniel Souroujon. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 20.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Identification of Risk Factors and Symptoms of COVID-19: Analysis of Biomedical Literature and Social Media Data

Jouhyun Jeon<sup>1</sup>, PhD; Gaurav Baruah<sup>1</sup>, PhD; Sarah Sarabadani<sup>1</sup>, MASc; Adam Palanica<sup>1</sup>, PhD

Klick Labs, Klick Applied Sciences, Toronto, ON, Canada

**Corresponding Author:**

Jouhyun Jeon, PhD

Klick Labs

Klick Applied Sciences

175 Bloor Street East, Suite 300

Toronto, ON, M4W 3R8

Canada

Phone: 1 416 214 4977

Email: [cjeon@klick.com](mailto:cjeon@klick.com)

## Abstract

**Background:** In December 2019, the COVID-19 outbreak started in China and rapidly spread around the world. Lack of a vaccine or optimized intervention raised the importance of characterizing risk factors and symptoms for the early identification and successful treatment of patients with COVID-19.

**Objective:** This study aims to investigate and analyze biomedical literature and public social media data to understand the association of risk factors and symptoms with the various outcomes observed in patients with COVID-19.

**Methods:** Through semantic analysis, we collected 45 retrospective cohort studies, which evaluated 303 clinical and demographic variables across 13 different outcomes of patients with COVID-19, and 84,140 Twitter posts from 1036 COVID-19-positive users. Machine learning tools to extract biomedical information were introduced to identify mentions of uncommon or novel symptoms in tweets. We then examined and compared two data sets to expand our landscape of risk factors and symptoms related to COVID-19.

**Results:** From the biomedical literature, approximately 90% of clinical and demographic variables showed inconsistent associations with COVID-19 outcomes. Consensus analysis identified 72 risk factors that were specifically associated with individual outcomes. From the social media data, 51 symptoms were characterized and analyzed. By comparing social media data with biomedical literature, we identified 25 novel symptoms that were specifically mentioned in tweets but have been not previously well characterized. Furthermore, there were certain combinations of symptoms that were frequently mentioned together in social media.

**Conclusions:** Identified outcome-specific risk factors, symptoms, and combinations of symptoms may serve as surrogate indicators to identify patients with COVID-19 and predict their clinical outcomes in order to provide appropriate treatments.

(*J Med Internet Res* 2020;22(10):e20509) doi:[10.2196/20509](https://doi.org/10.2196/20509)

## KEYWORDS

SARS-CoV-2; COVID-19; risk factor; symptom; diagnosis; treatment; biomedical literature; social media; Twitter; tweets

## Introduction

COVID-19 is an emerging infectious disease that has quickly spread worldwide. Since its outbreak in China in December 2019, over 4 million cases have been confirmed across more than 200 countries [1] (as of May 20, 2020), with the number of cases continuing to increase. Several studies have characterized possible symptoms (physical or mental features indicating a disease condition) and risk factors (variables

associated with an increased risk of disease) for patients infected with COVID-19 [2,3]. However, the majority of retrospective studies have been based on a single outcome from a single center and counted the number of aggregate cases [4], providing a scattered and incomplete picture of the risk factors for disease severity. Furthermore, uncharacterized or uncommon symptoms have made COVID-19 difficult to diagnose, making it difficult to provide appropriate treatment to patients. Lack of a vaccine or optimized treatment raises the importance of early and definitive diagnosis for this disease. Additionally, there are

limited hospital resources to triage patients based on symptoms to determine who is more or less likely to require intensive treatment (eg, intensive care unit [ICU] admission or intubation).

All of these uncertainties suggest there are urgent needs for a low-cost and efficient method of gathering COVID-19 symptom- and risk factor-related data as quickly as possible to reduce the medical and economic burden in our society. Instead of conducting time-consuming and costly clinical trials to examine patients, an alternative research avenue involves scraping public social media data to investigate potential risk factors of COVID-19 development. Social media provides an efficient method of gathering large amounts of representative data on the general public, in a cost-effective, scalable, and convenient manner any time of day, especially in remote or unattended regions. Here, we systematically investigated published biomedical literature to identify risk factors associated with outcomes of patients with COVID-19. We then gathered public Twitter data from COVID-19-positive users to expand the scientific literature and also examine rare or uncommon symptoms that were not previously well characterized.

## Methods

### Compiling Biomedical Literature on COVID-19, and Identifying Clinical and Demographic Variables

COVID-19 (COVID-19 Open Research Dataset) [5] was used to find biomedical literature on COVID-19. We compiled retrospective studies that investigated clinical and demographic variables in various outcomes of COVID-19. To do this, we collected literature published between January 2020 and March 2020. We then generated two sets of keywords. The first set of keywords represented cohort-based and retrospective studies, and comprised the following keywords: “epidemiological characteristics,” “clinical characteristics,” “risk factors,” “clinical features,” “cohort,” “clinical course,” “clinical findings,” “risk of death,” “pathological characteristics,” “retrospective,” and “mortality risk.” The second set represented COVID-19 (“novel coronavirus,” “coronavirus,” “COVID-19,” “SARS-COV-2,” “severe acute respiratory syndrome coronavirus 2,” “2019 novel coronavirus,” “2019-ncov,” and “coronavirus disease 2019”).

From the semantic analysis, we found 116 articles that contained both keywords and were likely to be relevant to our study. We next extracted 535 tables in those articles using Camelot [6] based on the notion that tables listed clinical and demographical variables, and their associated statistics. Articles without tabulated information ( $n=12$ ) were removed. After careful manual curation of tables, reported data, and article themselves, 45 studies were chosen (Table S1 in Multimedia Appendix 1). In total, 304 clinical and demographic variables were evaluated in 45 studies. They were composed of 92 comorbidities/complications, 49 treatments, 124 lab findings, 34 symptoms, and 4 demographic variables (age, sex, alcohol drinking history, and smoking history). Literature collection was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines (Figure S1 in Multimedia Appendix 1) [7].

### Association Between Clinical and Demographic Variables and Clinical Outcomes

To explore the consistency between individual variables and clinical outcomes, we defined three types of associations. Positive associations indicated that a variable in the outcome group had a hazard or odd ratio  $>1$ , or a higher value with statistical significance ( $P$  value  $<.05$ ) compared to the control group. In case a statistical test was not performed, we decided that there was a positive association when the outcome group had a 1.5-fold higher value than the control group. Negative association indicated that a variable in the outcome group had a hazard or odd ratio  $<1$ , or a lower value with statistical significance ( $P$  value  $<.05$ ) compared to the control group. In case a statistical test was not performed, we decided that there was a negative association when the outcome group had a 1.5-fold lower value than the control group. When there was no significant change between the outcome group and the control group, the variable and clinical outcomes were assigned a “no association” designation. In terms of sex, when there were more males compared to females, we assumed there was a positive association with an outcome based on the case studies of sex and age of patients with COVID-19 in Italy [8] and New York City [9] (as of April 14, 2020). To identify outcome-specific risk factors in biomedical literature, we performed consensus analysis. Risk factors were the variables associated with an increased risk of disease or infection. Seven outcomes, which were studied at least twice, and 107 variables tested two or more times in studies were used for further analyses. Clinical and demographic variables with positive associations in  $>50\%$  of studies, which investigated the same output, were defined as outcome-specific risk factors.

### Compiling Social Media Data and Identifying Symptoms of COVID-19-Positive Users

Twitter was used as the social media source. To identify COVID-19-positive users, we first collected users who used one of these phrases: “my positive COVID test,” “my positive COVID diagnosis,” “I am positive for COVID,” “I tested positive for COVID,” and “I have COVID-19” in their original tweets between January 2020 and March 2020. In total, 1036 users were identified as self-identified patients with COVID-19. We then collected additional tweets generated 14 days before and 14 days after the original tweets mentioning users’ COVID-19 status ( $n=84,140$  tweets). To identify symptoms that were mentioned in tweets, we applied two symptom extraction methods. The Amazon Comprehend Medical tool [10] was applied to an entire set of tweets. Symptoms were physical or mental features indicating a disease condition. We considered two medical entities (symptoms and signs) as symptoms that users mentioned. We also implemented a symptom extraction model using Scispacy, version 0.2.4 (Python Software Foundation). Scispacy can handle scientific document and extracts medical and clinical terminology [11]. The model was trained on publicly available, domain-specific corpus of medical notes, which consists of 1500 PubMed articles with over 10,000 disease and related chemical terms. The model identifies medical name entities in tweet texts. We considered the medical entity “disease” as a symptom that users mentioned. In total, 51

symptoms from 574 COVID-19–positive users were identified from both symptom identification methods.

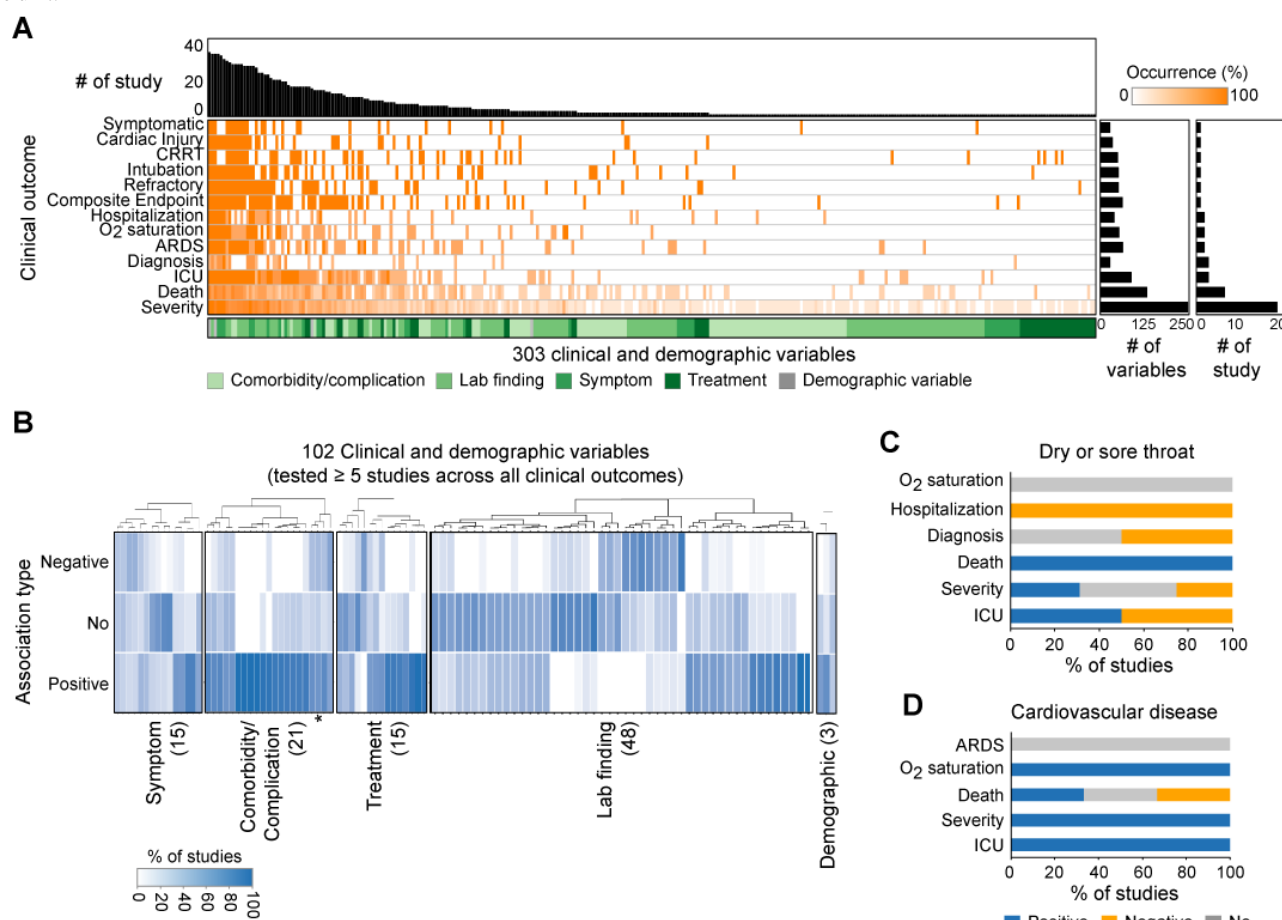
## Results

### Landscape of Clinical and Demographic Variables of COVID-19 in Biomedical Literature

To understand the clinical and demographic properties of COVID-19, we performed a meta-analysis of 45 recently published biomedical studies (Figure 1A, and Table S1 in Multimedia Appendix 1). The literature evaluated 299 clinical variables (92 comorbidities/complications, 124 laboratory findings, 49 treatment options, and 34 symptoms) and 4 demographic variables (age, sex, alcohol drinking history, and smoking history) in 13 clinical outcomes. Seven outcomes (disease severity, death, ICU admission, diagnosis, acute

respiratory distress syndrome [ARDS], O<sub>2</sub> saturation, and hospitalization) were studied at least twice (Figure 1). On average, each study examined 72 variables, and 102 variables were assessed in at least five studies. Age and sex were measured in more than 80% of the studies. Diabetes and hypertension were the most commonly measured comorbidities (>50% of studies). Fever, cough, myalgia/fatigue, chest tightness/dyspnea, diarrhea, and headache/dizziness were the most commonly measured symptoms (>50% of studies). Eleven laboratory test values that measured liver and kidney function (eg, alanine aminotransferase and aspartate aminotransferase) and hematologic index (eg, lymphocytes, platelets, and neutrophils) were examined in >50% of studies. Therapy involving antiviral agents, antibiotics, and oxygen inhalation were used in more than 30% of the studies (Table S2 in Multimedia Appendix 1).

**Figure 1.** Properties of clinical and demographic variables of COVID-19. (A) Landscape of clinical and demographic variables. (B) Association between variables and overall clinical outcomes. Number of variables tested in  $\geq 5$  studies appears in brackets. Asterisk indicated association types of cancer. (C) and (D) Association types of clinical outcomes of COVID-19; association types of (C) dry or sore throat and (D) cardiovascular disease depending on different clinical outcomes were shown. CRRT: continuous renal replacement therapy; ARDS: acute respiratory distress syndrome; ICU: intensive care unit.



We next investigated the association between identified variables and clinical outcomes of patients with COVID-19. We considered 102 frequently tested variables ( $\geq 5$  studies), and examined the proportion of studies that showed positive, negative, and no associations between clinical outcomes and a given variable (Table S3 in Multimedia Appendix 1). Positive associations indicated higher values (eg, disease severity increases as patients get older), while negative associations

indicated lower values (eg, disease severity increases as basophil count decreases) of variables associated with clinical outcomes (see Methods for details). A no-association designation indicated there was no relation between variables and outcomes.

We found that the majority of variables ( $n=95$ , 93%) had inconsistent associations across clinical outcomes showing mixed association types (Figure 1B). Of those, 46 (45%) variables had all three types of associations. For example, cancer



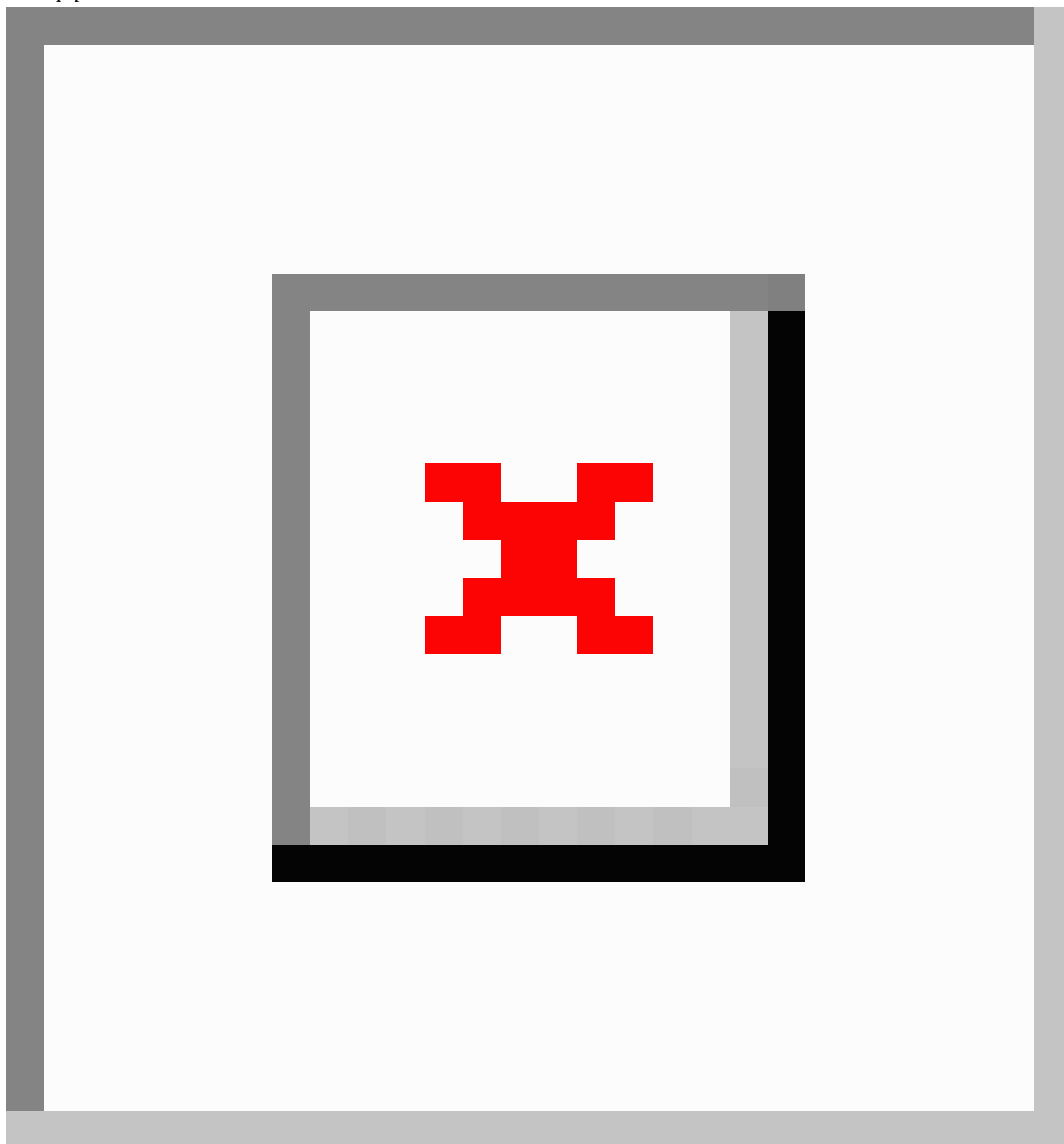
showed positive, negative, and no associations in 58% (n=11), 26% (n=3), and 16% (n=5) of studies, respectively (marked with an asterisk in [Figure 1B](#)). In total, 43% (9/21) of comorbidity/complication, 40% (6/15) of treatment, and 73% (11/15) of symptom variables showed all three types of associations. Laboratory findings showed relatively more consistent associations with clinical outcomes: 38% (18/48) of variables showed all association types. Furthermore, we found that variables had unique association types depending on different clinical outcomes. Dry or sore throat, one known symptom of COVID-19, showed positive, negative, and no associations in death, hospitalization, and O<sub>2</sub> saturation, respectively ([Figure 1C](#)). Meanwhile, it showed mixed associations with other clinical outcomes, such as disease severity, ICU admission, and diagnosis. Cardiovascular disease, one common comorbidity of COVID-19, showed mixed associations in death, positive association with ICU and disease severity, but no association with ARDS ([Figure 1D](#)).

### Consensus Identification of Outcome-Specific Clinical and Demographic Variables

According to biomedical literature at the time of publication, there were no effective treatment options, or well-identified symptoms, comorbidities, and lab findings to predict outcomes of COVID-19. Therefore, it seemed relevant to find a set of clinical and demographic variables that were specific for individual outcomes for better guidance on disease detection, treatment, and control. To generalize the importance of clinical and demographic variables, a consensus (level of agreement) analysis was performed. We collected 107 variables that were

tested at least twice in a given outcome and defined them as outcome-specific risk factors when they showed positive associations with a given outcome in more than half of studies ([Figure 2](#), and Table S4 in [Multimedia Appendix 1](#)). In total, we characterized 72 outcome-specific risk factors from the literature. As shown in [Figure 2A](#), different sets of variables were specifically associated with individual outcomes. Arrhythmia, thyroid disease (comorbidity/complication), confusion/fluster, tonsil swelling, enlargement of lymph nodes/sinus (symptom), and levels of interleukin (IL)-10 and N-terminal pro-brain natriuretic peptide (NT-proBNP; lab finding) were specifically associated with the severity of disease progression. Level of prothrombin time was a specific risk factor for ICU admission. For death, Sequential Organ Failure Assessment (SOFA) score (lab finding) and anemia (symptom) showed positive associations. Fever was specifically associated with O<sub>2</sub> saturation. We also observed 15 variables that were associated with several clinical outcomes ( $\geq 3$  outcomes; [Figure 2B](#)). Age was a risk factor for ARDS, disease severity, death, ICU admission, and O<sub>2</sub> saturation, but not for diagnosis and hospitalization. Sex (male) was a specific variable for ARDS, disease severity, death, and ICU admission. Three lab findings (D-dimer, C reactive protein, and lactate dehydrogenase) were specifically associated with four outcomes. Diabetes and hypertension (comorbidity/complication) were specific risk factors for disease severity and death. Identified outcome-specific variables could be surrogate risk factors to identify patients with COVID-19 and determine their treatment options.

**Figure 2.** Consensus identification of outcome-specific clinical and demographic variables. (A) Outcome-specific clinical and demographic variables in a given outcome of COVID-19. Variables that were only specific for one clinical outcome were shown in the red-dashed box. Clinical and demographic variables that were specific for at least three outcomes are presented in (B). Blue coloring indicates identified outcome-specific variables (risk factors). ICU: SOFA: Sequential Organ Failure Assessment; ARDS: acute respiratory distress syndrome; IL-10: interleukin 10; NT-proBNP: N-terminal pro-brain natriuretic peptide.



### Expanding the COVID-19 Symptom Landscape and Identifying Novel Symptoms Using Social Media

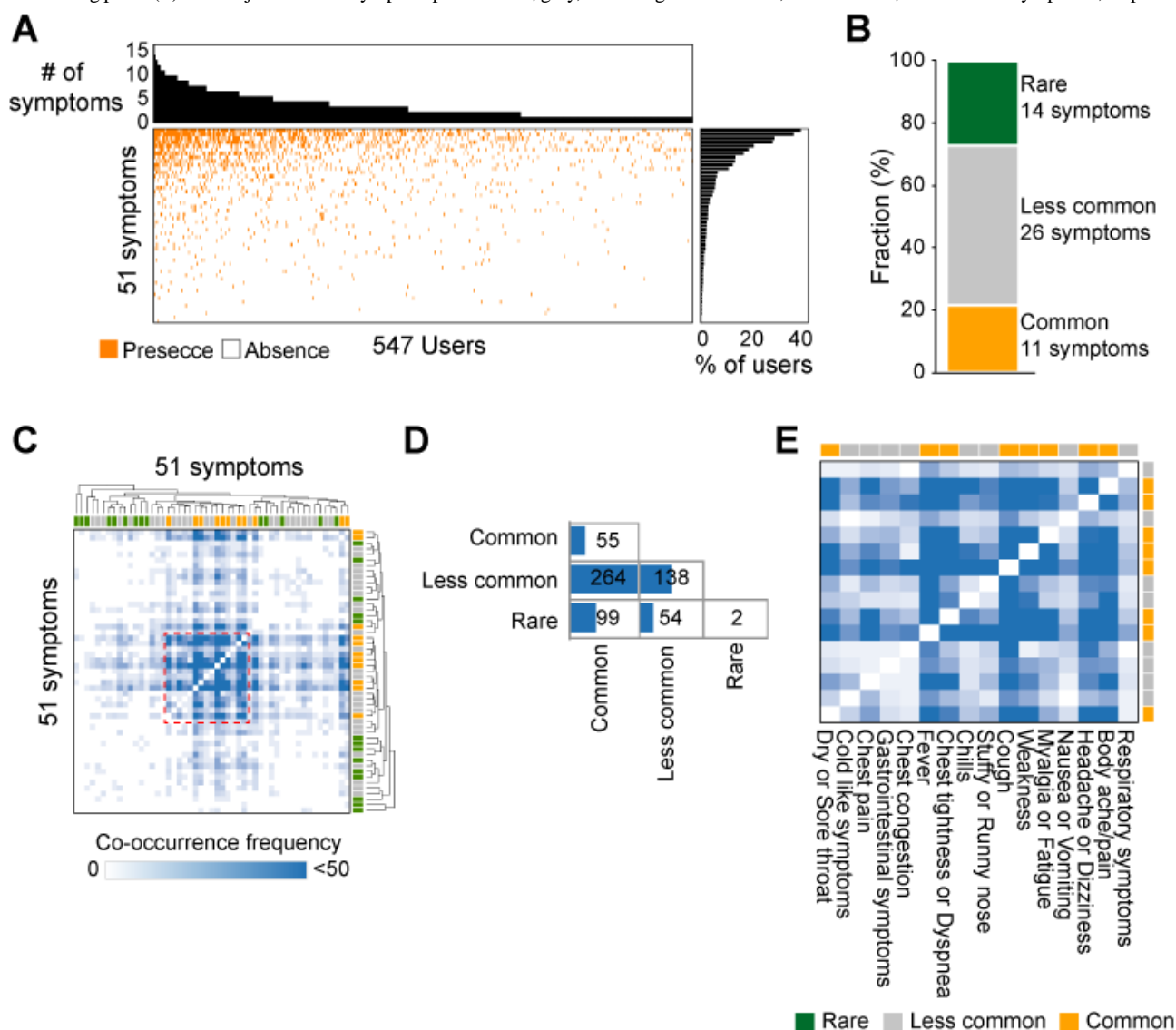
Early identification of symptoms is important for the successful treatment of disease [12]. Although COVID-19 showed heterogeneous and uncharacterized symptoms, a limited number of symptoms known to be associated with infectious diseases, such as fever, cough, and fatigue, were considered in the biomedical literature. Social media can provide rapid and efficient surveillance of disease risk and outbreaks [13,14]. Therefore, we decided to expand the symptom landscape by

integrating social media data with biomedical literature. We first identified 1036 twitter users who introduced themselves as COVID-19–positive patients, and selected 574 users (55%) who openly and voluntarily discussed their COVID-19 symptoms (see Methods for details). In total, 51 symptoms were identified in social media data (Figure 3A, and Table S5 in Multimedia Appendix 1). On average, individual users mentioned 3 different symptoms (range 1–15). We grouped 51 symptoms into three categories based on their frequency of mention (Figure 3B). Common symptoms (n=11) were mentioned by >10% of users. Many were nonspecific symptoms

of respiratory infections, such as fever, cough, and chest tightness/dyspnea (Table S5 in [Multimedia Appendix 1](#)). However, 14 symptoms were potentially COVID-19-specific and rarely reported in social media (ie, in <1% of users). Sputum, dehydration, anemia, urination problem, hair loss, enlargement of lymph nodes or sinus, and oral problems (eg, abrasions in mouth, mouth ulcers, sensitive teeth, toothache, and dry mouth) were defined as rare symptoms. In all, 26 symptoms were observed in 1% to 10% of users (less common). They included chills, chest pain, gastrointestinal symptoms, arthritis, anorexia, allergy-like symptoms, dyssomnias, ear- and eye-related problems, and skin problems such as blister, dry skin, chapped lips, rash, and itching. We next examined the symptoms that were mentioned together in social media ([Figure 3C](#)). We identified 612 co-occurring symptom pairs (Table S6 in [Multimedia Appendix 1](#)). In total, 264 (43%) symptoms ranged from common to less common ([Figure 3D](#)). One major

cluster of symptoms was frequently mentioned together ([Figure 3E](#)). They were composed of 8 common symptoms, such as dry or sore throat, fever, chest tightness/dyspnea, cough, weakness, myalgia/fatigue, headache/dizziness, and body ache/pain (eg, neck and back pain and general body ache), and 8 less common symptoms, such as cold-like symptoms, chest pain and congestion, gastrointestinal symptoms, chills, stuffy or runny nose, nausea/vomiting, and respiratory symptoms such as lung pain. In total, 99 (16%) pairs were between common symptoms and rare symptoms. Cough, fever, headache/dizziness and body ache/pain (common symptoms) co-occurred with sputum, dehydration, and anemia at least 10 times (Table S6 in [Multimedia Appendix 1](#)). There were also two pairs of rare symptoms ([Figure 3D](#)): anemia-weight gain and anemia-urination problems (eg, urinary retention and weakened bladder).

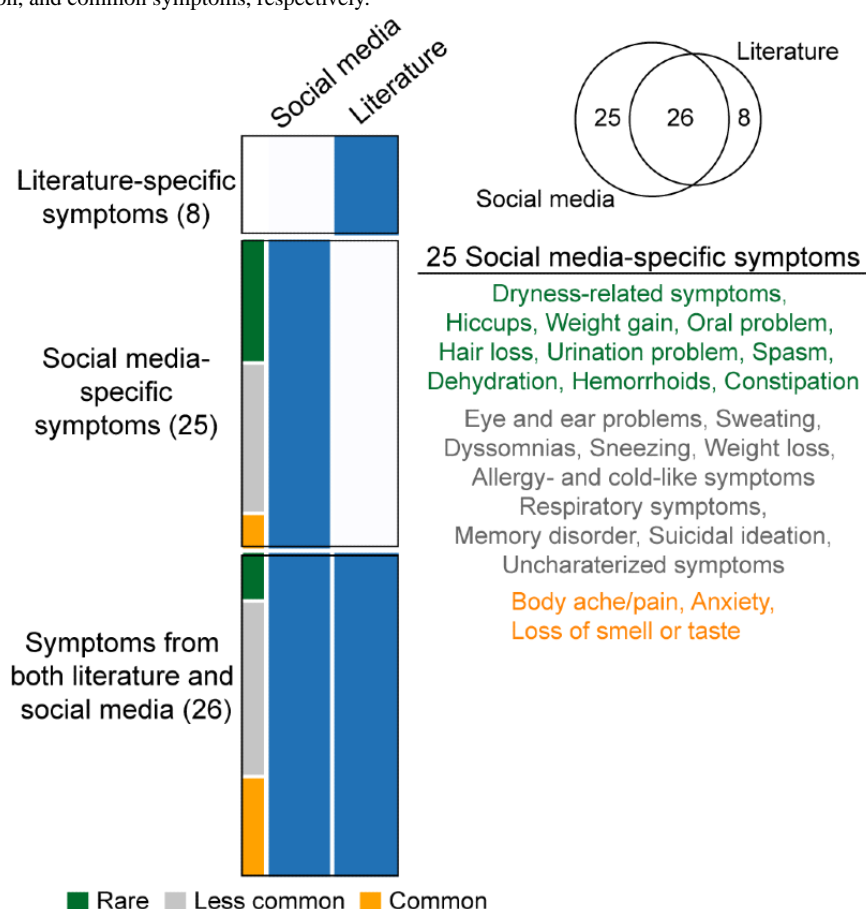
**Figure 3.** COVID-19-related symptoms extracted from social media data. (A) Landscape of symptoms identified from social media data. Orange and white indicate the presence and absence of symptoms in a given user, respectively. (B) Fraction of common, less common, and rare symptoms. Common symptoms were mentioned from >10% of users, and rare symptoms were mentioned from <1% of users. (C) Co-occurrence of symptoms. One major cluster was shown in the red-dashed box. (D) Number of symptoms pairs depending on mentioning frequency. Blue bars (bottom) indicate the number of co-occurring pairs. (E) One major cluster of symptom pairs. Green, gray, and orange indicate rare, less common, and common symptoms, respectively.



Finally, we identified novel symptoms potentially related to COVID-19. We first extended the repertoire of symptoms by integrating social media data with biomedical literature. In total, 59 symptoms were identified (Figure 4, and Table S7 in Multimedia Appendix 1). Indeed, social media identified more symptoms ( $n=51$ ) than the literature ( $n=34$ ), and most symptoms (26/34, 76%) in the literature were equally observed in social media. Interestingly, 25 (42%) were novel symptoms that were only mentioned in social media and were not considered in the literature. Loss of smell or taste, and body ache/pain were frequently mentioned common symptoms in social media. Problems involving eyes (eg, dry eye and eye pain) and ears

(eg, ear pain and earache), sweating, sneezing, and allergy-like symptoms were mentioned at a moderate frequency only in social media. Of 14 rare symptoms, 10 were only observed in social media: dehydration, dryness-related symptoms (eg, feeling dry), oral problem, hair loss, urination problem, spasm, hemorrhoids, constipation, hiccups, and weight gain. These social media-specific rare symptoms would be potential novel symptoms for COVID-19, when we considered that 4 rare symptoms (ie, abdominal pain, anemia, sputum, and enlargement of lymph nodes or sinus) were already evaluated in the literature (Figure 4).

**Figure 4.** Identification of novel symptoms of COVID-19, and comparison of symptoms between biomedical literature and social media data. Symptoms that were observed in the literature or social media are colored in blue; 25 social media-specific symptoms are presented. Green, gray, and orange indicated rare, less common, and common symptoms, respectively.



## Discussion

### Principal Findings

Our meta-analysis based on biomedical literature showed the inconsistency of clinical and demographic variables across clinical outcomes. From the consensus analysis, we identified outcome-specific risk factors that may be helpful to identify patients and predict their specific clinical outcomes. Social media data expanded the repertoire of COVID-19 symptoms from biomedical literature. In addition to more commonly reported symptoms, social media data revealed loss of smell or taste, body ache/pain, and back/neck pain, as well as other less common and rare symptoms such as urination problems, dehydration, allergy-like symptom, and ear- and eye-related

problems. Indeed, loss of smell or taste was recently proposed as one of the key features of a COVID-19 diagnosis model [15], and body ache/pain was suggested by the Centers for Disease Control and Prevention. COVID-19 is an ongoing health problem, and the symptoms that medical institutes and ordinary people should consider will be evolving as more studies are published and social media data are explored. This evolution may help improve the definitions of symptom types (common, less common, and rare symptoms) and social media-specific symptoms.

From social media data, we observed that certain combinations of symptoms were frequently observed among patients with COVID-19. Interestingly, we identified two pairs composed of rare symptoms (anemia-weight gain and anemia-urination

problems, such as urination retention and weakened bladder). It has been shown that COVID-19 attacks hemoglobin in red blood cells and restricts oxygen transportation [16]. A persistent reduction of oxygen transportation leads to the development of anemia [17]. Urinary bladder is enriched with ACE2 positive cells and proposed as a target organ for COVID-19 invasion [18]. Our results highlighted that symptom combinations could guide the reliable identification of patients with COVID-19 rather than a single common symptom that may result in false positives.

### Limitations and Future Work

One of the limitations of our study is the self-reported nature of social media data and the lack of more detailed information from the patients. We observed that 55% of social media users who were self-identified patients with COVID-19 mentioned symptoms, and an additional 8% mentioned potential comorbidities. Thus, only 63% of social media users communicated information on COVID-19 conditions, which means that at least 37% of users could be either false positives (they were not COVID-19-positive users) or asymptomatic patients. Alternatively, it is possible that we have not captured all COVID-19-positive patients in our social media collection due to the limited amount of keyword searches. Nevertheless, various studies have indicated that between 4% and 78% of all COVID-19-positive patients were asymptomatic [19], and this seems to vary widely based on age of patients, test location, and time of testing after infection [20-23]. Thus, our research is in line with other studies demonstrating the vast range of patients with COVID-19 who show or report no symptoms. It should also be noted that Twitter was the source of the social media data we examined, and perhaps more symptoms would be discovered if we analyzed other sources. Twitter does have a wide, representative user base around the world, and provides open source information that can be easily gathered, but future research could examine alternate social media sources.

Although social media may lack depth of patient information, it provides an effective method of collecting breadth of data. Social media data can be gathered noninvasively across the world, 24 hours a day, and is an extremely efficient method [24] for rapidly disseminating new knowledge related to COVID-19. In other words, clinicians and scientists can collect new patient information from various regional locations, as well as quickly circulate public service announcements for a wide range of audiences. Social media hubs provide a useful alternate source of patient data to explore clinical characteristics of various disease states and populations.

Another limitation of our research involves the limited number of available biomedical studies on individual outcomes. Of the 13 reported clinical outcomes for COVID-19, 7 were studied at least two times, limiting opportunities to perform more systematic and consensus analyses of the landscape of risk factors and symptoms. In fact, we observed there was no significant risk factors associated with diagnosis and hospitalization indicating the current lack of clinical understanding at the early stage of COVID-19. Use of additional literature that will be published in the future and electronic health record studies [25] may refine the assessment of risk factors and symptoms, and increase the accuracy of patient identification for different clinical outcomes.

### Conclusion

In this meta-data study, we demonstrated the extensive variability present in clinical and demographic variables across COVID-19 outcomes, and the usefulness of gathering social media data as an effective and alternative way to uncover less common or other types of symptoms that have not been previously reported. Our findings show the practicality and feasibility of employing social media data for investigating new disease states. These practices could be incorporated into routine procedures for early COVID-19 identification and determination of clinical outcomes, in order to provide appropriate interventions and treatments.

### Acknowledgments

We would like to thank all members of the Klick Applied Sciences team for their helpful advice and assistance. We especially thank Maheedhar Kolla, Jeremy J Jurksztowicz, JJ Mifsud, and Michael Lieberman for technical support and scientific advice for the study.

### Authors' Contributions

JJ and GB designed the study. JJ, GB, and SS collected the data. JJ and AP analyzed the data. JJ interpreted the data and wrote the manuscript with contributions from AP. All authors revised the manuscript and contributed to the final review and editing, and have approved the final manuscript. This research was internally funded and received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Conflicts of Interest

None declared.

Multimedia Appendix 1  
Supplementary materials.

[DOCX File, 1315 KB - [jmir\\_v22i10e20509\\_app1.docx](#)]



## References

1. COVID-19 Coronavirus Pandemic. Worldometer. 2020. URL: <https://www.worldometers.info/coronavirus> [accessed 2020-09-30]
2. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: [10.1056/NEJMoa2002032](https://doi.org/10.1056/NEJMoa2002032)] [Medline: [32109013](https://pubmed.ncbi.nlm.nih.gov/32109013/)]
3. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* 2020 Jul 01;180(7):934-943 [FREE Full text] [doi: [10.1001/jamainternmed.2020.0994](https://doi.org/10.1001/jamainternmed.2020.0994)] [Medline: [32167524](https://pubmed.ncbi.nlm.nih.gov/32167524/)]
4. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* 2020 May 19;20(5):533-534. [doi: [10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)] [Medline: [32087114](https://pubmed.ncbi.nlm.nih.gov/32087114/)]
5. Kohlmeier S, Lo K, Wang LL, Yang YY. COVID-19 Open Research Dataset Challenge (CORD-19). Zenodo 2020 Mar:e [FREE Full text] [doi: [10.5281/zenodo.3727291](https://doi.org/10.5281/zenodo.3727291)]
6. atlanhq / camelot. GitHub. 2020. URL: <https://github.com/atlanhq/camelot> [accessed 2020-09-30]
7. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018 Jan 23;319(4):388-396. [doi: [10.1001/jama.2017.19163](https://doi.org/10.1001/jama.2017.19163)] [Medline: [29362800](https://pubmed.ncbi.nlm.nih.gov/29362800/)]
8. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA* 2020 May 12;323(18):1775-1776. [doi: [10.1001/jama.2020.4683](https://doi.org/10.1001/jama.2020.4683)] [Medline: [32203977](https://pubmed.ncbi.nlm.nih.gov/32203977/)]
9. Age, Sex, Existing Conditions of COVID-19 Cases and Deaths. Worldometer. 2020. URL: <https://www.worldometers.info/coronavirus/coronavirus-age-sex-demographics> [accessed 2020-09-30]
10. Amazon Comprehend Medical. AWS. 2020. URL: <https://aws.amazon.com/comprehend/medical/> [accessed 2020-09-30]
11. Neumann M, King D, Beltagy I, Ammar W. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. 2019 Presented at: Proceedings of the 18th BioNLP Workshop and Shared Task; August 2019; Florence, Italy p. 319-327 URL: <https://www.aclweb.org/anthology/W19-5034.pdf> [doi: [10.18653/v1/W19-5034](https://doi.org/10.18653/v1/W19-5034)]
12. Kostopoulou O, Rosen A, Round T, Wright E, Douiri A, Delaney B. Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients. *Br J Gen Pract* 2015 Jan;65(630):e49-e54. [doi: [10.3399/bjgp15X683161](https://doi.org/10.3399/bjgp15X683161)] [Medline: [25548316](https://pubmed.ncbi.nlm.nih.gov/25548316/)]
13. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Prev Med* 2014 Jun;63:112-115 [FREE Full text] [doi: [10.1016/j.ypmed.2014.01.024](https://doi.org/10.1016/j.ypmed.2014.01.024)] [Medline: [24513169](https://pubmed.ncbi.nlm.nih.gov/24513169/)]
14. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS One* 2013;8(12):e83672 [FREE Full text] [doi: [10.1371/journal.pone.0083672](https://doi.org/10.1371/journal.pone.0083672)] [Medline: [24349542](https://pubmed.ncbi.nlm.nih.gov/24349542/)]
15. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020 Jul 11;26(7):1037-1040. [doi: [10.1038/s41591-020-0916-2](https://doi.org/10.1038/s41591-020-0916-2)] [Medline: [32393804](https://pubmed.ncbi.nlm.nih.gov/32393804/)]
16. Liu W, Li H. COVID-19: Attacks the 1-Beta Chain of Hemoglobin and Captures the Porphyrin to Inhibit Human Heme Metabolism. Preprint. ChemRxiv 2020 Apr 13:e [FREE Full text] [doi: [10.26434/chemrxiv.11938173.v7](https://doi.org/10.26434/chemrxiv.11938173.v7)]
17. Rhodes C, Varacallo M. Physiology, Oxygen Transport. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2020:e.
18. Puliatti S, Eissa A, Eissa R, Amato M, Mazzone E, Dell'Oglio P, et al. COVID-19 and urology: a comprehensive review of the literature. *BJU Int* 2020 Jun 06;125(6):E7-E14. [doi: [10.1111/bju.15071](https://doi.org/10.1111/bju.15071)] [Medline: [32249538](https://pubmed.ncbi.nlm.nih.gov/32249538/)]
19. COVID-19: What proportion are asymptomatic? CEBM. 2020. URL: <https://www.cebm.net/covid-19/covid-19-what-proportion-are-asymptomatic> [accessed 2020-09-30]
20. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, CDC COVID-19 Investigation Team. Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility - King County, Washington, March 2020. *MMWR Morb Mortal Wkly Rep* 2020 Apr 03;69(13):377-381 [FREE Full text] [doi: [10.15585/mmwr.mm6913e1](https://doi.org/10.15585/mmwr.mm6913e1)] [Medline: [32240128](https://pubmed.ncbi.nlm.nih.gov/32240128/)]
21. Day M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ* 2020 Mar 23;368:m1165. [doi: [10.1136/bmj.m1165](https://doi.org/10.1136/bmj.m1165)] [Medline: [32205334](https://pubmed.ncbi.nlm.nih.gov/32205334/)]
22. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung S, Hayashi K, et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int J Infect Dis* 2020 May;94:154-155 [FREE Full text] [doi: [10.1016/j.ijid.2020.03.020](https://doi.org/10.1016/j.ijid.2020.03.020)] [Medline: [32179137](https://pubmed.ncbi.nlm.nih.gov/32179137/)]
23. Quilty BJ, Clifford S, Flasche S, Eggo RM, CMMID nCoV working group. Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). *Euro Surveill* 2020 Feb;25(5):e [FREE Full text] [doi: [10.2807/1560-7917.ES.2020.25.5.2000080](https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000080)] [Medline: [32046816](https://pubmed.ncbi.nlm.nih.gov/32046816/)]
24. Chan AKM, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. *Anaesthesia* 2020 Mar 30:e [FREE Full text] [doi: [10.1111/anae.15057](https://doi.org/10.1111/anae.15057)] [Medline: [32227594](https://pubmed.ncbi.nlm.nih.gov/32227594/)]

25. The OpenSAFELY Collaborative, Williamson E, Walker A, Bhaskaran K, Bacon S, Bates C, et al. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. Preprint. medRxiv 2020 May 7:e [FREE Full text] [doi: [10.1101/2020.05.06.20092999](https://doi.org/10.1101/2020.05.06.20092999)]

## Abbreviations

**ARDS:** acute respiratory distress syndrome

**CORD-19:** COVID-19 Open Research Dataset

**ICU:** intensive care unit

**IL:** interleukin

**NT-proBNP:** N-terminal pro-brain natriuretic peptide

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analysis

**SOFA:** Sequential Organ Failure Assessment

*Edited by G Eysenbach; submitted 20.05.20; peer-reviewed by M Hoenigl, P Banik; comments to author 03.09.20; revised version received 08.09.20; accepted 13.09.20; published 02.10.20.*

*Please cite as:*

*Jeon J, Baruah G, Sarabadani S, Palanica A*

*Identification of Risk Factors and Symptoms of COVID-19: Analysis of Biomedical Literature and Social Media Data*

*J Med Internet Res 2020;22(10):e20509*

*URL:* <https://www.jmir.org/2020/10/e20509>

*doi:* [10.2196/20509](https://doi.org/10.2196/20509)

*PMID:* [32936770](https://pubmed.ncbi.nlm.nih.gov/32936770/)

©Jouhyun Jeon, Gaurav Baruah, Sarah Sarabadani, Adam Palanica. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 02.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Association of Web-Based Physical Education With Mental Health of College Students in Wuhan During the COVID-19 Outbreak: Cross-Sectional Survey Study

Cheng-Hu Deng<sup>1</sup>, MA; Jing-Qiang Wang<sup>2</sup>, MA; Li-Ming Zhu<sup>3</sup>, MA; He-Wang Liu<sup>4</sup>, MA; Yu Guo<sup>5</sup>, MD, PhD; Xue-Hua Peng<sup>5</sup>, MD, PhD; Jian-Bo Shao<sup>5</sup>, MD, PhD; Wei Xia<sup>5</sup>, MD, PhD

<sup>1</sup>Department of Physical Education, Wuhan University of Technology, Wuhan, China

<sup>2</sup>Department of Physical Education, Hubei Business College, Wuhan, China

<sup>3</sup>Department of Physical Education, Jiangnan University, Wuhan, China

<sup>4</sup>Department of Physical Education, Huazhong Agricultural University, Wuhan, China

<sup>5</sup>Department of Imaging Center, Wuhan Children's Hospital (Wuhan Maternal and Child Healthcare Hospital), Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

**Corresponding Author:**

Wei Xia, MD, PhD

Department of Imaging Center

Wuhan Children's Hospital (Wuhan Maternal and Child Healthcare Hospital)

Tongji Medical College, Huazhong University of Science and Technology

Hubei

Wuhan 430015

China

Phone: 86 1 338 755 3260

Email: [176778380@qq.com](mailto:176778380@qq.com)

## Abstract

**Background:** The COVID-19 outbreak has affected people's health worldwide. For college students, web-based physical education is a challenge, as these course are normally offered outdoors.

**Objective:** The aim of this study was to use data from a web-based survey to evaluate the relationship between the mental health status of college students and their sports-related lifestyles. Problems related to web-based physical education were also examined.

**Methods:** A web-based survey was conducted by snowball sampling from May 8 to 11, 2020. Demographic data, mental health status, and sports-related lifestyles of college students in Wuhan as well as issues related to web-based physical education were collected. Mental health status was assessed by the Depression, Anxiety, and Stress Scale (DASS-21).

**Results:** The study included 1607 respondents from 267 cities. The average scores of the DASS-21 subscales (2.46 for depression, 1.48 for anxiety, and 2.59 for stress) were significantly lower in our study than in a previous study ( $P<.05$ ). Lower DASS-21 scores were significantly correlated with regular exercise, maintaining exercise habits during the outbreak of COVID-19, exercising more than 1 to 2 times a week, exercise duration  $>1$  hour, and  $>2000$  pedometer steps (all  $P<.05$ ). None of the three forms of web-based physical education was preferred by more than 50% of respondents. Frequent technical problems were confronted by 1087/1607 students (67.6%). Shape-up exercises (846/1607, 52.6%), a designed combination of exercises (710/1607, 44.2%), and Chinese kung fu (559/1607, 34.8%) were suggested sports for web-based physical education.

**Conclusions:** Mental status was significantly correlated with regular exercise and sufficient exercise duration. Professional physical guidance is needed for college students in selected sports. Exercises not meeting students' preferences, frequent technical problems, and the distant interaction involved in web-based physical education were the main problems that should be solved in future.

(*J Med Internet Res* 2020;22(10):e21301) doi:[10.2196/21301](https://doi.org/10.2196/21301)

**KEYWORDS**

COVID-19; college students; mental status; physical education; young adults; web-based education; global health; web-based survey; physical activity; mental health

## Introduction

In December 2019, a novel coronavirus, SARS-CoV-2, was first recognized in Wuhan and then quickly spread worldwide, infecting millions of people [1]. Furthermore, uninfected people were also greatly affected and were required to adopt a totally different lifestyle because of quarantine in Wuhan [2]. Many basic necessities of living, as well as education, were transferred to the internet [3,4]. Web-based education was not new in many fields [5]; however, it is novel in sports education, which requires essential interaction between teachers and students.

Previous studies have stated that mental illness symptoms are common among university students according to the Depression, Anxiety, and Stress Scale (DASS-21) [6,7]. Dogra's study [8] described that vigorous physical activity is associated with lower possibility of poor mental health and depressive symptoms. As physical education plays an important role in relieving pressure not only on the body but also on the mind [9], it is particularly essential during the pandemic [10]. According to a report by the United Nations Educational Scientific and Cultural Organization (UNESCO), more than 160 countries have shut down schools [11]; therefore, web-based physical education is an issue that most countries must face. It has been reported that SARS-CoV-2 will coexist with humans for a long time [12], which suggests that web-based physical education could last for a relatively long time in the future. A recent scientometric analysis reviewed research topics related to COVID-19; however, the impact of COVID-19 on physical education was not researched [13].

Since February, web-based physical education, which is the only choice in a quarantined city, has been adopted by many universities and colleges in Wuhan [14]. Whether the web-based physical education was efficient or a waste of time may be reflected by the mental and physical state of students. Several studies have reported the mental state of the general population, medical staff, or college students during the peak of the COVID-19 epidemic; these studies have proved the existence of anxiety, depression, and stress in different populations [15-18]. However, as far as we know, no report has evaluated the mental and physical state of students in universities and colleges in Wuhan after the implementation of web-based physical education.

As a newly developed mode of physical education, web-based physical education presents challenges and opportunities. Because of the characteristics of physical education, suitable sports should be chosen while considering the restrictions of sports fields and related equipment. Also, the performance of network platforms and teaching resources for web-based physical education remains a question. Furthermore, the extended effects of web-based physical education on family members when the activities are performed at home is a concern, which may be related to the similar atmosphere to class learning.

Therefore, the primary aim of our study is to describe the association between the mental health and physical activity of students in Wuhan universities and colleges after a 3-month web-based physical education program. The secondary aim of our study is to identify the problems that students encountered

during the web-based study, and we hope that these problems can be solved by educators in the near future.

## Methods

### Study Design and Study Population

This cross-sectional study was conducted from May 8 to 11, 2020. A snowball strategy was employed to recruit students in Wuhan universities and colleges after a 3-month web-based physical education program.

### Procedure

To minimize face-to-face interactions as recommended by the Chinese government, an anonymous questionnaire was completed through a web-based survey platform (SurveyStar, Changsha Ranxing Science and Technology). The web-based survey was first disseminated by teachers to students through the educational platform, and the students were encouraged to spread it to other students. This study was approved by the institutional review board of Wuhan Children's Hospital (WHCH 2020029). Informed consent was requested at the very beginning of this web-based survey, and informed consent was provided by all the respondents.

### Measurements and Outcomes

A structured questionnaire was employed to evaluate the mental health status and sports-related lifestyles of students in Wuhan universities and colleges as well as their issues related to web-based physical education. There were 45 items included in the questionnaire, and approximately 6 minutes were required to complete it.

Demographic data included gender, age, grade, residential location, and BMI, which was calculated as weight in kilograms divided by height in meters squared. The categorization of BMI was based on reference data for Chinese people ( $BMI < 18.5$ ,  $18.5 \leq BMI < 23$ ,  $23 \leq BMI < 26$ ,  $26 \leq BMI$ ) [19]. The mental health status of students was evaluated by the DASS-21, which consists of depression, anxiety, and stress subscales [20]. Sports-related lifestyle variables included exercise habits, changes in weight, frequency and duration of exercise, preferred sports, pedometer steps, and access to sports facilities. The minimal amount of weight to indicate weight loss or weight increase was 3 kg. The average number of pedometer steps was calculated as the average number of steps taken per day during the past month according to a mobile phone app, WeChat (Tencent). There was no control for other physical activities or exercises performed in addition to web-based physical education. Variables for characteristics of web-based physical education included different exercise classes offered, frequency of web-based physical education, influence on other family members which may be related to the similar atmosphere to class learning, problems confronted, and suggested sports for web-based physical education.

In addition to the total score of the DASS-21 scale, the scores of the three subscales were calculated as follows [15]: the depression subscale consisted of questions 3, 5, 10, 13, 16, 17, and 21, classified into normal (0-9), mild depression (10-12), moderate depression (13-20), severe depression (21-27), and

extremely severe depression (28-42); the anxiety subscale consisted of questions 2, 4, 7, 9, 15, 19, and 20, classified into normal (0-6), mild anxiety (7-9), moderate anxiety (10-14), severe anxiety (15-19), and extremely severe anxiety (20-42); the stress subscale consisted of questions 1, 6, 8, 11, 12, 14, and 18, classified into normal (0-10), mild stress (11-18), moderate stress (19-26), severe stress (27-34), and extremely severe stress (35-42).

### Statistical Analysis

Respondents who did not complete the questionnaire were not included in our study; we deleted the respondents with missing data. Descriptive statistics were employed for categorical variables, including demographic data, sports-related lifestyle, and issues related to web-based physical education. Each categorical variable was presented as the percentage of responses to the corresponding question, which was calculated by dividing the number of respondents per response by the number of total responses to the question. The scores of the DASS-21 scale and its subscales were expressed as mean (SD) as well as median (IQR). The Cronbach alpha values of the reliability of the DASS-21 and its subscales were calculated as a measure of internal consistency. We performed a descriptive comparison of the mean scores of the subscales of the DASS-21 between our study and Wang's first study [15], which described the mental health state of the general population during the peak of the COVID-19 epidemic. Linear regressions were used to

analyze the univariate associations between sports-related lifestyle variables and the DASS-21 scales. All the tests were two-tailed, with a significance level of  $P < .05$ . Statistical analysis was performed using SPSS 19.0 (IBM Corporation).

## Results

### Survey Respondents

A total of 1673 surveys were received; as 66 respondents did not complete the questionnaire, we included 1607 (96.1%) respondents from 267 cities in the study. Questionnaires were received on the first day (May 8) from 292 respondents, on the second day (May 9) from 1130 respondents, on the third day (May 10) from 130 respondents, and on the fourth day (May 11) from 55 respondents.

### Demographic Data

The demographic features of all the students in our study are shown in Table 1. Most respondents were male (1041/1607, 64.8%) and were aged 18 to 22 years (1573/1607, 97.9%). Most of the students were freshmen or sophomores (1524/1607, 94.8%), and urban areas (723, 45%) were the most common place of residence of the college students. The BMI values of most of the 1607 college students were in the normal range (969, 60.3%), while the BMI values of 638 respondents (39.7%) were out of the normal range.

**Table 1.** The demographic characteristics of the respondents (N=1607), n (%).

Demographic characteristic	Value
<b>Gender</b>	
Male	1041 (64.8)
Female	566 (35.2)
<b>Age (years)</b>	
<18	20 (1.2)
18-22	1573 (97.9)
>22	14 (0.9)
<b>Grade</b>	
Freshman	784 (48.8)
Sophomore	740 (46.0)
Junior	70 (4.4)
Senior	13 (0.8)
<b>Place of residence</b>	
Urban	723 (45.0)
Rural-urban	431 (26.8)
Rural	453 (28.2)
<b>BMI, kg/m<sup>2</sup></b>	
<18.5	275 (17.1)
18.5-22	969 (60.3)
23-25	231 (14.4)
≥26	132 (8.2)



## Mental Health Status

The respondents' mental health status was measured by the DASS-21 scale. The mean (SD) of the total DASS-21 score was 6.52 (7.86). The mean (SD) values of the depression, anxiety, and stress subscales were 2.46 (3.02), 1.48 (2.35), and 2.59 (3.09), respectively. The median (IQR) of the total score on the DASS-21 was 4 (1-10). The median (IQR) values of the depression, anxiety, and stress subscales were 1 (0-4), 0 (0-2), and 1 (0-4), respectively. The comparison between the mean

(SD) values in our study and Wang's study [15] is as follows: for the depression subscale, 2.46 (3.02) versus 6.25 (7.16); for the anxiety subscale, 1.48 (2.35) versus 6.16 (6.57); for the stress subscale, 2.59 (3.09) versus 7.76 (7.74). The classification of responses in the different groups for the depression, anxiety, and stress subscales are shown in Table 2. The Cronbach alpha value of the reliability of the DASS-21 was .94, and the Cronbach alpha values for the depression, anxiety, and stress subscales were .84, .85, and .86, separately.

**Table 2.** Classification of responses to the depression, anxiety, and stress subscales of the DASS-21 (N=1607), n (%).

DASS-21 <sup>a</sup> subscale	Normal	Mild	Moderate	Severe	Extremely severe
Depression	1551 (96.5)	36 (2.2)	19 (1.2)	1 (0.1)	0 (0.0)
Anxiety	1519 (94.5)	71 (4.4)	14 (0.9)	2 (0.1)	1 (0.1)
Stress	1574 (97.9)	30 (1.9)	3 (0.2)	0 (0.0)	0 (0.0)

<sup>a</sup>DASS-21: Depression, Anxiety, and Stress Scale.

## Sports-Related Lifestyle and Exercise Status

The sports-related lifestyle variables of the college students who responded to the survey are shown in Table 3. Although most of the 1607 students (1088, 67.7%) exercised regularly, 1279 (79.6%) of the students were disturbed by the outbreak of

COVID-19 and spent less time on sports (826, 51.4%) and/or gained weight (592, 36.8%). Exercising <3 times a week was observed in 1010/1607 students (62.9%), and <2000 average pedometer steps were observed in 1155/1607 students (71.9%). Restrictions on access to sports facilities were experienced by 1198/1607 students (74.5%).

**Table 3.** Sports-related lifestyle variables of the survey respondents after the outbreak of COVID-19 (N=1607), n (%).

Sports-related lifestyle variable	Value
<b>Regular exercise<sup>a</sup></b>	
Yes	1088 (67.7)
No	519 (32.3)
<b>Negative influence of COVID-19 on exercise habits</b>	
No, exercise habits were maintained	328 (20.4)
Yes, but only slightly	889 (55.3)
Yes, it has a great impact on my exercise habits	390 (24.3)
<b>Time spent on sports after the outbreak of COVID-19<sup>b</sup></b>	
Less	826 (51.4)
Same	460 (28.6)
More	321 (20.0)
<b>Weight change after the outbreak of COVID-19<sup>c</sup></b>	
Less	321 (20.0)
Same	694 (43.2)
More	592 (36.8)
<b>Frequency of exercise</b>	
Occasionally or never	549 (34.2)
1 to 2 times per week	461 (28.7)
≥3 times per week	377 (23.4)
Every day	220 (13.7)
<b>Average duration of exercise performed in a week (hours)</b>	
<1	1253 (78.0)
>1	354 (22.0)
<b>Favorite sport after the outbreak of COVID-19</b>	
High-intensity interval training	119 (7.4)
Shape-up exercises	200 (12.4)
Strength training	312 (19.4)
Ball game	288 (17.9)
Walking	665 (41.4)
Body combat	23 (1.4)
<b>Average pedometer steps<sup>d</sup></b>	
0-500	413 (25.7)
501-2000	742 (46.2)
2001-4000	224 (13.9)
>4000	228 (14.2)
<b>Restricted access to sports facilities</b>	
Yes	1198 (74.5)
No	409 (25.5)

<sup>a</sup>Exercising regularly was defined as ≥3 times a week and ≥60 minutes each time.

<sup>b</sup>Time spent on sports was defined as time spent on all types of sports, including web-based physical education.

<sup>c</sup>The minimum amount of weight to indicate weight loss or weight increase was 3 kilograms.

<sup>d</sup>The average number of pedometer steps was calculated as the average number of steps taken per day during the past month according to the WeChat mobile phone app.

The differences in the association between sports-related lifestyle variables and the scores of the DASS-21 and its subscales are represented in Table 4. The respondents who exercised regularly had lower scores on the DASS-21 and all its subscales (for depression,  $B=-1.257$ ,  $t=-7.962$ ,  $P<.001$ ; for anxiety,  $B=-0.700$ ,  $t=-5.636$ ,  $P<.001$ ; for stress,  $B=-1.013$ ,  $t=-6.211$ ,  $P<.001$ ; for total score,  $B=-2.969$ ,  $t=-7.197$ ,  $P<.001$ ), as well as the respondents who maintained their exercise habits during the outbreak (for depression,  $B=-2.017$ ,  $t=-9.171$ ,  $P<.001$ ; for anxiety,  $B=-1.211$ ,  $t=-6.988$ ,  $P<.001$ ; for stress,  $B=-2.198$ ,  $t=-9.788$ ,  $P<.001$ ; for total score,  $B=-5.427$ ,  $t=-9.491$ ,  $P<.001$ ) or were influenced little by the outbreak (for depression,  $B=-1.301$ ,  $t=-7.299$ ,  $P<.001$ ; for anxiety,  $B=-0.783$ ,  $t=-5.572$ ,  $P<.001$ ; for stress,  $B=-1.446$ ,  $t=-7.941$ ,  $P<.001$ ; for total score,  $B=-3.530$ ,  $t=-7.616$ ,  $P<.001$ ). The respondents who exercised more than 1 to 2 times a week demonstrated significantly lower scores on the DASS-21 and all its subscales compared to respondents who exercised occasionally, with all  $P<.05$ . The respondents who exercised  $>1$  hour had lower total scores ( $B=-1.350$ ,  $t=-2.861$ ,  $P=.004$ ) and lower scores on the depression ( $B=-0.588$ ,  $t=-3.248$ ,  $P=.001$ ) and stress ( $B=-0.503$ ,  $t=-2.708$ ,  $P=.007$ ) subscales of the DASS-21 compared to respondents who exercised  $<1$  hour. Respondents with  $>2000$  average pedometer steps had significantly lower scores on the

DASS-21 and all its subscales compared to respondents with  $<599$  average steps, with all  $P<.05$ .

### Issues of Web-Based Physical Education

Three main modes of web-based physical education were adopted by universities and colleges in Wuhan. Interaction between teachers and students was the most common mode (1056/1607, 65.7%). Moreover, this was the only mode in which students could interact with the teacher; the other two modes involved unilateral teaching. Web-based physical education was accessed once per week or less by 1256/1607 students (78.2%). Surprisingly, the family members of 728/1607 students (45.3%) were motivated to exercise because of web-based physical education. Many problems arose during web-based physical education, which were confronted by 1087/1607 students (67.6%). Considering convenience and availability, the respondents suggested that shape-up exercises, designed combinations of exercise by teachers for a specific purpose, and Chinese kung fu were suitable sports for web-based physical education. Here, the designed combinations of exercise were combinations of various physical education exercises that were designed by teachers, such as a combination of shape-up exercises and Chinese kung fu. Detailed information is summarized in Table 5.

**Table 4.** Associations between sports-related lifestyle variables and scores on the DASS-21 and its subscales (N=1607).

Sports-related lifestyle variable	Depression subscale score			Anxiety subscale score			Stress subscale score			Total DASS-21 <sup>a</sup> score		
	B	t	P value	B	t	P value	B	t	P value	B	t	P value
<b>Exercise regularly<sup>b</sup></b>												
Yes	-.195	-7.962	<.001	-.139	-5.636	<.001	-.153	-6.211	<.001	-.177	-7.197	<.001
No	Ref. <sup>c</sup>	N/A <sup>d</sup>	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A
<b>Negative influence of COVID-19 on exercise habits</b>												
No, exercise habits were maintained	-.270	-9.171	<.001	-.208	-6.988	<.001	-.287	-9.788	<.001	-.279	-9.491	<.001
Yes, but only slightly	-.215	-7.299	<.001	-.166	-5.572	<.001	-.233	-7.941	<.001	-.223	-7.616	<.001
Yes, it has a great impact on my exercise habits	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A
<b>Time spent on sports after the outbreak of COVID-19<sup>e</sup></b>												
More	-.075	-2.854	.004	-.037	-1.398	.16	-.046	-1.750	.08	-.058	-2.202	.03
Same	-.102	-3.896	<.001	-.043	-1.645	.10	-.100	-3.807	<.001	-.091	-3.485	.001
Less	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A
<b>Weight change after the outbreak of COVID-19<sup>f</sup></b>												
Less	.018	0.633	.53	.012	0.442	.66	-.009	-0.322	.75	.007	0.249	.80
Same	-.035	-1.275	.20	-.036	-1.300	.20	-.081	-2.938	.003	-.056	-2.034	.04
More	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A
<b>Frequency of exercise</b>												
Every day	-.109	-3.998	<.001	-.095	-3.477	.001	-.086	-3.155	.002	-.104	-3.820	<.001
≥3 times per week	-.146	-5.212	<.001	-.096	-3.387	.001	-.122	-4.337	<.001	-.133	-4.724	<.001
1 to 2 times per week	-.112	-3.946	<.001	-.089	-3.124	.002	-.084	-2.949	.003	-.102	-3.613	<.001
Occasionally or never	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A
<b>Average duration of exercise performed in a week</b>												
>1 hour	-.081	-3.248	.001	-.046	-1.833	.07	-.067	-2.708	.007	-.071	-2.861	.004
<1 hour	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A
<b>Average pedometer steps<sup>g</sup></b>												
>4000	-.123	-4.283	<.001	-.100	-3.487	.001	-.115	-4.026	<.001	-.122	-4.275	<.001
2001-4000	-.103	-3.620	<.001	-.062	-2.175	.03	-.085	-2.971	.003	-.092	-3.212	.001
501-2000	-.078	-2.566	.01	-.044	-1.443	.15	-.063	-2.078	.04	-.068	-2.236	.03
0-500	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A
<b>Restricted access to sports facilities</b>												
No	-.040	-1.602	.11	-.027	-1.065	.29	-.084	-3.368	.001	-.056	-2.258	.02
Yes	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A	Ref.	N/A	N/A

<sup>a</sup>DASS-21: Depression, Anxiety, and Stress Scale.<sup>b</sup>Exercising regularly was defined as ≥3 times a week and ≥60 minutes each time.<sup>c</sup>Ref.: reference.<sup>d</sup>N/A: not applicable.<sup>e</sup>Time spent on sports was defined as time spent on all types of sports, including web-based physical education.

<sup>f</sup>The minimum amount of weight to indicate weight loss or weight increase was 3 kilograms.

<sup>g</sup>The average number of pedometer steps was calculated as the average number of steps taken per day during the past month according to the WeChat mobile phone app.

**Table 5.** The conditions and problems of web-based physical education (N=1607), n (%).

Conditions and problems of web-based physical education	Value
<b>Participation in web-based physical education</b>	
Watching recorded video	841 (52.3)
Watching real-time video	820 (51.0)
Communicating with a teacher on an education platform	1056 (65.7)
<b>Preferred type of web-based physical education</b>	
Watching recorded video	497 (30.9)
Watching real-time video	536 (33.4)
Communicating with a teacher on an education platform	574 (35.7)
<b>Frequency of participation in web-based physical education</b>	
Once every two weeks	25 (1.6)
Once per week	1231 (76.6)
Twice per week	351 (21.8)
<b>Has web-based physical education motivated other family members to exercise?</b>	
Yes	728 (45.3)
No	879 (54.7)
<b>Problems confronted during web-based physical education</b>	
Network instability	752 (46.8)
Lack of familiarity with software	384 (23.9)
No interaction with teacher	481 (29.9)
Lack of self-control	608 (37.8)
Inability to keep up with the lesson	73 (4.5)
<b>Frequency of technical problems (network, software, platform)</b>	
Every class	142 (8.8)
>4 times per month	270 (16.8)
≤4 times per month	675 (42.0)
Never	520 (32.4)
<b>Would you like to continue to participate in web-based physical education?</b>	
Yes	760 (47.3)
No	847 (52.7)
<b>Suitable sports suggested for web-based physical education</b>	
Shape-up exercise	846 (52.6)
Designed combination of exercises	710 (44.2)
Chinese kung fu	559 (34.8)
Rhythmic sport	459 (28.6)
Table tennis	364 (22.7)



## Discussion

### Principal Findings

The main purposes of our study were to determine the mental status of college students using the DASS-21 after 3 months of web-based physical education and to evaluate the relationship of the students' mental health status with their sports-related lifestyle. A web-based survey was employed to collect related information. The results showed that the average scores on the DASS-21 subscales were significantly lower than in a previous study. Lower DASS-21 scores were significantly correlated with positive sports-related lifestyle. Furthermore, web-based physical education was unsatisfactory due to several issues related to technology and content.

### Relationship Between Mental Health Status and Sports-Related Lifestyle

Previous studies have proven that the general population, including college students, suffered numerous negative effects induced by the outbreak of COVID-19 [21,22]. According to the DASS-21, depression, anxiety, and stress could be observed in various populations during the outbreak of COVID-19 in China, which lasted for at least 4 weeks [15]; these populations included the general population [21], general workforce [23], psychiatric patients [24], and health care professionals [25]. All groups in the general population were required to change and rebuild their lifestyles. For college students, in addition to the changes in their ordinary lifestyles, a brand new education style was rapidly established. Physical education, typically an outdoor course, was required to be conducted on the internet. Whether physical education is essential during the COVID-19 epidemic and how to suitably provide it are issues that most countries will be facing for a long period of time.

In Wang's study [15], which reported a higher psychological impact of COVID-19 in respondents aged 12-21.4 years, it was suggested that respondents in this age group might be affected by prolonged school closure and require web-based education support. Comparing our study with Wang's study, we observed lower scores on the DASS-21 subscales, which were obtained 3 months after web-based physical education was established. The difference between our study and the previous study may result from the restoration of web-based education support, of which physical education was an important part. To demonstrate the importance of exercise in our study, linear regression was employed to analyze the correlation between the DASS-21 scores and sports-related lifestyle variables. Unsurprisingly, lower scores on the DASS-21 and its subscales were observed in respondents who exercised regularly and maintained their exercise habits during the outbreak of COVID-19. The respondents who exercised more than 1 to 2 times a week, had an exercise duration >1 hour, and had >2000 average pedometer steps had significantly lower scores on the DASS-21 and all its subscales compared to other participants. These data suggest that exercise, especially regular exercise with sufficient duration, is related to a lower risk of mental disturbance, which is in accordance with a previous study [26]. It is surprising that although more than 70% respondents accumulated <2000 pedometer steps per day, the mean DASS-21 scores were low.

We believe that the low number of steps per day may be related to the exercises chosen for the physical classes during COVID-19 confinement. Some exercises may not involve many steps, such as tai chi and shape-up exercises. According to a large cross-section study, which verified that all exercise types are significantly associated with lower mental health burden [27], the exercise itself mattered rather than the type. Strong evidence from a meta-analysis supported that exercise can protect populations from depression regardless of age and geographical region [28], which may also apply to the COVID-19 pandemic. Therefore, we suggest that a positive sports-related lifestyle is significantly associated with mental health during the confines of the COVID-19 pandemic.

### Issues Related to Web-Based Physical Education

It is unfortunate that 1198/1607 respondents (74.5%) were restricted from using sports facilities, which may be an obstacle preventing them from exercising regularly. Moreover, the BMI of 638/1607 respondents (39.7%) was out of the normal range, which may be related to an unhealthy lifestyle [29]. These results suggest that the respondents were in need of professional guidance for physical education, considering the available facilities and equipment. Interestingly, 728/1607 respondents (45.3%) observed that other family members were motivated to exercise by web-based physical education. This may be related to the similar atmosphere to class learning, in which the other participating family members can be considered as classmates or companions. This also suggests that web-based physical education is not only a new learning style for college students but may also be a new lifestyle for the general population [30,31].

As stated above, effective web-based physical education is essential for lifestyle rebuilding not only for college students but also for the general population as part of behavior therapy [32], health education [33], and promotion by local health authorities [34]. However, as a totally new mode of physical education, web-based learning involved several issues that must be noted. As far as we know, only three types of web-based physical education were available to the college students, including watching recorded video, watching real-time video, and communicating with a teacher on an education platform. However, none of these types was preferred by more than 50% of respondents, and more than 50% of respondents did not want to experience physical education on the internet. However, in contrast, motivation of family members to exercise by web-based physical education was observed by nearly 50% of respondents. This suggests that web-based physical education is welcomed by the general population but cannot meet the requirements of college students. We believe that the dissatisfaction of college students may result from the comparison between web-based physical classes and face-to-face classes, whereas the general population would be more interested in trying this new style without having experienced face-to-face classes. Moreover, frequent technical problems were confronted by 1087/1607 respondents (67.6%), including network instability (752/1607, 46.8%) and unfamiliarity with software (384/1607, 23.9%), which further worsened the experience of the physical course. Furthermore, lack of interaction with the teacher (481/1607, 29.9%), lack of self-control (608/1607, 37.8%), and

inability to keep up with the lesson (73/1607, 4.5%) could be observed in the respondents, which was frustrating for the college students. However, Soffer's study [35] suggested that in many aspects of the examined effectiveness, web-based education was at least as effective as a face-to-face course [35]. We suppose that web-based physical education is promising if substantial improvements are made, such as technology support, optional exercises, and accommodation of students' preferences. According to Chekroud's study [27], all exercise types were significantly associated with lower mental health burden; numerous types of exercises could be chosen for web-based physical education. Considering the access to sports facilities and available sports equipment, only a few sports were suggested for web-based physical education, including shape-up exercises, a designed combination of exercises, Chinese kung fu, and rhythmic sport, and table tennis; meanwhile, ordinary physical courses could not be conducted properly, such as ball games and athletic events. Improving the physical fitness of college students with limited sports is the major issue of web-based physical education; to address this issue, we may learn from other courses and search for help from other fields, such as virtual reality [36].

### Limitations

There are several limitations of our study. First, due to anonymity and confidentiality requirements, a prospective study

could not be performed through the web-based snowball sampling survey, and the respondents to the survey may not be a representative sample of Chinese students. Second, mental health was evaluated by the DASS-21 scale instead of by mental health professionals; floor effects could not be excluded, although both methods of evaluation are based on the respondents' feelings and self-reporting. Third, due to the inherent nature of a cross-sectional study, we could only verify the association between sports-related lifestyle and mental health and could not verify the causal relationship. Finally, the assessment of sports-related lifestyle variables depends on non-standardized questions that have not been validated.

### Conclusions

The mental status of most of the college students in Wuhan who responded to our survey was normal. The mental status of the students was significantly correlated with regular exercise and sufficient exercise duration. Therefore, professional physical guidance is needed for college students as well as the general population. Considering the restrictions on sports facilities and equipment, selected sports were suggested for web-based physical education to improve physical fitness. However, web-based physical education is still far from satisfactory. Exercises not meeting students' preferences, frequent technical problems, and distant interactions are the main problems that should be solved in future.

### Acknowledgments

This research was supported by Hubei Province Health and Family Planning Scientific Research Project (WJ2018H0160) and the Wuhan Children's Hospital Foundation (2020FE002). The funding organizations had no role in the study design, data collection, data analysis, manuscript preparation, or publication decision. The work is the responsibility of the authors.

### Authors' Contributions

CD and WX conceived and designed the study. YG, XP, and JS contributed to the literature search. CD, JW, LZ, and HL collected the data. CD and WX analyzed the data. CD drafted the manuscript. WX revised the manuscript. WX had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors reviewed and approved the final version of the manuscript.

### Conflicts of Interest

None declared.

### References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* 2020 Feb;395(10223):470-473. [doi: [10.1016/s0140-6736\(20\)30185-9](https://doi.org/10.1016/s0140-6736(20)30185-9)]
2. Kraemer MUG, Yang C, Gutierrez B, Wu C, Klein B, Pigott DM, Open COVID-19 Data Working Group, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 2020 May 01;368(6490):493-497 [FREE Full text] [doi: [10.1126/science.abb4218](https://doi.org/10.1126/science.abb4218)] [Medline: [32213647](https://pubmed.ncbi.nlm.nih.gov/32213647/)]
3. Sahu P. Closure of Universities Due to Coronavirus Disease 2019 (COVID-19): Impact on Education and Mental Health of Students and Academic Staff. *Cureus* 2020 Apr 04;12(4):e7541 [FREE Full text] [doi: [10.7759/cureus.7541](https://doi.org/10.7759/cureus.7541)] [Medline: [32377489](https://pubmed.ncbi.nlm.nih.gov/32377489/)]
4. Moszkowicz D, Duboc H, Dubertret C, Roux D, Bretagnol F. Daily medical education for confined students during coronavirus disease 2019 pandemic: A simple videoconference solution. *Clin Anat* 2020 Sep 22;33(6):927-928 [FREE Full text] [doi: [10.1002/ca.23601](https://doi.org/10.1002/ca.23601)] [Medline: [32253771](https://pubmed.ncbi.nlm.nih.gov/32253771/)]
5. Gewin V. Five tips for moving teaching online as COVID-19 takes hold. *Nature* 2020 Apr;580(7802):295-296. [doi: [10.1038/d41586-020-00896-7](https://doi.org/10.1038/d41586-020-00896-7)] [Medline: [32210377](https://pubmed.ncbi.nlm.nih.gov/32210377/)]

6. Lovell GP, Nash K, Sharman R, Lane BR. A cross-sectional investigation of depressive, anxiety, and stress symptoms and health-behavior participation in Australian university students. *Nurs Health Sci* 2015 Mar 06;17(1):134-142. [doi: [10.1111/nhs.12147](https://doi.org/10.1111/nhs.12147)] [Medline: [24799077](https://pubmed.ncbi.nlm.nih.gov/24799077/)]
7. Schofield MJ, O'Halloran P, McLean SA, Forrester-Knauss C, Paxton SJ. Depressive Symptoms Among Australian University Students: Who Is at Risk? *Aust Psychol* 2016 Mar 18;51(2):135-144. [doi: [10.1111/ap.12129](https://doi.org/10.1111/ap.12129)]
8. Dogra S, MacIntosh L, O'Neill C, D'Silva C, Shearer H, Smith K, et al. The association of physical activity with depression and stress among post-secondary school students: A systematic review. *Mental Health and Physical Activity* 2018 Mar;14:146-156. [doi: [10.1016/j.mhpa.2017.11.001](https://doi.org/10.1016/j.mhpa.2017.11.001)]
9. Steptoe A, Butler N. Sports participation and emotional wellbeing in adolescents. *Lancet* 1996 Jun;347(9018):1789-1792. [doi: [10.1016/s0140-6736\(96\)91616-5](https://doi.org/10.1016/s0140-6736(96)91616-5)]
10. Gilat R, Cole BJ. COVID-19, Medicine, and Sports. *Arthrosc Sports Med Rehabil* 2020 Apr 11;2(3):e175-e176 [FREE Full text] [doi: [10.1016/j.asmr.2020.04.003](https://doi.org/10.1016/j.asmr.2020.04.003)] [Medline: [32292914](https://pubmed.ncbi.nlm.nih.gov/32292914/)]
11. Global monitoring of school closure caused by COVID-19. Webpage in Chinese. UNESECO. URL: <https://zh.unesco.org/themes/education-emergencies/coronavirus-school-closures> [accessed 2020-10-01]
12. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 2020 May 22;368(6493):860-868 [FREE Full text] [doi: [10.1126/science.abb5793](https://doi.org/10.1126/science.abb5793)] [Medline: [32291278](https://pubmed.ncbi.nlm.nih.gov/32291278/)]
13. Tran BX, Ha GH, Nguyen LH, Vu GT, Hoang MT, Le HT, et al. Studies of Novel Coronavirus Disease 19 (COVID-19) Pandemic: A Global Analysis of Literature. *Int J Environ Res Public Health* 2020 Jun 08;17(11):4095 [FREE Full text] [doi: [10.3390/ijerph17114095](https://doi.org/10.3390/ijerph17114095)] [Medline: [32521776](https://pubmed.ncbi.nlm.nih.gov/32521776/)]
14. Wang C, Cheng Z, Yue X, McAleer M. Risk Management of COVID-19 by Universities in China. *JRFM* 2020 Feb 19;13(2):36. [doi: [10.3390/jrfm13020036](https://doi.org/10.3390/jrfm13020036)]
15. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre R, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. *Brain Behav Immun* 2020 Jul;87:40-48 [FREE Full text] [doi: [10.1016/j.bbi.2020.04.028](https://doi.org/10.1016/j.bbi.2020.04.028)] [Medline: [32298802](https://pubmed.ncbi.nlm.nih.gov/32298802/)]
16. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Res* 2020 May;287:112934 [FREE Full text] [doi: [10.1016/j.psychres.2020.112934](https://doi.org/10.1016/j.psychres.2020.112934)] [Medline: [32229390](https://pubmed.ncbi.nlm.nih.gov/32229390/)]
17. Chen Q, Liang M, Li Y, Guo J, Fei D, Wang L, et al. Mental health care for medical staff in China during the COVID-19 outbreak. *Lancet Psychiatry* 2020 Apr;7(4):e15-e16. [doi: [10.1016/s2215-0366\(20\)30078-x](https://doi.org/10.1016/s2215-0366(20)30078-x)]
18. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. *Lancet* 2020 Feb 22;395(10224):e37-e38 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30309-3](https://doi.org/10.1016/S0140-6736(20)30309-3)] [Medline: [32043982](https://pubmed.ncbi.nlm.nih.gov/32043982/)]
19. Ko GT, Tang J, Chan JC, Sung R, Wu MM, Wai HP, et al. Lower BMI cut-off value to define obesity in Hong Kong Chinese: an analysis based on body fat assessment by bioelectrical impedance. *Br J Nutr* 2001 Feb;85(2):239-242. [doi: [10.1079/bjn2000251](https://doi.org/10.1079/bjn2000251)] [Medline: [11242492](https://pubmed.ncbi.nlm.nih.gov/11242492/)]
20. Norton PJ. Depression Anxiety and Stress Scales (DASS-21): psychometric analysis across four racial groups. *Anxiety Stress Coping* 2007 Sep;20(3):253-265. [doi: [10.1080/10615800701309279](https://doi.org/10.1080/10615800701309279)] [Medline: [17999228](https://pubmed.ncbi.nlm.nih.gov/17999228/)]
21. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. *Int J Environ Res Public Health* 2020 Mar 06;17(5):1729 [FREE Full text] [doi: [10.3390/ijerph17051729](https://doi.org/10.3390/ijerph17051729)] [Medline: [32155789](https://pubmed.ncbi.nlm.nih.gov/32155789/)]
22. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. *Lancet Psychiatry* 2020 Apr;7(4):300-302. [doi: [10.1016/s2215-0366\(20\)30073-0](https://doi.org/10.1016/s2215-0366(20)30073-0)]
23. Tan W, Hao F, McIntyre RS, Jiang L, Jiang X, Zhang L, et al. Is returning to work during the COVID-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of Chinese workforce. *Brain Behav Immun* 2020 Jul;87:84-92 [FREE Full text] [doi: [10.1016/j.bbi.2020.04.055](https://doi.org/10.1016/j.bbi.2020.04.055)] [Medline: [32335200](https://pubmed.ncbi.nlm.nih.gov/32335200/)]
24. Hao F, Tan W, Jiang L, Zhang L, Zhao X, Zou Y, et al. Do psychiatric patients experience more psychiatric symptoms during COVID-19 pandemic and lockdown? A case-control study with service and research implications for immunopsychiatry. *Brain Behav Immun* 2020 Jul;87:100-106 [FREE Full text] [doi: [10.1016/j.bbi.2020.04.069](https://doi.org/10.1016/j.bbi.2020.04.069)] [Medline: [32353518](https://pubmed.ncbi.nlm.nih.gov/32353518/)]
25. Chew NWS, Lee GKH, Tan BYQ, Jing M, Goh Y, Ngiam NJH, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. *Brain Behav Immun* 2020 Aug;88:559-565 [FREE Full text] [doi: [10.1016/j.bbi.2020.04.049](https://doi.org/10.1016/j.bbi.2020.04.049)] [Medline: [32330593](https://pubmed.ncbi.nlm.nih.gov/32330593/)]
26. Brière FN, Yale-Soulière G, Gonzalez-Sicilia D, Harbec M, Morizot J, Janosz M, et al. Prospective associations between sport participation and psychological adjustment in adolescents. *J Epidemiol Community Health* 2018 Jul 14;72(7):575-581. [doi: [10.1136/jech-2017-209656](https://doi.org/10.1136/jech-2017-209656)] [Medline: [29540466](https://pubmed.ncbi.nlm.nih.gov/29540466/)]
27. Chekroud SR, Gueorguieva R, Zheutlin AB, Paulus M, Krumholz HM, Krystal JH, et al. Association between physical exercise and mental health in 1.2 million individuals in the USA between 2011 and 2015: a cross-sectional study. *Lancet Psychiatry* 2018 Sep;5(9):739-746. [doi: [10.1016/S2215-0366\(18\)30227-X](https://doi.org/10.1016/S2215-0366(18)30227-X)]

28. Schuch FB, Vancampfort D, Firth J, Rosenbaum S, Ward PB, Silva ES, et al. Physical Activity and Incident Depression: A Meta-Analysis of Prospective Cohort Studies. *Am J Psychiatry* 2018 Jul 01;175(7):631-648. [doi: [10.1176/appi.ajp.2018.17111194](https://doi.org/10.1176/appi.ajp.2018.17111194)] [Medline: [29690792](https://pubmed.ncbi.nlm.nih.gov/29690792/)]
29. Pietrobelli A, Pecoraro L, Ferruzzi A, Heo M, Faith M, Zoller T, et al. Effects of COVID-19 Lockdown on Lifestyle Behaviors in Children with Obesity Living in Verona, Italy: A Longitudinal Study. *Obesity (Silver Spring)* 2020 Aug 30;28(8):1382-1385 [FREE Full text] [doi: [10.1002/oby.22861](https://doi.org/10.1002/oby.22861)] [Medline: [32352652](https://pubmed.ncbi.nlm.nih.gov/32352652/)]
30. Sallis JF, McKenzie TL, Beets MW, Beighle A, Erwin H, Lee S. Physical Education's Role in Public Health: Steps Forward and Backward Over 20 Years and HOPE for the Future. *Res Quart Exerc Sport* 2012 Jun 01;83(2):125-135. [doi: [10.5641/027013612800745329](https://doi.org/10.5641/027013612800745329)]
31. Hills AP, Dengel DR, Lubans DR. Supporting public health priorities: recommendations for physical education and physical activity promotion in schools. *Prog Cardiovasc Dis* 2015;57(4):368-374. [doi: [10.1016/j.pcad.2014.09.010](https://doi.org/10.1016/j.pcad.2014.09.010)] [Medline: [25269062](https://pubmed.ncbi.nlm.nih.gov/25269062/)]
32. Ho CS, Chee CY, Ho RC. Mental Health Strategies to Combat the Psychological Impact of COVID-19 Beyond Paranoia and Panic. *Ann Acad Med Singap* 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: [32200399](https://pubmed.ncbi.nlm.nih.gov/32200399/)]
33. Tran BX, Dang AK, Thai PK, Le HT, Le XTT, Do TTT, et al. Coverage of Health Information by Different Sources in Communities: Implication for COVID-19 Epidemic Response. *Int J Environ Res Public Health* 2020 May 20;17(10):3577 [FREE Full text] [doi: [10.3390/ijerph17103577](https://doi.org/10.3390/ijerph17103577)] [Medline: [32443712](https://pubmed.ncbi.nlm.nih.gov/32443712/)]
34. Tran BX, Phan HT, Nguyen TPT, Hoang MT, Vu GT, Thi Lei H, et al. Reaching further by Village Health Collaborators: The informal health taskforce of Vietnam for COVID-19 responses. *J Glob Health* 2020 Jun;10(1):010354 [FREE Full text] [doi: [10.7189/jogh.10.010354](https://doi.org/10.7189/jogh.10.010354)] [Medline: [32509285](https://pubmed.ncbi.nlm.nih.gov/32509285/)]
35. Soffer T, Nachmias R. Effectiveness of learning in online academic courses compared with face-to-face courses in higher education. *J Comput Assist Learn* 2018 Mar 22;34(5):534-543. [doi: [10.1111/jcal.12258](https://doi.org/10.1111/jcal.12258)]
36. Evans DJ, Bay BH, Wilson TD, Smith CF, Lachman N, Pawlina W. Going Virtual to Support Anatomy Education: A STOPGAP in the Midst of the Covid-19 Pandemic. *Anat Sci Educ* 2020 May 05;13(3):279-283. [doi: [10.1002/ase.1963](https://doi.org/10.1002/ase.1963)] [Medline: [32277598](https://pubmed.ncbi.nlm.nih.gov/32277598/)]

## Abbreviations

**DASS-21:** Depression, Anxiety, and Stress Scale

*Edited by G Eysenbach; submitted 10.06.20; peer-reviewed by R Ho, C Müller; comments to author 23.07.20; revised version received 01.08.20; accepted 30.09.20; published 05.10.20.*

### *Please cite as:*

Deng CH, Wang JQ, Zhu LM, Liu HW, Guo Y, Peng XH, Shao JB, Xia W

Association of Web-Based Physical Education With Mental Health of College Students in Wuhan During the COVID-19 Outbreak: Cross-Sectional Survey Study

*J Med Internet Res* 2020;22(10):e21301

URL: <http://www.jmir.org/2020/10/e21301/>

doi: [10.2196/21301](https://doi.org/10.2196/21301)

PMID: [32997639](https://pubmed.ncbi.nlm.nih.gov/32997639/)

©Cheng-Hu Deng, Jing-Qiang Wang, Li-Ming Zhu, He-Wang Liu, Yu Guo, Xue-Hua Peng, Jian-Bo Shao, Wei Xia. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Original Paper

# Dynamic Panel Surveillance of COVID-19 Transmission in the United States to Inform Health Policy: Observational Statistical Study

James Francis Oehmke<sup>1</sup>, PhD; Charles B Moss<sup>2</sup>, PhD; Lauren Nadya Singh<sup>1</sup>, MPH; Theresa Bristol Oehmke<sup>3</sup>, MS; Lori Ann Post<sup>1</sup>, PhD

<sup>1</sup>Buehler Center for Health Policy and Economics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

<sup>2</sup>Food and Resource Economics Department, University of Florida, Gainesville, FL, United States

<sup>3</sup>Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA, United States

**Corresponding Author:**

Lori Ann Post, PhD

Buehler Center for Health Policy and Economics

Feinberg School of Medicine

Northwestern University

420 E Superior St

Chicago, IL,

United States

Phone: 1 203 980 7107

Email: [lori.post@northwestern.edu](mailto:lori.post@northwestern.edu)

## Abstract

**Background:** The Great COVID-19 Shutdown aimed to eliminate or slow the spread of SARS-CoV-2, the virus that causes COVID-19. The United States has no national policy, leaving states to independently implement public health guidelines that are predicated on a sustained decline in COVID-19 cases. Operationalization of “sustained decline” varies by state and county. Existing models of COVID-19 transmission rely on parameters such as case estimates or  $R_0$  and are dependent on intensive data collection efforts. Static statistical models do not capture all of the relevant dynamics required to measure sustained declines. Moreover, existing COVID-19 models use data that are subject to significant measurement error and contamination.

**Objective:** This study will generate novel metrics of speed, acceleration, jerk, and 7-day lag in the speed of COVID-19 transmission using state government tallies of SARS-CoV-2 infections, including state-level dynamics of SARS-CoV-2 infections. This study provides the prototype for a global surveillance system to inform public health practice, including novel standardized metrics of COVID-19 transmission, for use in combination with traditional surveillance tools.

**Methods:** Dynamic panel data models were estimated with the Arellano-Bond estimator using the generalized method of moments. This statistical technique allows for the control of a variety of deficiencies in the existing data. Tests of the validity of the model and statistical techniques were applied.

**Results:** The statistical approach was validated based on the regression results, which determined recent changes in the pattern of infection. During the weeks of August 17-23 and August 24-30, 2020, there were substantial regional differences in the evolution of the US pandemic. Census regions 1 and 2 were relatively quiet with a small but significant persistence effect that remained relatively unchanged from the prior 2 weeks. Census region 3 was sensitive to the number of tests administered, with a high constant rate of cases. A weekly special analysis showed that these results were driven by states with a high number of positive test reports from universities. Census region 4 had a high constant number of cases and a significantly increased persistence effect during the week of August 24-30. This change represents an increase in the transmission model  $R$  value for that week and is consistent with a re-emergence of the pandemic.

**Conclusions:** Reopening the United States comes with three certainties: (1) the “social” end of the pandemic and reopening are going to occur before the “medical” end even while the pandemic is growing. We need improved standardized surveillance techniques to inform leaders when it is safe to open sections of the country; (2) varying public health policies and guidelines unnecessarily result in varying degrees of transmission and outbreaks; and (3) even those states most successful in containing the pandemic continue to see a small but constant stream of new cases daily.



**KEYWORDS**

COVID-19; models; surveillance; reopening America; contagion; metrics; surveillance; health policy; public health

## Introduction

Without question, SARS-CoV-2, the novel coronavirus that causes COVID-19 [1,2], has resulted in an unprecedented pandemic in modern history with significant morbidity and mortality [3-7]. Although some countries have had success in controlling COVID-19 [8-11], others have encountered much difficulty [12-17], resulting in significant adverse outcomes [18-21]. Beyond the overall implications related to infection and death [9,11,22-29], the COVID-19 pandemic has a deleterious impact on the global economy [27,30,31], violence [32-37], mental health [38-43], and food security [44-47], and disproportionately affects vulnerable populations such as the elderly [48-52], the poor [53,54], and racial and ethnic minorities [55-61]. We must establish COVID-19 control through good policy [10,12,62-68]; unfortunately, different states have implemented various and inconsistent COVID-19 policies [67,69-76] in the absence of a national plan [77,78]. Without a COVID-19 vaccine [79-81], we need *systematic public health surveillance* [7,82-89] to inform policies and guidelines for COVID-19 control and prevention such as quarantines, social distancing, face masks, crowd control, and hygiene to prevent viral spread [90-98]. Good surveillance can safely inform our leaders when, how, and where our country can reopen [76,99-103].

According to Teutsch and Churchill [85], public health surveillance is the “systematic, ongoing assessment of the health of a community, based on the collection, interpretation, and use of health data and information. Surveillance provides information necessary for public health decision making” (pg 1). Surveillance does not rely on a single indicator; it depends on a variety of metrics to identify high-priority COVID-19 health events such as incidence, prevalence, mortality, severity, cost, preventability, and communicability [104]. We need to meet these objectives of a surveillance system to prevent infectious diseases [104]. The United States must address several public health surveillance objectives, specifically to detect outbreaks (eg, the distribution and spread of COVID-19) and evaluate control strategies [104]. A surveillance system also includes “the functional capacity for data collection and analysis, as well as the timely dissemination of data” (pg 1) [87]. To this end, our study aims to create novel, validated metrics of speed, acceleration, and jerk in COVID-19 transmission in the United States.

The Great COVID-19 Shutdown refers to the variety of “lockdown” [105] public health policies adopted by countries around the globe to prevent the further spread of COVID-19, ranging from strict and complete quarantines [106-108] to disorganized and piecemeal closures [105]. It worked in places when it was implemented properly and in a timely manner, such as China, South Korea, Singapore, and Vietnam [10,11,62,109]. Some countries eliminated COVID-19, defined as achieving zero new cases over 14 days, while others flattened the curve

[64,110-113]. Governments that failed to effectively close down public movement and interactions resulted in increases in SARS-CoV-2 infections [50,114-120]. The United States had no national policy and was late in responding to the looming pandemic [105]. In fact, COVID-19 was technically classified as an epidemic by the Centers for Disease Control and Prevention when it accounted for >7.3% of all deaths in the United States. According to the National Center for Health Statistics, this was reached during the week of March 29-April 4 when COVID-19 accounted for 13.87% of all causes of death [121].

In response to the large death toll exacted by the epidemic, states independently implemented public health guidelines [14,19,70,71,122-125] regarding closures, social distancing, masks, and hand hygiene, which begs the question: when is it safe to reopen [126]? Reopening guidelines are predicated on a sustained decline in COVID-19 cases; however, operationalization of “sustained decline” varies by state [127]. Existing contagion models for COVID-19 rely on parameters such as case estimates or  $R_0$  and use intensive data collection efforts [128,129]. “Static” statistical models do not capture all of the relevant dynamics required to measure sustained declines [130-135]. Moreover, existing COVID-19 models use data that are subject to significant measurement error and other contaminants. Estimates of new SARS-CoV-2 infections suffer from undercounts due to asymptomatic carriers [136,137], access to testing [9,138-140], testing delays [141], testing sensitivity and specificity [142-145], and access to health care [60,146-150]. Surveillance systems and any enumeration of COVID-19 cases will err on the side of severity, meaning the most severe cases are more likely to be captured, the consequence of which is a significant undercount [71,104,130,151-156].

The conventional approach to modeling the spread of diseases such as COVID-19 is to posit an underlying contagion model and then to seek accurate direct measurement of the model parameters such as effective transmission rates or other parameters, often through labor-intensive methods relying on contact tracing to determine the spread of the virus in a sample population. For viral epidemics with an incubation period of up to 14 days, it takes weeks if not months to generate accurate parameter estimates even for simple contagion models [130]. For example, Li et al [157] provided early estimates of contagion parameters for COVID-19 using Wuhan data from contact tracing and methods developed by Lipsitch [158] but with weak statistical properties. It estimated the serial interval distribution and  $R_0$  from only six pairs of cases. These models also rely on underlying assumptions about immunity, common propensity for infection, well-mixed populations, etc [159]. Improvements in the models typically focus on relaxing these assumptions, for example, disaggregating the population by geography and modeling within-geography and cross-geographical personal interactions [160]. For example, Martcheva [161] provides an

excellent dynamic analysis of a wide variety of contagion models and their possible dynamics. Unfortunately, the study had limited options for the statistical inference of parameter values from actual data.

In contrast, we take an empirical approach that focuses on statistical modeling of widely available empirical data such as the number of confirmed cases or the number of tests conducted that can inform estimates of the current value of critical parameters like the infection rate or reproduction rate. We explicitly recognize that the data generating process for the reported data contain an underlying contagion component, a political-economic component such as availability of accurate test kits, a social component such as how strongly people adhere to social-distancing and shelter-in-place policies, and a sometimes inaccurate data reporting process that may obscure the underlying contagion process. We therefore seek a statistical approach that can provide meaningful information despite the complex and sometimes obfuscating data generation process. Our approach is consistent with the principles of evidence-based medicine, including controlling for complex pathways that may include socioeconomic factors such as mediating variables, and policy recommendations “based on the best available knowledge, derived from diverse sources and methods” (pg S58) [162].

There are two primary advantages to this empirical approach. First, we can apply the empirical model relatively quickly to a short data set. This advantage stems from the panel nature of the model. We used US states as the cross-sectional variable, so that a week’s data from all US states provides a reasonable sample size. In addition to enabling parameter estimation early in a pandemic, using this property we tested to see if there has been a shift in the transmission or reproductive rates of the transmission process in the past week, that is, whether there is statistical evidence that the US pandemic is peaking.

The second advantage is that the approach directly measures and informs policy-relevant variables. For example, the White House issued guidance on reopening the US economy that depends on a decrease in the documented number of cases and in the proportion of positive test results over a 14-day period, among other criteria and considerations [163]. As noted above, the number and proportion of positive test results are the outcome of a data generating process that includes not just the underlying transmission process but a multitude of mediating factors as well as idiosyncrasies of the data collection and reporting process. We specifically modeled the number of positive test results in our empirical model, which provides evidence of direct use in policy dialog.

This study has two objectives: (1) to create a proof-of-concept COVID-19 surveillance system using the United States as a prototype for a global system; and (2) to validate novel surveillance metrics/techniques including speed, acceleration, and jerk to better inform public health leaders how the pandemic is spreading or changing course.

## Methods

### Overview of Methodology

First, we will provide standard surveillance metrics including new counts of SARS-CoV-2 infections, moving 7-day averages of SARS-CoV-2 infections, rates of SARS-CoV-2 infections per 100,000 population, new numbers of COVID-19 deaths, moving 7-day averages of COVID-19 deaths, and rates of COVID-19 deaths per 100,000 population plus testing and positive testing ratios. Standard surveillance metrics are useful and allow us to compare data even though standard techniques are limited to more severe cases and suffer from data contamination.

Second, to address these data limitations we will validate novel surveillance metrics of (1) speed, (2) acceleration, and (3) jerk (change in acceleration). The basic question we are trying to inform is: how are we doing this week relative to previous weeks? From a public health perspective, in the midst of a pandemic, we would like (at least) three affirmative responses: (1) there are fewer new cases per day this week than last week, (2) the number of new cases is declining from day to day, and (3) the day-to-day decline in the number of cases is even bigger this week than last week. Additionally, we would like some indicative information about significant shifts in how the pandemic is progressing — positive shifts could be the first indicators of the emergence of a new or recurrent hotspot, and positive shifts could be first indicators of successful public health policy.

This study derives indicators to inform the three questions specified in the study objective above. Next, we provide a regression-based decomposition of the indicators. While it is beyond the scope of this study to determine the underlying causes of the pandemic and its trajectory over time, we provide a decomposition into proximate contributory factors such as whether an acceleration is due to a “natural” progression of the pandemic (eg, due to an increasing infectious population) or to a shift in an underlying model parameter (eg, a parameter shift that could be associated with reopening, other health policy changes, a viral mutation, the end of summer vacation for K-12 schools, or other underlying causes). Other factors can affect acceleration by “shifting” the underlying parameters (eg, the virus can mutate to become more or less infectious, states can impose lockdowns, social pressures can encourage or discourage people from wearing masks and social distancing, etc). Therefore, we use the regression analysis to provide a decomposition of speed, acceleration, and jerk into proximate contributory factors. Finally, this study is an innovation over traditional agnostic surveillancesystems in that we go beyond presenting metrics of the transmission of COVID-19 by providing probable scenarios regarding the context in which the disease is spreading.

The COVID Tracking Project [164] compiles data from multiple state sources on the web [165]; data for the most recent 36 days were accessed from the GitHub repository [166]. After accounting for lagged and differenced regressors, this resulted in a panel of 50 states plus the District of Columbia with 29 days in each panel (n=1352). Following Oehmke et al [167],

an empirical difference equation was specified in which the number of positive cases in each state at each day is a function of the prior number of cases, the level of testing, and weekly shift variables that measure whether the contagion was growing faster, at the same rate, or slower compared to the previous weeks. This resulted in a dynamic panel model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in STATA/MP, version 16.1 (StataCorp LLC).

Arellano-Bond estimation of difference equations has several statistical advantages: (1) it allows for statistical examination of the model's predictive ability and the validity of the model specification; (2) it corrects for autocorrelation and heteroscedasticity; (3) it has good properties for data with a small number of time periods and large number of states; (4) it corrects for omitted variables issues and provides a statistical test of correction validity. With these advantages, the method is applicable to ascertaining and statistically validating changes in the evolution of the pandemic within a period of one week or less, such as changes in the reproduction rate [167-174].

### Speed: New Cases Per Day

The basic indicator of the pandemic's status on a given day is the number of new cases on that day. Since new cases per day is a rate (value per unit of time), we will adopt physics nomenclature and refer to this as the speed of the pandemic. This is consistent with heuristic descriptions of the pandemic as spreading rapidly (ie, a large number of new cases per day) or slowly (ie, a small number of new cases per day). The public health ideal is to bring the speed of the pandemic to zero.

We report the number of new cases for each state both as a number per day and as a number per 100,000 population per day (table and column references).

For mathematical formality, we write:

$$N_{it}$$

where we have suppressed the  $i$  subscript of the previous section. We will be reporting surveillance numbers for each state and for the District of Columbia.

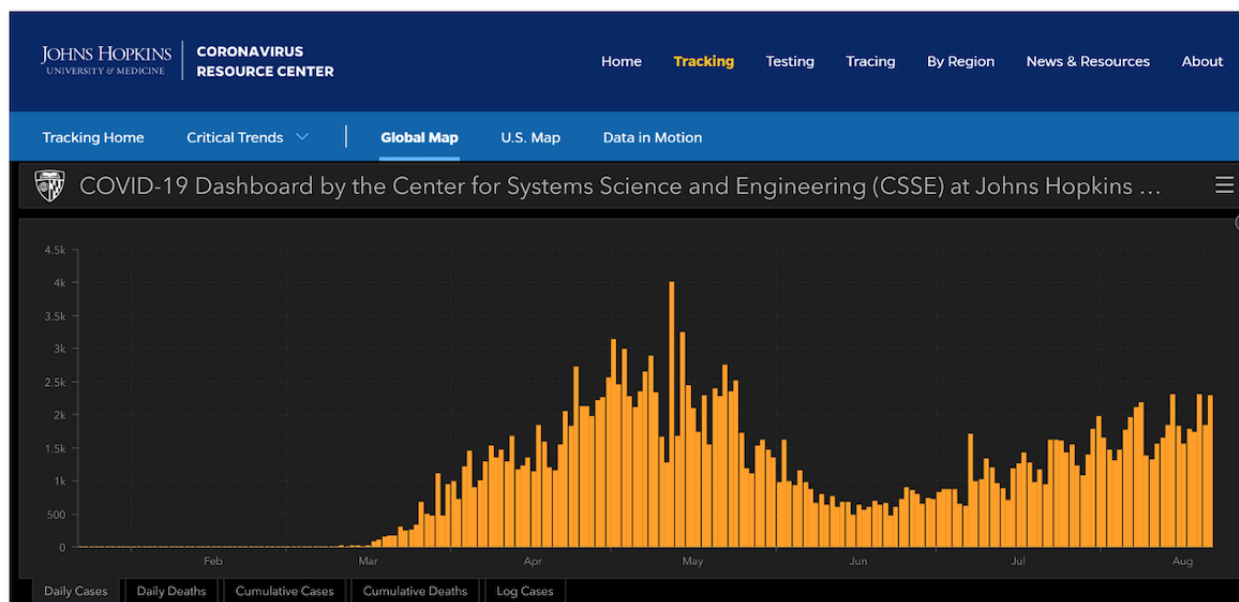
### Acceleration

We are also interested in whether the number of cases per day is increasing, peaking, or decreasing, and why. Again, we will adopt physics nomenclature and refer to this datum as the acceleration. Since acceleration is difficult to ascertain on a daily basis, and there are weekend effects, etc, in the data, we report the weekly average for the acceleration as:

$$D.N_{it}$$

where  $D.$  is the difference operator. A positive acceleration indicates an increasing number of cases per day, and a negative acceleration (deceleration) indicates a decreasing number of cases per day. An acceleration of 0 is indicative of a peak, valley, or inflection point depending in part on whether the previous acceleration was positive or negative. For example, acceleration in Illinois changed from positive to zero in mid-May, indicating a peak, and from negative to zero toward the end of June, indicating a valley (Figure 1) [69,175].

**Figure 1.** The number of positives per day in Illinois, according to the COVID-19 Dashboard of the Center for Systems Science and Engineering [175].



We provide a regression-based decomposition of accelerations into proximate components. That is, this is the systematic component of changes in acceleration, where  $t$  denotes the end date for the most recent week. Subtracting Oehmke et al's [167] equation (3) at time  $t-1$  from the same equation at time  $t$  results in:

$$D.N_{it}$$

where we have suppressed the error terms and added a term for the "weekend effect." We refer to the term containing  $Pos_{t-1}$  as the 1-day persistence effect. This, in turn, comprises a natural progression effect measured by  $\beta_0 D.Pos_{t-1}$  that represents the effect of a change over time in the number of new positive

results where the magnitude of the effect is calibrated at the prior week's parameter  $\beta_0$ , and a shift effect  $\beta_2 D.Pos_{t-1}$  that measures the effect of the week's shift in the parameter from  $\beta_0$  to  $\beta_0 + \beta_2$ . The second term in this equation is the 7-day persistence effect and is analogous to the 1-day persistence effect, including its decomposition into a natural progression effect  $y_0 D.Pos_{t-7}$  and a shift effect  $y_2 D.Pos_{t-7}$ . The next part of equation 2 represents the portion of acceleration that is composed of changes in the contemporaneous component of the model.

The analogous expression for 1 week prior and 2 weeks prior are:



The expression for 2 weeks prior,  $D.Pos_{t-14}$ , represents the baseline and does not contain any shift parameters. The shift parameters  $\beta_1$ ,  $\beta_2$ ,  $y_1$ , and  $y_2$  represent shifts in the most recent 2 weeks relative to the week ending at time  $t-14$ .

The expressions for  $D.Pos_t$  from equations 3 or 4 are easily adapted from time  $t$  to time  $t-j$  for each week and averaged over the week to provide a decomposition of acceleration as defined by equation 2.

### Jerk: The Change in Acceleration

We now address the question of whether the day-to-day increase (or decrease) in new cases the current week is bigger or smaller than the day-to-day increase (or decrease) in new cases of the past week.

Formally, for the current week we are interested in is:



The first term to the right of the definitional equality is the average growth in the number of daily positive cases for the current week ending at time  $t$ , and the second term is the average increase in the number of daily positive cases for the prior week. Using physics nomenclature, the difference between these two acceleration rates is the "jerk." A positive jerk indicates that the acceleration in the number of daily cases this week is greater than the average growth last week. Such a finding would be

consistent with a scenario in which the pandemic was experiencing explosive growth; where a policy shift such as reopening had augmented the acceleration of the pandemic, possibly including a shift from deceleration to acceleration; or where a megaevent had "jerked" the acceleration upward, among other scenarios.

Using equation four, for the most recent week ending at time  $t$ , we can write:



The top row contains the 1-day persistence effect's contribution to the jerk. The first term on the right side of the equation represents the natural progression of the 1-day persistence effect on acceleration due to changes across weeks in the daily change in the number of new cases per day. The last term in the first row represents structural shifts in the 1-day persistence effect. The second row is analogous to the first row, except that it represents the 7-day persistence effect's contribution to the jerk. The third row represents the contribution of contemporaneous effects to the jerk.

The analogous equation for the prior week is:



Equations 6 and 7 are easily averaged over the week to provide a decomposition of jerk as defined by equation 5.

## Results

### Regional Regression

#### Findings

We group the states according to Census region and present regression results for each region below. The biweekly surveillance products will be based on these regressions.

For Region 1 (Northeast), the regression Wald statistic shows that the model was statistically significant ( $\chi^2_{10}=132$ ,  $P<.001$ ), and the Sargan test fails to reject the validity of the overidentifying restrictions. ( $\chi^2_{252}=258$ ,  $P=.38$ ) (Table 1).



**Table 1.** Arellano-Bond dynamic panel data modeling of the number of daily infections reported by state, August 2-30, 2020.

Variable	Region 1		Region 2		Region 3		Region 4	
	Coefficient	P value	Coefficient	P value	Coefficient	P value	Coefficient	P value
L1Pos	0.084	.22	−0.102	.02	−0.012	.77	0.273	<.001
L1shiftAug17	−0.129	.16	0.069	.24	0.221	.02	−0.213	.03
L1shiftAug24	−0.112	.19	0.093	.10	−0.021	.84	−0.737	<.001
L7Pos	0.151	.02	0.288	<.001	0.269	<.001	0.006	.93
L7shiftAug17	−0.014	.87	−0.024	.67	−0.334	<.001	0.018	.79
L7shiftAug24	0.004	.96	0.046	.49	−0.265	.003	0.397	.02
Tests	0.003	.12	0.047	<.001	0.091	<.001	0.017	.048
Tests_squared	7.12E-09	.61	−5.05E-07	<.001	−4.89E-07	<.001	2.74E-08	.48
Tests_per_10K	1.072	.04	8.023	.002	−15.986	<.001	— <sup>a</sup>	—
Weekend	−14.751	.20	23.581	.28	−33.948	.58	51.977	.34
Constant	124.637	<.001	46.461	.89	429.167	<.001	397.678	<.001
Wald statistic for regression	$\chi^2_{10}=132$	<.001	$\chi^2_{10}=590$	<.001	$\chi^2_{10}=475$	<.001	$\chi^2_{10}=316$	<.001
Sargan statistic for validity	$\chi^2_{252}=258$	.38	$\chi^2_{338}=373$	.09	$\chi^2_{483}=446$	.89	$\chi^2_{368}=370$	.46

<sup>a</sup>Region 4 did not include the *Tests\_per\_100K* variable due to collinearity.

The coefficient on the first lag of the dependent variable is not statistically significant, nor are the shift parameters for the weeks of August 17 and August 24 for this coefficient. The coefficient on the 7th lag is positive and statistically significant (0.151,  $P=.02$ ). Neither of the shift parameters for the weeks of August 17 and August 24 are statistically significant. Of the variables representing the number of tests administered, the number per 100,000 population is significant (1.072,  $P=.04$ ). The weekend variable is not significant. The constant is positive and significant (124.637,  $P<.001$ ).

For Region 2 (Midwest), the regression Wald statistic shows that the model was statistically significant ( $\chi^2_{10}=590$ ,  $P<.001$ ), and the Sargan test fails to reject the validity of the overidentifying restrictions ( $\chi^2_{338}=373$ ,  $P=.09$ ).

The coefficient on the first lag of the dependent variable is not statistically significant. The shift for the week of August 17 for this coefficient is positive and statistically significant (0.221,  $P=.02$ ), but the shift for the week of August 24 is not significant. The coefficient on the 7th lag of the dependent variable is positive and significant (0.269,  $P<.001$ ). Neither of the weekly shift variables for this coefficient are significant. The tests, tests squared, and tests per 10,000 population are all statistically significant (0.047,  $P<.001$ ;  $-5.05E-07$ ,  $P<.001$ ; 8.023,  $P=.002$ ). Neither the weekend variable nor the constant are significant.

For Region 3 (South), the regression Wald statistic shows that the model was statistically significant ( $\chi^2_{10}=475$ ,  $P<.001$ ), and the Sargan test fails to reject the validity of the overidentifying restrictions ( $\chi^2_{483}=446$ ,  $P=.89$ ).

The coefficient on the first lag of the dependent variable is negative and statistically significant (−0.102,  $P=.02$ ). Neither

of the weekly shift variables for this coefficient are significant. The coefficient on the 7th lag of the dependent variable is positive and significant (0.288,  $P<.001$ ). The shifts for the weeks of August 17 and August 24 are negative and significant (−0.334,  $P<.001$ ; and −0.265,  $P=.003$ , respectively). The tests, tests squared, and tests per 10,000 population are all statistically significant (0.091,  $P<.001$ ;  $-4.89E-07$ ,  $P<.001$ ; and  $-15.986$ ,  $P<.001$ , respectively). The weekend variable is not significant. The constant is positive and significant (429.167,  $P<.001$ ).

For Region 4 (West), the regression Wald statistic shows that the model was statistically significant ( $\chi^2_{10}=316$ ,  $P<.001$ ), and the Sargan test fails to reject the validity of the overidentifying restrictions ( $\chi^2_{368}=370$ ,  $P=.46$ ).

The coefficient on the first lag of the dependent variable is negative and statistically significant (0.273,  $P<.001$ ). The shifts for the week of August 17 and August 24 for this coefficient are negative and statistically significant (−0.213,  $P=.03$ ; and −0.737,  $P<.001$ , respectively). The coefficient on the 7th lag of the dependent variable is not significant. The shift for the week of August 24 for this coefficient is positive and significant (0.397,  $P=.02$ ), but the shift for the week of August 17 is not significant. Of the test variables, only the coefficient on the number of tests administered is significant (0.017,  $P=.048$ ). The weekend variable is not significant. The constant is positive and significant (397.678,  $P<.001$ ).

### Interpretation

Region 1 appears to be fairly calm, with the only statistically significant persistence effect being a small 7-day lag effect. Region 2 is slightly less calm, but with a larger and statistically significant persistence effect and a noticeable positive effect of both the number of tests and the number of tests per 10,000.



Region 3 has the largest constant (average of state-specific effects) and the largest coefficient on tests, suggesting that the number of people newly tested for the virus is an important explanatory factor for the number of new cases. Region 4 has a high constant (average state-specific value) and significant shifts in both the 1-day and 7-day persistence values.

## University Reopenings

### Regression Results

A significant advantage of the panel data approach is that it can provide statistically valid quantifications of shifts in a fairly

short period such as 1 week. Perhaps the biggest pandemic issue during the week of August 24 was the high number of cases reported on university campuses as they reopened. We address this with an additional regression analysis. Six states in Region 3 reported 500 or more cases; at least one other university in these states reported 200 or more cases (Alabama, Florida, Georgia, North Carolina, South Carolina, and Texas). To inform this university effect, we split Region 3 into two groups of states—one with a high prevalence of university COVID-19 positives (denoted as group 3a) and another comprising the remaining Region 3 states (denoted as group 3b)—and then ran the regression analysis on the two groups (Table 2).

**Table 2.** Arellano-Bond dynamic panel data modeling of the number of daily infections reported by states in Region 3, grouped by the university effect, August 2-30, 2020.

Variable	Group 3a (with university effect)		Group 3b (without university effect)	
	Coefficient	P value	Coefficient	P value
L1Pos	−0.023	.76	0.037	.34
L1shiftAug17	0.249	.13	−0.029	.74
L1shiftAug24	−0.075	.69	0.005	.95
L7Pos	0.268	<.001	0.213	<.001
L7shiftAug17	−0.364	.007	−0.100	.22
L7shiftAug24	−0.252	.12	0.092	.26
Tests	0.121	<.001	0.029	<.001
Tests_squared	−6.59E-09	<.001	−5.61E-07	<.001
Tests_per_10K	−39.704	.005	−4.402	.06
Constant	910.482	.008	245.307	<.001
Wald statistic for regression	$\chi^2_9=169$	<.001	$\chi^2_9=491$	<.001
Sargan statistic for validity	$\chi^2_{165}=149$	.81	$\chi^2_{310}=301$	.63

For each group, the Wald statistic shows that the model was statistically significant ( $\chi^2_9=169$ ,  $P<.001$ ; and  $\chi^2_9=491$ ,  $P<.001$ , respectively), and the Sargan test fails to reject the validity of the overidentifying restrictions ( $\chi^2_{165}=149$ ,  $P=.81$ ;  $\chi^2_{310}=301$ ,  $P=.63$ ).

Without belaboring the individual coefficients, there are two important differences between the two groups. First is the coefficient on *Tests*, which numerically is the most important of the three test coefficients; group 3a (0.121,  $P<.001$ ) is more than four times the size of the coefficient for group 3b (0.029,  $P<.001$ ). The second important difference is that the constant for group a is more than three times the size of the constant for group b.

### Interpretation

The larger coefficient on *Tests* means that a higher percentage of tests are associated with positive results, possibly as large as 10% for group a (considering only the linear term). The larger value of the constant (which is an average of state-specific effects) means that there are larger state-specific risk factors, possibly related to the degree of “lockdown” and social compliance with recommendations such as social distancing or wearing masks. Coupling these two effects suggests that for the

week of August 24, the university effect is mostly due to increases in the number of asymptomatic students who got tested for the first time as they returned to university. This is consistent with the comparison of regional results across regions. It also suggests that the following week may be much worse if a significant fraction of the students are infectious and fail to practice social distancing, etc, thereby infecting others, who will likely show up in that week’s numbers.

These results may also help to explain spikes in other states, such as Iowa, Kansas, North Dakota, and South Dakota (which is also potentially affected by the Sturgis Motorcycle Rally), which all had significant numbers of COVID-19 cases at universities.

### Surveillance Results

Surveillance results are presented in Tables 3 and 4. The seven data elements in this proof-of-concept surveillance system are calculated as weekly averages and the speed, acceleration, and jerk are normalized per 100,000 population to compare the transmission of COVID-19 from week to week. These surveillance system data elements include (1) average weekly number of daily tests; (2) average weekly number of daily tests per 100,000 population; (3) average weekly number of daily positive tests; (4) average weekly number of daily positive tests

per 100,000 population referred to as *speed*; (5) weekly average of day-to-day change in the number of positives per day per 100,000 population, referred to as *acceleration*; (6) change in acceleration, referred to as *jerk*, which is the acceleration in the current week minus the acceleration in the prior week; a sustained positive jerk is typically associated with explosive growth; and finally, (7) the 7-day lag, which is the number of new cases of COVID-19 reported today per 100,000 population (ie, today's speed) that are associated with new cases reported 7 days ago (ie, last week's speed), and measures how much the increase in speed from last week persists into this week. Data are presented according to US Census regions. Data element 1 is reported as a number while 2-7 are reported as a rate, which better allows for comparison between US states.

The innovation of this study is the novel metrics we derived to measure how COVID-19 is spreading and changes in terms of transmission rates. These measures should be considered in combination with traditional static numbers including transmission rates and death rates. These novel metrics measure how fast the rates are changing, accounting for their data limitations.

As an example, we tracked the transmission of COVID-19 for the state of Illinois for the week from August 17 to 23, 2020. Illinois had a weekly average of 48,181 COVID-19 tests daily, also expressed as a weekly average of 380 tests per 100,000 population per day. Illinois had a weekly average of 2026 positive tests per day. The speed of the COVID-19 transmission is measured as an increase of 15.99 persons infected per 100,000 population per day. For the week of August 17 to 23 in Illinois, COVID-19 acceleration was 0.37, which means that every day there were .37 more new cases per 100,000 than the day before, or 2.6 more cases per day per 100,000 over the course of the

week. The jerk is 0.17, which means that acceleration was increasing: this increased acceleration accounted for 1.4 of the 2.6 additional cases per day per 100,000. Finally, the 7-day lag effect for speed is 3.58, which means that persistence or echo effects accounted for 3.58 or 22% of the 15.99 new daily positive cases per 100,000, which indicates an important but moderate persistence or echo effect for the week of August 17.

We see significant differences in COVID-19 transmission the following week (August 24-30, 2020). Illinois experienced a decrease in weekly average tests to 44,719 daily COVID-19 tests, also expressed as a weekly average of 353 tests per 100,000 population per day. This is 27 fewer tests per 100,000 population from last week. Illinois had a weekly average of 1923 positive tests per day, a decrease from the prior week, also expressed as a speed of 15.18 persons newly infected per day per 100,000 population. During the week of August 24-30, the acceleration decreased from the previous week to 0.11 and the jerk was negative (−0.26), indicating a leveling off of growth in new cases. Finally, the 7-day lag effect on speed is 5.35, which means that the persistence or echo effects accounted for 5.35 or over one-third of the 15.18 new daily positive cases per 100,000. The increased importance of echo effects rather than new cases from other (new) causes is consistent with a leveling off of COVID-19 growth in Illinois during the week of August 24-30.

In summary, the week of August 17-23 showed an increasing COVID-19 speed with positive acceleration and jerk. The week of August 24-30 exhibited a moderation in speed with lower acceleration and negative jerk. This is indicative of a leveling off or an inflection point: the pandemic in Illinois may be starting to decline, or this could be simply a pause before a continued increase in COVID-19 speed.

**Table 3.** Surveillance metrics for the week of August 17-23, 2020.

State	Tests per day, n (weekly average)	Daily tests per 100K people, n (daily average for the week)	Positives, n (reported number of new positive test results or confirmed cases per day per 100K people, weekly average)	Speed, n (daily positives per 100K people, weekly average)	Acceleration (day-to-day change in the number of positives per day, weekly average, per 100K people)	Jerk (week-over-week change in acceleration, per 100K people)	7-day persistence effect on speed (number of new cases per day per 100K people)
<b>Region 1</b>							
CT	16,936	475	127	3.56	0.49	-0.16	0.32
ME	3068	228	24	1.78	-0.06	-0.15	0.18
MA	14,815	215	309	4.48	-0.63	-0.57	0.61
NH	1591	117	17	1.25	0.07	0.13	0.23
NJ	22,687	255	291	3.28	0.39	0.93	0.59
NY	78,995	406	604	3.11	-0.03	-0.09	0.47
PA	13,737	107	655	5.12	-0.05	0.07	0.86
RI	5884	555	107	10.10	0.18	0.16	1.12
VT	1273	204	6	0.96	-0.05	-0.07	0.18
<b>Region 2</b>							
IL	48,181	380	2026	15.99	0.37	0.17	3.58
IN	10,136	151	788	11.71	-0.26	0.38	3.41
IA	4398	139	550	17.42	-0.43	-1.01	4.35
KS	4654	160	594	20.40	6.63	-2.55	4.20
MI	30,346	304	650	6.51	0.41	0.49	2.09
MN	9467	168	633	11.23	-0.06	0.09	2.85
MO	9888	161	1086	17.69	-0.61	-3.11	6.20
NE	2498	129	220	11.37	-0.73	-1.56	3.89
ND	1584	208	184	24.16	-0.06	-1.09	4.91
OH	22035	189	931	7.96	0.03	0.35	2.40
SD	1141	129	143	16.18	-0.24	-0.69	2.85
WI	8511	146	708	12.17	-0.54	-0.73	3.51
<b>Region 3</b>							
AL	10,749	219	947	19.31	-0.95	5.15	-1.33
AR	6236	207	558	18.50	-1.41	1.77	-1.12
DE	1637	168	63	6.51	0.18	0.44	-0.83
DC	3313	469	53	7.49	-0.10	0.69	-0.61
FL	28,001	130	3879	18.06	-0.54	1.09	-1.74
GA	23,802	224	2417	22.76	-0.18	1.58	-1.77
KY	5339	119	602	13.47	2.81	2.87	-0.89
LA	15,107	325	718	15.44	0.13	4.65	-1.29
MD	12927	214	556	9.19	0.14	1.09	-0.72
MS	2015	68	823	27.64	1.18	1.88	-1.53
NC	21,975	210	1452	13.84	0.31	0.59	-0.77
OK	8220	208	689	17.41	0.08	-0.13	-1.11
SC	6362	124	784	15.24	0.22	1.22	-1.08

State	Tests per day, n (weekly average)	Daily tests per 100K people, n (daily average for the week)	Positives, n (re- ported number of new positive test results or con- firmed cases per day per 100K people, weekly average)	Speed, n (daily positives per 100K people, weekly average)	Acceleration (day-to-day change in the number of posi- tives per day, weekly average, per 100K people)	Jerk (week- over-week change in accel- eration, per 100K people)	7-day persistence ef- fect on speed (number of new cases per day per 100K people)
TN	26,836	393	1461	21.40	-0.22	0.12	-1.48
TX	32,712	113	5994	20.67	-1.16	-2.07	-1.56
VA	16,720	196	897	10.51	-0.07	-0.14	-0.71
WV	5836	338	101	5.85	-0.17	0.03	-0.46
<b>Region 4</b>							
AK	3704	506	71	9.73	-0.82	-0.94	0.28
AZ	8414	116	652	8.96	-1.32	-1.46	0.31
CA	106,128	269	6015	15.22	-0.40	-0.22	0.59
CO	10,060	175	292	5.07	-0.01	0.32	0.15
HI	2412	170	219	15.45	0.02	-0.49	0.36
ID	2008	112	312	17.47	-0.09	2.06	0.57
MT	1261	118	97	9.07	-0.51	-0.88	0.26
NV	3824	124	614	19.92	-0.77	-0.24	0.57
NM	5696	272	143	6.81	0.44	0.50	0.19
OR	4432	105	239	5.67	-0.06	0.00	0.16
UT	3758	117	352	10.98	-0.13	0.07	0.27
WA	11,587	152	419	5.50	-0.17	-0.08	0.17
WY	685	118	42	7.23	-0.57	-1.11	0.14

**Table 4.** Surveillance metrics for the week of August 24-30, 2020.

State	Tests per day, n (weekly average)	Daily tests per 100K people, n (daily average for the week)	Positives, n (reported number of new positive test results or confirmed cases per day per 100K people, weekly average)	Speed, n (daily positives per 100K people, weekly average)	Acceleration (day-to-day change in the number of positives per day, weekly average, per 100K people)	Jerk (week-over-week change in acceleration, per 100K people)	7-day persistence effect on speed (number of new cases per day per 100K people)
<b>Region 1</b>							
CT	21,027	590	195	5.48	0.81	0.33	0.55
ME	3994	297	25	1.88	0.05	0.12	0.28
MA	24,300	353	410	5.95	0.41	1.04	0.69
NH	1890	139	21	1.54	-0.07	-0.15	0.19
NJ	26,762	301	302	3.40	0.05	-0.34	0.51
NY	82,233	423	623	3.20	0.09	0.12	0.48
PA	13,769	108	637	4.97	0.06	0.10	0.79
RI	4963	469	60	5.66	-1.19	-1.36	1.57
VT	1890	303	8	1.35	0.16	0.21	0.15
<b>Region 2</b>							
IL	44,719	353	1923	15.18	0.11	-0.26	5.35
IN	12,508	186	1054	15.66	0.56	0.82	3.92
IA	5017	159	921	29.18	1.95	2.38	5.83
KS	7366	253	838	28.78	7.63	1.00	6.82
MI	30,189	302	817	8.18	0.90	0.48	2.18
MN	8822	156	801	14.20	0.54	0.60	3.75
MO	8486	138	1226	19.97	1.51	2.11	5.92
NE	2798	145	282	14.57	1.20	1.93	3.80
ND	1297	170	261	34.23	1.46	1.52	8.08
OH	30,424	260	1066	9.12	0.35	0.32	2.66
SD	1270	144	292	33.04	3.86	4.10	5.41
WI	8464	145	728	12.50	0.22	0.76	4.07
<b>Region 3</b>							
AL	8485	173	1454	29.65	2.38	3.33	0.09
AR	6712	222	612	20.27	0.49	1.90	0.08
DE	1831	188	66	6.81	-0.98	-1.16	0.03
DC	3149	446	53	7.47	-0.45	-0.34	0.03
FL	24,425	114	3002	13.98	-0.26	0.28	0.08
GA	22,229	209	2146	20.21	-0.69	-0.51	0.10
KY	9483	212	643	14.40	-2.59	-5.40	0.06
LA	14,987	322	703	15.13	1.23	1.10	0.07
MD	12,335	204	527	8.72	-0.19	-0.34	0.04
MS	4988	168	683	22.95	0.10	-1.08	0.13
NC	23,543	224	1573	15.00	-0.57	-0.88	0.06
OK	7438	188	694	17.53	0.36	0.29	0.08
SC	7220	140	905	17.58	1.06	0.84	0.07



State	Tests per day, n (weekly average)	Daily tests per 100K people, n (daily average for the week)	Positives, n (reported number of new positive test results or confirmed cases per day per 100K people, weekly average)	Speed, n (daily positives per 100K people, weekly average)	Acceleration (day-to-day change in the number of positives per day, weekly average, per 100K people)	Jerk (week-over-week change in acceleration, per 100K people)	7-day persistence effect on speed (number of new cases per day per 100K people)
TN	20,545	301	1311	19.20	-2.13	-1.91	0.10
TX	36,669	126	4688	16.17	-0.28	0.87	0.09
VA	14,649	172	969	11.35	0.07	0.15	0.05
WV	4990	289	120	6.92	0.46	0.63	0.03
<b>Region 4</b>							
AK	2771	379	75	10.31	-0.25	0.57	3.92
AZ	6939	95	508	6.98	0.33	1.65	3.61
CA	98,685	250	5177	13.10	-0.26	0.14	6.14
CO	9257	161	308	5.35	-0.07	-0.06	2.05
HI	2536	179	255	17.99	0.25	0.23	6.23
ID	2435	136	288	16.11	0.00	0.09	7.04
MT	5130	480	130	12.17	0.48	0.99	3.66
NV	3065	100	472	15.34	-0.39	0.37	8.03
NM	6766	323	125	5.97	-0.48	-0.93	2.75
OR	4789	114	231	5.48	0.12	0.17	2.29
UT	4382	137	391	12.20	0.66	0.79	4.42
WA	11,760	154	380	4.99	1.80	1.96	2.22
WY	1486	257	34	5.95	0.00	0.57	2.92

## Discussion

### Principal Findings

The dynamic panel data model is a statistically validated analysis of reported COVID-19 transmissions and an important addition to the epidemiological toolkit for understanding the progression of the pandemic. It is important to recognize that surveillance systems require a variety of metrics. Systematic surveillance with standardized measures of decreases and increases in COVID-19 transmission coupled with health policies and guidelines add a critical tool to the epidemiologic arsenal to combat COVID-19.

The specific findings of the modeling exercise confirm that SARS-CoV-2 infection rates are persistent but changeable, and for most states increasing during the period between June 13-19, 2020. We find that for every 100 new COVID-19 cases from June 13-19, the following day would result in 26 new cases, meaning there is a significant reduction each day. However, it is important to recognize that this is an average across states and that state and local experiences will vary, which we measured. From June 20-26, on average in the United States, every 100 new cases on Monday was associated with 65 new cases on Tuesday, indicating the contagion increased 2.5-fold the rate from the prior week. The American pandemic has been ramping up in the past 2 weeks.

Remarkably, the US states diverged into three distinct patterns: (1) decline, (2) constant, and (3) increases consistent with outbreaks. In the 30 states with increasing cases, over the course of 2 weeks, there was a 3.6-fold increase in new infections while the states that had sustained declines in cases decreased by 2.5-fold. Again, these are averages among the three classifications of decline, constant, and increases, but these data could be further refined to show how much each state contributed to increases and decreases. Further investigation could usefully model state and local differences in infection rates, as well as ascertain quickly whether the pandemic will continue to re-emerge in the United States, or whether infection rates will reverse track and decline again even though states reopen.

The strengths of this study are the derived new metrics of the transmission of COVID-19. The limitation of this proof-of-concept surveillance system is that it includes only dynamic cases of COVID-19 infections; a full surveillance system should also include static cases. For example, Table 2 refers only to dynamic, new infections.

Based on the empirical evidence that our metrics of the COVID-19 contagion is a good standardization of increases and decreases for public health surveillance purposes, our future work will focus on the surveillance of 195 countries in eight global regions as defined by the World Bank. When possible,

we will provide subcountry-level metrics of the COVID-19 contagion beginning with US states and Canadian provinces. Our surveillance system will include estimates of speed,

acceleration, and jerk in acceleration along with traditional surveillance metrics.

## Conflicts of Interest

None declared.

## References

- Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents* 2020 Mar;55(3):105924 [FREE Full text] [doi: [10.1016/j.ijantimicag.2020.105924](https://doi.org/10.1016/j.ijantimicag.2020.105924)] [Medline: [32081636](https://pubmed.ncbi.nlm.nih.gov/32081636/)]
- Yuen K, Ye Z, Fung S, Chan C, Jin D. SARS-CoV-2 and COVID-19: The most important research questions. *Cell Biosci* 2020 Mar 16;10(1):40 [FREE Full text] [doi: [10.1186/s13578-020-00404-4](https://doi.org/10.1186/s13578-020-00404-4)] [Medline: [32190290](https://pubmed.ncbi.nlm.nih.gov/32190290/)]
- Kumari T, Shukla V. COVID-19: Towards confronting an unprecedented pandemic. *IJBI* 2020;02(01):01-10. [doi: [10.46505/ijbi.2020.2101](https://doi.org/10.46505/ijbi.2020.2101)]
- Yan Z. Unprecedented pandemic, unprecedented shift, and unprecedented opportunity. *Hum Behav Emerg Technol* 2020 Apr 06;2(2):110-112 [FREE Full text] [doi: [10.1002/hbe2.192](https://doi.org/10.1002/hbe2.192)] [Medline: [32427197](https://pubmed.ncbi.nlm.nih.gov/32427197/)]
- Liu Y, Liu B, Cui J. COVID-19 Evolves in Human Hosts. *Preprints* 2020:e [FREE Full text] [doi: [10.20944/preprints202003.0316.v1](https://doi.org/10.20944/preprints202003.0316.v1)]
- Velavan TP, Meyer CG. The COVID-19 epidemic. *Trop Med Int Health* 2020 Mar 16;25(3):278-280 [FREE Full text] [doi: [10.1111/tmi.13383](https://doi.org/10.1111/tmi.13383)] [Medline: [32052514](https://pubmed.ncbi.nlm.nih.gov/32052514/)]
- Heymann DL, Shindo N. COVID-19: what is next for public health? *The Lancet* 2020 Feb;395(10224):542-545. [doi: [10.1016/s0140-6736\(20\)30374-3](https://doi.org/10.1016/s0140-6736(20)30374-3)]
- Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. *Swiss Med Wkly* 2020 Mar 09;150:w20225 [FREE Full text] [doi: [10.4414/smww.2020.20225](https://doi.org/10.4414/smww.2020.20225)] [Medline: [32191813](https://pubmed.ncbi.nlm.nih.gov/32191813/)]
- Cohen J, Kupferschmidt K. Countries test tactics in 'war' against COVID-19. *Science* 2020 Mar 20;367(6484):1287-1288. [doi: [10.1126/science.367.6484.1287](https://doi.org/10.1126/science.367.6484.1287)] [Medline: [32193299](https://pubmed.ncbi.nlm.nih.gov/32193299/)]
- La V, Pham T, Ho M, Nguyen M, P. Nguyen K, Vuong T, et al. Policy Response, Social Media and Science Journalism for the Sustainability of the Public Health System Amid the COVID-19 Outbreak: The Vietnam Lessons. *Sustainability* 2020 Apr 07;12(7):2931. [doi: [10.3390/su12072931](https://doi.org/10.3390/su12072931)]
- Park S, Choi GJ, Ko H. Information Technology-Based Tracing Strategy in Response to COVID-19 in South Korea-Privacy Controversies. *JAMA* 2020 Jun 02;323(21):2129-2130. [doi: [10.1001/jama.2020.6602](https://doi.org/10.1001/jama.2020.6602)] [Medline: [32324202](https://pubmed.ncbi.nlm.nih.gov/32324202/)]
- Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health* 2020 Apr;8(4):e488-e496. [doi: [10.1016/s2214-109x\(20\)30074-7](https://doi.org/10.1016/s2214-109x(20)30074-7)]
- Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020 Apr 07;323(13):1239-1242. [doi: [10.1001/jama.2020.2648](https://doi.org/10.1001/jama.2020.2648)] [Medline: [32091533](https://pubmed.ncbi.nlm.nih.gov/32091533/)]
- Parodi SM, Liu VX. From Containment to Mitigation of COVID-19 in the US. *JAMA* 2020 Apr 21;323(15):1441-1442. [doi: [10.1001/jama.2020.3882](https://doi.org/10.1001/jama.2020.3882)] [Medline: [32167525](https://pubmed.ncbi.nlm.nih.gov/32167525/)]
- Watkins J. Preventing a covid-19 pandemic. *BMJ* 2020 Feb 28;368:m810. [doi: [10.1136/bmj.m810](https://doi.org/10.1136/bmj.m810)] [Medline: [32111649](https://pubmed.ncbi.nlm.nih.gov/32111649/)]
- Del Rio C, Malani PN. COVID-19-New Insights on a Rapidly Changing Epidemic. *JAMA* 2020 Apr 14;323(14):1339-1340. [doi: [10.1001/jama.2020.3072](https://doi.org/10.1001/jama.2020.3072)] [Medline: [32108857](https://pubmed.ncbi.nlm.nih.gov/32108857/)]
- Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, et al. COVID-19: towards controlling of a pandemic. *The Lancet* 2020 Mar;395(10229):1015-1018. [doi: [10.1016/s0140-6736\(20\)30673-5](https://doi.org/10.1016/s0140-6736(20)30673-5)]
- Cheng VCC, Wong S, Chen JHK, Yip CCY, Chuang VWM, Tsang OTY, et al. Escalating infection control response to the rapidly evolving epidemiology of the coronavirus disease 2019 (COVID-19) due to SARS-CoV-2 in Hong Kong. *Infect Control Hosp Epidemiol* 2020 May 05;41(5):493-498 [FREE Full text] [doi: [10.1017/ice.2020.58](https://doi.org/10.1017/ice.2020.58)] [Medline: [32131908](https://pubmed.ncbi.nlm.nih.gov/32131908/)]
- Spinelli A, Pellino G. COVID-19 pandemic: perspectives on an unfolding crisis. *Br J Surg* 2020 Jun 23;107(7):785-787 [FREE Full text] [doi: [10.1002/bjs.11627](https://doi.org/10.1002/bjs.11627)] [Medline: [32191340](https://pubmed.ncbi.nlm.nih.gov/32191340/)]
- Grech V. Unknown unknowns - COVID-19 and potential global mortality. *Early Hum Dev* 2020 May;144:105026 [FREE Full text] [doi: [10.1016/j.earlhumdev.2020.105026](https://doi.org/10.1016/j.earlhumdev.2020.105026)] [Medline: [32247898](https://pubmed.ncbi.nlm.nih.gov/32247898/)]
- Xu S, Li Y. Beware of the second wave of COVID-19. *The Lancet* 2020 Apr;395(10233):1321-1322. [doi: [10.1016/s0140-6736\(20\)30845-x](https://doi.org/10.1016/s0140-6736(20)30845-x)]
- Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, et al. Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). *JAMA Cardiol* 2020 Jul 01;5(7):811-818 [FREE Full text] [doi: [10.1001/jamacardio.2020.1017](https://doi.org/10.1001/jamacardio.2020.1017)] [Medline: [32219356](https://pubmed.ncbi.nlm.nih.gov/32219356/)]

23. Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations. World Health Organization. 2020 Mar. URL: <https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations> [accessed 2020-09-17]
24. Ather A, Patel B, Ruparel NB, Diogenes A, Hargreaves KM. Reply to "Coronavirus Disease 19 (COVID-19): Implications for Clinical Dental Care". J Endod 2020 Sep;46(9):1342 [FREE Full text] [doi: [10.1016/j.joen.2020.08.005](https://doi.org/10.1016/j.joen.2020.08.005)] [Medline: [32810475](https://pubmed.ncbi.nlm.nih.gov/32810475/)]
25. Connors J, Levy J. COVID-19 and its implications for thrombosis and anticoagulation. Blood 2020 Jun 04;135(23):2033-2040 [FREE Full text] [doi: [10.1182/blood.2020060000](https://doi.org/10.1182/blood.2020060000)] [Medline: [32339221](https://pubmed.ncbi.nlm.nih.gov/32339221/)]
26. Cao X. COVID-19: immunopathology and its implications for therapy. Nat Rev Immunol 2020 May 9;20(5):269-270 [FREE Full text] [doi: [10.1038/s41577-020-0308-3](https://doi.org/10.1038/s41577-020-0308-3)] [Medline: [32273594](https://pubmed.ncbi.nlm.nih.gov/32273594/)]
27. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: [10.1016/j.ijssu.2020.04.018](https://doi.org/10.1016/j.ijssu.2020.04.018)] [Medline: [32305533](https://pubmed.ncbi.nlm.nih.gov/32305533/)]
28. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: [10.1136/bmj.m1182](https://doi.org/10.1136/bmj.m1182)] [Medline: [32213507](https://pubmed.ncbi.nlm.nih.gov/32213507/)]
29. Chew M, Ong L, Koh F, Ng A, Tan Y, Ong BC. Guest post: Lessons in preparedness? The response to the COVID-19 pandemic by a surgical department in Singapore. BJS. 2020 Mar. URL: <https://cuttingedgeblog.com/2020/03/29/guest-post-lessons-in-preparedness-the-response-to-the-covid-19-pandemic-by-a-surgical-department-in-singapore/> [accessed 2020-09-17]
30. Barro R, Ursúa J, Weng J. The Coronavirus and the Great Influenza Pandemic: Lessons from the "spanish flu" for the coronavirus's potential effects on mortality and economic activity. National Bureau of Economic Research 2020 Apr:e [FREE Full text] [doi: [10.3386/w26866](https://doi.org/10.3386/w26866)]
31. Baker S, Bloom N, Davis S, Kost K, Sammon M, Viratyosin T. The unprecedented stock market impact of COVID-19. National Bureau of Economic Research 2020 Jun:e [FREE Full text] [doi: [10.3386/w26945](https://doi.org/10.3386/w26945)]
32. Bradbury-Jones C, Isham L. The pandemic paradox: The consequences of COVID-19 on domestic violence. J Clin Nurs 2020 Jul 22;29(13-14):2047-2049 [FREE Full text] [doi: [10.1111/jocn.15296](https://doi.org/10.1111/jocn.15296)] [Medline: [32281158](https://pubmed.ncbi.nlm.nih.gov/32281158/)]
33. Taub A. A New Covid-19 Crisis: Domestic Abuse Rises Worldwide. The New York Times. 2020 Apr. URL: <https://www.nytimes.com/2020/04/06/world/coronavirus-domestic-violence.html> [accessed 2020-09-17]
34. Usher K, Bhullar N, Durkin J, Gyamfi N, Jackson D. Family violence and COVID-19: Increased vulnerability and reduced options for support. Int J Ment Health Nurs 2020 Aug 07;29(4):549-552 [FREE Full text] [doi: [10.1111/inm.12735](https://doi.org/10.1111/inm.12735)] [Medline: [32314526](https://pubmed.ncbi.nlm.nih.gov/32314526/)]
35. Boserup B, McKenney M, Elkbuli A. Alarming trends in US domestic violence during the COVID-19 pandemic. Am J Emerg Med 2020 Apr 28:e [FREE Full text] [doi: [10.1016/j.ajem.2020.04.077](https://doi.org/10.1016/j.ajem.2020.04.077)] [Medline: [32402499](https://pubmed.ncbi.nlm.nih.gov/32402499/)]
36. COVID-19 and violence against women: what the health sector/system can do. World Health Organization. 2020 Apr. URL: <https://www.who.int/reproductivehealth/publications/vaw-covid-19/en/> [accessed 2020-09-17]
37. Leslie E, Wilson R. Sheltering in Place and Domestic Violence: Evidence from Calls for Service during COVID-19. SSRN Journal 2020 May:e [FREE Full text] [doi: [10.2139/ssrn.3600646](https://doi.org/10.2139/ssrn.3600646)]
38. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: [10.1056/nejmp2008017](https://doi.org/10.1056/nejmp2008017)]
39. Yao H, Chen J, Xu Y. Patients with mental health disorders in the COVID-19 epidemic. The Lancet Psychiatry 2020 Apr;7(4):e21. [doi: [10.1016/s2215-0366\(20\)30090-0](https://doi.org/10.1016/s2215-0366(20)30090-0)]
40. Holmes E, O'Connor R, Perry V, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. The Lancet Psychiatry 2020 Jun;7(6):547-560. [doi: [10.1016/S2215-0366\(20\)30168-1](https://doi.org/10.1016/S2215-0366(20)30168-1)]
41. Greenberg N, Docherty M, Gnanapragasam S, Wessely S. Managing mental health challenges faced by healthcare workers during covid-19 pandemic. BMJ 2020 Mar 26;368:m1211. [doi: [10.1136/bmj.m1211](https://doi.org/10.1136/bmj.m1211)] [Medline: [32217624](https://pubmed.ncbi.nlm.nih.gov/32217624/)]
42. Mental health and psychosocial considerations during the COVID-19 outbreak, 18 March 2020. World Health Organization. 2020 Mar. URL: <https://www.who.int/docs/default-source/coronaviruse/mental-health-considerations.pdf> [accessed 2020-09-17]
43. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. The Lancet Psychiatry 2020 Apr;7(4):300-302. [doi: [10.1016/s2215-0366\(20\)30073-0](https://doi.org/10.1016/s2215-0366(20)30073-0)]
44. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. The Lancet Public Health 2020 May;5(5):e243-e244. [doi: [10.1016/s2468-2667\(20\)30084-0](https://doi.org/10.1016/s2468-2667(20)30084-0)]
45. The Lancet. India under COVID-19 lockdown. The Lancet 2020 Apr;395(10233):1315. [doi: [10.1016/s0140-6736\(20\)30938-7](https://doi.org/10.1016/s0140-6736(20)30938-7)]
46. Gostin LO, Hodge JG, Wiley LF. Presidential Powers and Response to COVID-19. JAMA 2020 Apr 28;323(16):1547-1548. [doi: [10.1001/jama.2020.4335](https://doi.org/10.1001/jama.2020.4335)] [Medline: [32186661](https://pubmed.ncbi.nlm.nih.gov/32186661/)]

47. Coronavirus disease 2019 (COVID-19) Situation Report – 72. World Health Organization. 2020 Apr. URL: [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200401-sitrep-72-covid-19.pdf?sfvrsn=3dd8971b\\_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200401-sitrep-72-covid-19.pdf?sfvrsn=3dd8971b_2) [accessed 2020-09-17]
48. Shahid Z, Kalayanamitra R, McClafferty B, Kepko D, Ramgobin D, Patel R, et al. COVID-19 and Older Adults: What We Know. *J Am Geriatr Soc* 2020 May 20;68(5):926-929 [FREE Full text] [doi: [10.1111/jgs.16472](https://doi.org/10.1111/jgs.16472)] [Medline: [32255507](https://pubmed.ncbi.nlm.nih.gov/32255507/)]
49. Nikolich-Zugich J, Knox KS, Rios CT, Natt B, Bhattacharya D, Fain MJ. Correction to: SARS-CoV-2 and COVID-19 in older adults: what we may expect regarding pathogenesis, immune responses, and outcomes. *GeroScience* 2020 May 3;42(3):1013-1013. [doi: [10.1007/s11357-020-00193-1](https://doi.org/10.1007/s11357-020-00193-1)]
50. Jordan R, Adab P, Cheng K. Covid-19: risk factors for severe disease and death. *BMJ* 2020 Mar 26;368:m1198. [doi: [10.1136/bmj.m1198](https://doi.org/10.1136/bmj.m1198)] [Medline: [32217618](https://pubmed.ncbi.nlm.nih.gov/32217618/)]
51. Garnier-Crussard A, Forestier E, Gilbert T, Krolak-Salmon P. Novel Coronavirus (COVID-19) Epidemic: What Are the Risks for Older Patients? *J Am Geriatr Soc* 2020 May 12;68(5):939-940 [FREE Full text] [doi: [10.1111/jgs.16407](https://doi.org/10.1111/jgs.16407)] [Medline: [32162679](https://pubmed.ncbi.nlm.nih.gov/32162679/)]
52. Lloyd-Sherlock P, Ebrahim S, Geffen L, McKee M. Bearing the brunt of covid-19: older people in low and middle income countries. *BMJ* 2020 Mar 13;368:m1052. [doi: [10.1136/bmj.m1052](https://doi.org/10.1136/bmj.m1052)] [Medline: [32169830](https://pubmed.ncbi.nlm.nih.gov/32169830/)]
53. Ahmed F, Ahmed N, Pissarides C, Stiglitz J. Why inequality could spread COVID-19. *The Lancet Public Health* 2020 May;5(5):e240. [doi: [10.1016/s2468-2667\(20\)30085-2](https://doi.org/10.1016/s2468-2667(20)30085-2)]
54. Sumner A, Hoy C, Ortiz-Juarez E. Estimates of the Impact of COVID-19 on Global Poverty. United Nations University / UNU-WIDER 2020 Apr:1-14 [FREE Full text] [doi: [10.35188/unu-wider/2020/800-9](https://doi.org/10.35188/unu-wider/2020/800-9)]
55. Laurencin CT, McClinton A. The COVID-19 Pandemic: a Call to Action to Identify and Address Racial and Ethnic Disparities. *J Racial Ethn Health Disparities* 2020 Jun 18;7(3):398-402 [FREE Full text] [doi: [10.1007/s40615-020-00756-0](https://doi.org/10.1007/s40615-020-00756-0)] [Medline: [32306369](https://pubmed.ncbi.nlm.nih.gov/32306369/)]
56. Mogaji E. Financial Vulnerability During a Pandemic: Insights for Coronavirus Disease (COVID-19). *SSRN Journal* 2020:e. [doi: [10.2139/ssrn.3564702](https://doi.org/10.2139/ssrn.3564702)]
57. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. *JAMA* 2020 Jun 23;323(24):2466-2467. [doi: [10.1001/jama.2020.8598](https://doi.org/10.1001/jama.2020.8598)] [Medline: [32391864](https://pubmed.ncbi.nlm.nih.gov/32391864/)]
58. Khunti K, Singh A, Pareek M, Hanif W. Is ethnicity linked to incidence or outcomes of covid-19? *BMJ* 2020 Apr 20;369:m1548. [doi: [10.1136/bmj.m1548](https://doi.org/10.1136/bmj.m1548)] [Medline: [32312785](https://pubmed.ncbi.nlm.nih.gov/32312785/)]
59. Chowkwanyun M, Reed AL. Racial Health Disparities and Covid-19 — Caution and Context. *N Engl J Med* 2020 Jul 16;383(3):201-203. [doi: [10.1056/nejmp2012910](https://doi.org/10.1056/nejmp2012910)]
60. Yancy CW. COVID-19 and African Americans. *JAMA* 2020 May 19;323(19):1891-1892. [doi: [10.1001/jama.2020.6548](https://doi.org/10.1001/jama.2020.6548)] [Medline: [32293639](https://pubmed.ncbi.nlm.nih.gov/32293639/)]
61. Platt L, Warwick R. Are some ethnic groups more vulnerable to COVID-19 than others. Institute for Fiscal Studies / Nuffield Foundation. 2020 May. URL: <https://www.ifs.org.uk/inequality/chapter/are-some-ethnic-groups-more-vulnerable-to-covid-19-than-others/> [accessed 2020-06-19]
62. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA* 2020 May 19;323(19):1915-1923 [FREE Full text] [doi: [10.1001/jama.2020.6130](https://doi.org/10.1001/jama.2020.6130)] [Medline: [32275295](https://pubmed.ncbi.nlm.nih.gov/32275295/)]
63. Raoofi A, Takian A, Akbari Sari A, Olyaeemanesh A, Haghighi H, Aarabi M. COVID-19 Pandemic and Comparative Health Policy Learning in Iran. *Arch Iran Med* 2020 Apr 01;23(4):220-234. [doi: [10.34172/aim.2020.02](https://doi.org/10.34172/aim.2020.02)] [Medline: [32271594](https://pubmed.ncbi.nlm.nih.gov/32271594/)]
64. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* 2020 Mar;395(10228):931-934. [doi: [10.1016/s0140-6736\(20\)30567-5](https://doi.org/10.1016/s0140-6736(20)30567-5)]
65. Lee V, Chiew C, Khong W. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. *J Travel Med* 2020 May 18;27(3):e [FREE Full text] [doi: [10.1093/jtm/taaa039](https://doi.org/10.1093/jtm/taaa039)] [Medline: [32167146](https://pubmed.ncbi.nlm.nih.gov/32167146/)]
66. Nkengasong JN, Mankoula W. Looming threat of COVID-19 infection in Africa: act collectively, and fast. *The Lancet* 2020 Mar;395(10227):841-842. [doi: [10.1016/s0140-6736\(20\)30464-5](https://doi.org/10.1016/s0140-6736(20)30464-5)]
67. Dave D, Friedson A, Matsuzawa K, Sabia J. When Do Shelter-in-Place Orders Fight COVID-19 Best? Policy Heterogeneity Across States and Adoption Time. National Bureau of Economic Research 2020 Aug 03:e [FREE Full text] [doi: [10.1111/ecin.12944](https://doi.org/10.1111/ecin.12944)] [Medline: [32836519](https://pubmed.ncbi.nlm.nih.gov/32836519/)]
68. Loayza N, Pennings S. Macroeconomic policy in the time of COVID-19: A primer for developing countries. Washington, DC: World Bank; Mar 2020.
69. Adalja AA, Toner E, Inglesby TV. Priorities for the US Health Community Responding to COVID-19. *JAMA* 2020 Apr 14;323(14):1343-1344. [doi: [10.1001/jama.2020.3413](https://doi.org/10.1001/jama.2020.3413)] [Medline: [32125355](https://pubmed.ncbi.nlm.nih.gov/32125355/)]
70. Raifman J, Nocka K, Jones D, Bor J, Lipson S, Jay J, et al. COVID-19 US State Policy Database. Ann Arbor, MI: Inter-university Consortium for Political and Social Research; Sep 2020.
71. Fauci A, Lane H, Redfield R. Covid-19 - Navigating the Uncharted. *N Engl J Med* 2020 Mar 26;382(13):1268-1269 [FREE Full text] [doi: [10.1056/NEJMe2002387](https://doi.org/10.1056/NEJMe2002387)] [Medline: [32109011](https://pubmed.ncbi.nlm.nih.gov/32109011/)]



72. Haffajee RL, Mello MM. Thinking Globally, Acting Locally — The U.S. Response to Covid-19. *N Engl J Med* 2020 May 28;382(22):e75. [doi: [10.1056/nejmp2006740](https://doi.org/10.1056/nejmp2006740)]
73. Liu M, Thomadsen R, Yao S. Forecasting the Spread of COVID-19 under Different Reopening Strategies. *SSRN Journal* 2020:e. [doi: [10.2139/ssrn.3607977](https://doi.org/10.2139/ssrn.3607977)]
74. Hodge JG. Federal vs. State Powers in Rush to Reopen Amid the Coronavirus Pandemic. *SSRN Journal* 2020:e. [doi: [10.2139/ssrn.3587011](https://doi.org/10.2139/ssrn.3587011)]
75. Harris J. Reopening Under COVID-19: What to Watch For. *National Bureau of Economic Research* 2020:e [FREE Full text] [doi: [10.3386/w27166](https://doi.org/10.3386/w27166)]
76. Alhaery M. A COVID-19 Reopening Readiness Index: The Key to Opening up the Economy. *medRxiv* 2020:e [FREE Full text] [doi: [10.1101/2020.05.22.20110577](https://doi.org/10.1101/2020.05.22.20110577)]
77. Koonin L. Novel coronavirus disease (COVID-19) outbreak: Now is the time to refresh pandemic plans. *J Bus Contin Emer Plan* 2020 Jan 01;13(4):1-15. [Medline: [32160942](https://pubmed.ncbi.nlm.nih.gov/32160942/)]
78. Gostic K, Gomez A, Mummah R, Kucharski A, Lloyd-Smith J. Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *eLife* 2020;9:e55570 [FREE Full text] [doi: [10.7554/elife.55570](https://doi.org/10.7554/elife.55570)]
79. Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 Vaccines at Pandemic Speed. *N Engl J Med* 2020 May 21;382(21):1969-1973. [doi: [10.1056/nejmp2005630](https://doi.org/10.1056/nejmp2005630)]
80. Ahn D, Shin H, Kim M, Lee S, Kim H, Myoung J, et al. Current Status of Epidemiology, Diagnosis, Therapeutics, and Vaccines for Novel Coronavirus Disease 2019 (COVID-19). *J Microbiol Biotechnol* 2020 Mar 28;30(3):313-324 [FREE Full text] [doi: [10.4014/jmb.2003.03011](https://doi.org/10.4014/jmb.2003.03011)] [Medline: [32238757](https://pubmed.ncbi.nlm.nih.gov/32238757/)]
81. Miller A, Reandelar M, Fasciglione K, Roumenova V, Li Y, Otazu G. Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study. *medRxiv* 2020:e [FREE Full text] [doi: [10.1101/2020.03.24.20042937](https://doi.org/10.1101/2020.03.24.20042937)]
82. Hennessy CH, Moriarty DG, Zack MM, Scherr PA, Brackbill R. Measuring health-related quality of life for public health surveillance. *Public Health Rep* 1994;109(5):665-672 [FREE Full text] [Medline: [7938388](https://pubmed.ncbi.nlm.nih.gov/7938388/)]
83. Woodall WH. The Use of Control Charts in Health-Care and Public-Health Surveillance. *Journal of Quality Technology* 2018 Feb 05;38(2):89-104. [doi: [10.1080/00224065.2006.11918593](https://doi.org/10.1080/00224065.2006.11918593)]
84. Wharton M, Chorba T, Vogt R, Morse D, Buehler J. Case definitions for public health surveillance. *MMWR Recomm Rep* 1990 Oct 19;39(RR-13):1-43 [FREE Full text] [Medline: [2122225](https://pubmed.ncbi.nlm.nih.gov/2122225/)]
85. Teutsch SM, Churchill RE, editors. *Principles and Practice of Public Health Surveillance*. 2nd ed. New York, NY: Oxford University Press; 2000.
86. Lee LM, Teutsch SM, Thacker SB, St Louis ME, editors. *New York, NY: Oxford University Press*; 2010.
87. Thacker S, Berkelman R. Public health surveillance in the United States. *Epidemiol Rev* 1988;10(1):164-190. [doi: [10.1093/oxfordjournals.epirev.a036021](https://doi.org/10.1093/oxfordjournals.epirev.a036021)] [Medline: [3066626](https://pubmed.ncbi.nlm.nih.gov/3066626/)]
88. German R, Lee LM, Horan JM, Milstein B, Pertowski C, Waller MN, Guidelines Working Group Centers for Disease Control/Prevention (CDC). Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm Rep* 2001 Jul 27;50(RR-13):1-35; quiz CE1. [Medline: [18634202](https://pubmed.ncbi.nlm.nih.gov/18634202/)]
89. Declich S, Carter AO. Public health surveillance: historical origins, methods and evaluation. *Bull World Health Organ* 1994;72(2):285-304 [FREE Full text] [Medline: [8205649](https://pubmed.ncbi.nlm.nih.gov/8205649/)]
90. Mittal R, Ni R, Seo J. The flow physics of COVID-19. *J. Fluid Mech* 2020 May 01;894:e. [doi: [10.1017/jfm.2020.330](https://doi.org/10.1017/jfm.2020.330)]
91. Cheng KK, Lam TH, Leung CC. Wearing face masks in the community during the COVID-19 pandemic: altruism and solidarity. *The Lancet* 2020 Apr:e. [doi: [10.1016/s0140-6736\(20\)30918-1](https://doi.org/10.1016/s0140-6736(20)30918-1)]
92. Cheng VC, Wong S, Chuang VW, So SY, Chen JH, Sridhar S, et al. The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. *J Infect* 2020 Jul;81(1):107-114 [FREE Full text] [doi: [10.1016/j.jinf.2020.04.024](https://doi.org/10.1016/j.jinf.2020.04.024)] [Medline: [32335167](https://pubmed.ncbi.nlm.nih.gov/32335167/)]
93. Dalton C, Corbett S, Katelaris A. Pre-Emptive Low Cost Social Distancing and Enhanced Hygiene Implemented before Local COVID-19 Transmission Could Decrease the Number and Severity of Cases. *SSRN Journal* 2020:e. [doi: [10.2139/ssrn.3549276](https://doi.org/10.2139/ssrn.3549276)]
94. Klompas M, Morris CA, Sinclair J, Pearson M, Shenoy ES. Universal Masking in Hospitals in the Covid-19 Era. *N Engl J Med* 2020 May 21;382(21):e63. [doi: [10.1056/nejmp2006372](https://doi.org/10.1056/nejmp2006372)]
95. Javid B, Weekes M, Matheson N. Covid-19: should the public wear face masks? *BMJ* 2020 Apr 09;369:m1442. [doi: [10.1136/bmj.m1442](https://doi.org/10.1136/bmj.m1442)] [Medline: [32273278](https://pubmed.ncbi.nlm.nih.gov/32273278/)]
96. Mesa Vieira C, Franco OH, Gómez Restrepo C, Abel T. COVID-19: The forgotten priorities of the pandemic. *Maturitas* 2020 Jun;136:38-41 [FREE Full text] [doi: [10.1016/j.maturitas.2020.04.004](https://doi.org/10.1016/j.maturitas.2020.04.004)] [Medline: [32386664](https://pubmed.ncbi.nlm.nih.gov/32386664/)]
97. Greenhalgh T, Schmid MB, Czypionka T, Bassler D, Gruer L. Face masks for the public during the covid-19 crisis. *BMJ* 2020 Apr 09;369:m1435. [doi: [10.1136/bmj.m1435](https://doi.org/10.1136/bmj.m1435)] [Medline: [32273267](https://pubmed.ncbi.nlm.nih.gov/32273267/)]
98. Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y, et al. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infect Dis Model* 2020;5:293-308 [FREE Full text] [doi: [10.1016/j.idm.2020.04.001](https://doi.org/10.1016/j.idm.2020.04.001)] [Medline: [32355904](https://pubmed.ncbi.nlm.nih.gov/32355904/)]



99. Fairchild A, Gostin L, Bayer R. Vexing, Veiled, and Inequitable: Social distancing and the "rights" divide in the age of COVID-19. *Am J Bioeth* 2020 Jul 19;20(7):55-61. [doi: [10.1080/15265161.2020.1764142](https://doi.org/10.1080/15265161.2020.1764142)] [Medline: [32427528](https://pubmed.ncbi.nlm.nih.gov/32427528/)]
100. Bernstein J, Hutler B, Rieder T, Faden R, Han H, Barnhill A. An Ethics Framework for the COVID-19 Reopening Process. Johns Hopkins University. 2020 May. URL: <https://bioethics.jhu.edu/wp-content/uploads/2019/10/FINAL-SNF-Agora-Covid-19.pdf> [accessed 2020-07-20]
101. Halow B. Five Challenges To A Smooth Post-COVID Reopening. 2020. URL: <https://www.bashhalow.com/five-challenges-to-a-smooth-post-covid-reopening/> [accessed 2020-07-20]
102. O'Dowd A. Covid-19: Papers justifying government's plans to reopen schools are "inconclusive," say union bosses. *BMJ* 2020 May 27;369:m2108. [doi: [10.1136/bmj.m2108](https://doi.org/10.1136/bmj.m2108)] [Medline: [32461199](https://pubmed.ncbi.nlm.nih.gov/32461199/)]
103. Bhatia R, Jeffrey KM. A Step-by-Step Plan to Reopen California. *The Bold Italic*. 2020. URL: <https://thebolditalic.com/covid-19-next-steps-for-california-c01632c8e8b6> [accessed 2020-07-20]
104. Teutsch SM. Considerations in Planning a Surveillance System. In: Lee LS, Teutsch SM, Thacker SB, St. Louis ME, editors. New York, NY: Oxford University Press; 2010:18-28.
105. Ren X. Pandemic and lockdown: a territorial approach to COVID-19 in China, Italy and the United States. *Eurasian Geography and Economics* 2020 May 08;1-12. [doi: [10.1080/15387216.2020.1762103](https://doi.org/10.1080/15387216.2020.1762103)]
106. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* 2020 May 15;368(6492):742-746 [FREE Full text] [doi: [10.1126/science.abb4557](https://doi.org/10.1126/science.abb4557)] [Medline: [32269067](https://pubmed.ncbi.nlm.nih.gov/32269067/)]
107. Peng F, Tu L, Yang Y, Hu P, Wang R, Hu Q, et al. Management and Treatment of COVID-19: The Chinese Experience. *Can J Cardiol* 2020 Jun;36(6):915-930 [FREE Full text] [doi: [10.1016/j.cjca.2020.04.010](https://doi.org/10.1016/j.cjca.2020.04.010)] [Medline: [32439306](https://pubmed.ncbi.nlm.nih.gov/32439306/)]
108. Tang B, Xia F, Bragazzi N. Lessons drawn from China and South Korea for managing COVID-19 epidemic: insights from a comparative modeling study. *medRxiv* 2020:e [FREE Full text] [doi: [10.2471/blt.20.257238](https://doi.org/10.2471/blt.20.257238)]
109. Salzberger B, Glück T, Ehrenstein B. Successful containment of COVID-19: the WHO-Report on the COVID-19 outbreak in China. *Infection* 2020 Apr;48(2):151-153. [doi: [10.1007/s15010-020-01409-4](https://doi.org/10.1007/s15010-020-01409-4)] [Medline: [32185635](https://pubmed.ncbi.nlm.nih.gov/32185635/)]
110. Chudik A, Pesaran M, Rebucci A. Voluntary and Mandatory Social Distancing: Evidence on COVID-19 Exposure Rates from Chinese Provinces and Selected Countries. National Bureau of Economic Research. 2020. URL: <https://www.nber.org/papers/w27039> [accessed 2020-06-10]
111. Saez M, Tobias A, Varga D, Barceló MA. Effectiveness of the measures to flatten the epidemic curve of COVID-19. The case of Spain. *Sci Total Environ* 2020 Jul 20;727:138761 [FREE Full text] [doi: [10.1016/j.scitotenv.2020.138761](https://doi.org/10.1016/j.scitotenv.2020.138761)] [Medline: [32330703](https://pubmed.ncbi.nlm.nih.gov/32330703/)]
112. Kenyon C. Flattening-the-curve associated with reduced COVID-19 case fatality rates - an ecological analysis of 65 countries. *J Infect* 2020 Jul;81(1):e98-e99 [FREE Full text] [doi: [10.1016/j.jinf.2020.04.007](https://doi.org/10.1016/j.jinf.2020.04.007)] [Medline: [32305488](https://pubmed.ncbi.nlm.nih.gov/32305488/)]
113. De Leo S. Covid-19 testing strategies and lockdowns: the European closed curves, analysed by "skew-normal" distributions, the forecasts for the UK, Sweden, and the USA, and the ongoing outbreak in Brazil. *medRxiv* 2020:e [FREE Full text] [doi: [10.1101/2020.06.01.20119461](https://doi.org/10.1101/2020.06.01.20119461)]
114. Åslund A. Responses to the COVID-19 crisis in Russia, Ukraine, and Belarus. *Eurasian Geography and Economics* 2020 Jun 16;1-14. [doi: [10.1080/15387216.2020.1778499](https://doi.org/10.1080/15387216.2020.1778499)]
115. Russia's governors seem set up to fail on COVID-19. In: Emerald Expert Briefings. Bingley, United Kingdom: Emerald Publishing Limited; Apr 2020.
116. Forsyth IC. COVID-19: China's Chernobyl, China's Berlin Airlift, or Neither? Washington Headquarters Services. 2020 Jun. URL: <https://www.whs.mil/News/News-Display/Article/2210962/covid-19-chinas-chernobyl-chinas-berlin-airlift-or-neither/> [accessed 2020-07-20]
117. Mækelæ M, Reggev N, Dutra N, Tamayo R, Silva-Sobrinho R, Klevjer K, et al. Perceived efficacy of COVID-19 restrictions, reactions and their impact on mental health during the early phase of the outbreak in six countries. *R Soc Open Sci* 2020 Aug 12;7(8):200644. [doi: [10.1098/rsos.200644](https://doi.org/10.1098/rsos.200644)]
118. Belarus leader politically isolated by COVID-19 denial. In: Emerald Expert Briefings. Bingley, United Kingdom: Emerald Publishing Limited; Apr 2020.
119. Ukraine and Belarus seek loans to ride out COVID-19. In: Emerald Expert Briefings. Bingley, United Kingdom: Emerald Publishing Limited; May 2020.
120. COVID-19 forces difficult political choices on Ukraine. In: Emerald Expert Briefings. Bingley, United Kingdom: Emerald Publishing Limited; Mar 2020.
121. CDC. Provisional COVID-19 Death Counts by Week. National Center for Health Statistics. 2020. URL: <https://www.cdc.gov/nchs/nvss/vsr/covid19/index.htm> [accessed 2020-07-20]
122. Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. *The Lancet* 2020 Apr;395(10232):1243-1244. [doi: [10.1016/s0140-6736\(20\)30893-x](https://doi.org/10.1016/s0140-6736(20)30893-x)]
123. McKibbin W, Fernando R. The Global Macroeconomic Impacts of COVID-19: Seven Scenarios. *Asian Economic Papers* 2020 Aug 21;1-55. [doi: [10.1162/asep\\_a\\_00796](https://doi.org/10.1162/asep_a_00796)]
124. Baker S, Bloom N, Davis S, Terry S. Covid-induced economic uncertainty. National Bureau of Economic Research; 2020:e [FREE Full text] [doi: [10.3386/w26983](https://doi.org/10.3386/w26983)]

125. Omer SB, Malani P, Del Rio C. The COVID-19 Pandemic in the US: A Clinical Update. *JAMA* 2020 May 12;323(18):1767-1768. [doi: [10.1001/jama.2020.5788](https://doi.org/10.1001/jama.2020.5788)] [Medline: [32250388](https://pubmed.ncbi.nlm.nih.gov/32250388/)]
126. Brzezinski A, Deiana G, Kecht V, Van DD. The COVID-19 pandemic: government vs community action across the United States. *Covid Economics* 2020;7:156 [FREE Full text]
127. Ganz S. Does your state have 14-days of declining COVID-19 cases? Finding straightforward answers to a surprisingly tricky question. American Enterprise Institute 2020 May:e [FREE Full text]
128. Fox G, Trauer J, McBryde E. Modelling the impact of COVID-19 upon intensive care services in New South Wales. *Med J Aust* 2020:e [FREE Full text]
129. Choi S, Ki M. Estimating the reproductive number and the outbreak size of COVID-19 in Korea. *Epidemiol Health* 2020 Mar 12;42:e2020011 [FREE Full text] [doi: [10.4178/epih.e2020011](https://doi.org/10.4178/epih.e2020011)] [Medline: [32164053](https://pubmed.ncbi.nlm.nih.gov/32164053/)]
130. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine* 2020 May 05;172(9):577-582. [doi: [10.7326/m20-0504](https://doi.org/10.7326/m20-0504)]
131. Calafiore G, Novara C, Possieri C. A modified sir model for the COVID-19 contagion in Italy. *arXiv* 2020:e [FREE Full text]
132. Korolev I. Identification and estimation of the SEIRD epidemic model for COVID-19. *J Econom* 2020 Jul 30:e [FREE Full text] [doi: [10.1016/j.jeconom.2020.07.038](https://doi.org/10.1016/j.jeconom.2020.07.038)] [Medline: [32836680](https://pubmed.ncbi.nlm.nih.gov/32836680/)]
133. Pindyck R. COVID-19 and the welfare effects of reducing contagion. *National Bureau of Economic Research* 2020:e [FREE Full text] [doi: [10.3386/w27121](https://doi.org/10.3386/w27121)]
134. Stock J. Data gaps and the policy response to the novel coronavirus. *National Bureau of Economic Research* 2020:e [FREE Full text] [doi: [10.3386/w26902](https://doi.org/10.3386/w26902)]
135. He S, Tang S, Rong L. A discrete stochastic model of the COVID-19 outbreak: Forecast and control. *Math Biosci Eng* 2020;17:2804 [FREE Full text] [doi: [10.3934/mbe.2020153](https://doi.org/10.3934/mbe.2020153)]
136. Gandhi M, Yokoe D, Havlir D. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. *N Engl J Med* 2020 May 28;382(22):2158-2160 [FREE Full text] [doi: [10.1056/NEJMe2009758](https://doi.org/10.1056/NEJMe2009758)] [Medline: [32329972](https://pubmed.ncbi.nlm.nih.gov/32329972/)]
137. Mani N, Budak J, Lan K, Bryson-Cahn C, Zelikoff A, Barker GEC, et al. Prevalence of COVID-19 Infection and Outcomes Among Symptomatic Healthcare Workers in Seattle, Washington. *Clin Infect Dis* 2020 Jun 16;2020 [FREE Full text] [doi: [10.1093/cid/ciaa761](https://doi.org/10.1093/cid/ciaa761)] [Medline: [32548613](https://pubmed.ncbi.nlm.nih.gov/32548613/)]
138. Dodds C, Fakoya I. Covid-19: ensuring equality of access to testing for ethnic minorities. *BMJ* 2020 May 29;369:m2122. [doi: [10.1136/bmj.m2122](https://doi.org/10.1136/bmj.m2122)] [Medline: [32471825](https://pubmed.ncbi.nlm.nih.gov/32471825/)]
139. Schmitt-Grohé S, Teoh K, Uribe M. Testing Inequality in New York City. *National Bureau of Economic Research* 2020:e [FREE Full text] [doi: [10.3386/w27019](https://doi.org/10.3386/w27019)]
140. Voo T, Clapham H, Tam C. Ethical Implementation of Immunity Passports During the COVID-19 Pandemic. *J Infect Dis* 2020 Aug 04;222(5):715-718 [FREE Full text] [doi: [10.1093/infdis/jiaa352](https://doi.org/10.1093/infdis/jiaa352)] [Medline: [32582943](https://pubmed.ncbi.nlm.nih.gov/32582943/)]
141. Beeching N, Fletcher T, Beadsworth M. Covid-19: testing times. *BMJ* 2020 Apr 08;369:m1403. [doi: [10.1136/bmj.m1403](https://doi.org/10.1136/bmj.m1403)] [Medline: [32269032](https://pubmed.ncbi.nlm.nih.gov/32269032/)]
142. West CP, Montori VM, Sampathkumar P. COVID-19 Testing: The Threat of False-Negative Results. *Mayo Clin Proc* 2020 Jun;95(6):1127-1129 [FREE Full text] [doi: [10.1016/j.mayocp.2020.04.004](https://doi.org/10.1016/j.mayocp.2020.04.004)] [Medline: [32376102](https://pubmed.ncbi.nlm.nih.gov/32376102/)]
143. Kumleben N, Bhopal R, Cypionka T, Gruer L, Kock R, Stebbing J, et al. Test, test, test for COVID-19 antibodies: the importance of sensitivity, specificity and predictive powers. *Public Health* 2020 Aug;185:88-90 [FREE Full text] [doi: [10.1016/j.puhe.2020.06.006](https://doi.org/10.1016/j.puhe.2020.06.006)] [Medline: [32590234](https://pubmed.ncbi.nlm.nih.gov/32590234/)]
144. Cassaniti I, Novazzi F, Giardina F, Salinaro F, Sachs M, Perlini S, Members of the San Matteo Pavia COVID-19 Task Force. Performance of VivaDiag COVID-19 IgM/IgG Rapid Test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department. *J Med Virol* 2020 Mar 30:e [FREE Full text] [doi: [10.1002/jmv.25800](https://doi.org/10.1002/jmv.25800)] [Medline: [32227490](https://pubmed.ncbi.nlm.nih.gov/32227490/)]
145. Esbin MN, Whitney ON, Chong S, Maurer A, Darzacq X, Tjian R. Overcoming the bottleneck to widespread testing: a rapid review of nucleic acid testing approaches for COVID-19 detection. *RNA* 2020 Jul 01;26(7):771-783 [FREE Full text] [doi: [10.1261/rna.076232.120](https://doi.org/10.1261/rna.076232.120)] [Medline: [32358057](https://pubmed.ncbi.nlm.nih.gov/32358057/)]
146. Patel J, Nielsen F, Badiani A, Assi S, Unadkat V, Patel B, et al. Poverty, inequality and COVID-19: the forgotten vulnerable. *Public Health* 2020 Jun;183:110-111 [FREE Full text] [doi: [10.1016/j.puhe.2020.05.006](https://doi.org/10.1016/j.puhe.2020.05.006)] [Medline: [32502699](https://pubmed.ncbi.nlm.nih.gov/32502699/)]
147. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. *The Lancet Global Health* 2020 Apr;8(4):e480. [doi: [10.1016/s2214-109x\(20\)30068-1](https://doi.org/10.1016/s2214-109x(20)30068-1)]
148. Cafer A, Rosenthal M. COVID-19 in the Rural South: A Perfect Storm of Disease, Health Access, and Co-Morbidity. *APCRL Policy Briefs* 2020;2:e [FREE Full text]
149. Tsai J, Wilson M. COVID-19: a potential public health problem for homeless populations. *The Lancet Public Health* 2020 Apr;5(4):e186-e187. [doi: [10.1016/s2468-2667\(20\)30053-0](https://doi.org/10.1016/s2468-2667(20)30053-0)]
150. Harlem G. Corrigendum to: Descriptive analysis of social determinant factors in urban communities affected by COVID-19. *J Public Health (Oxf)* 2020 Sep 07:2020. [doi: [10.1093/pubmed/fdaa166](https://doi.org/10.1093/pubmed/fdaa166)] [Medline: [32894290](https://pubmed.ncbi.nlm.nih.gov/32894290/)]

151. Vetter P, Eckerle I, Kaiser L. Covid-19: a puzzle with many missing pieces. *BMJ* 2020 Feb 19;368:m627. [doi: [10.1136/bmj.m627](https://doi.org/10.1136/bmj.m627)] [Medline: [32075791](https://pubmed.ncbi.nlm.nih.gov/32075791/)]
152. Giamello J, Abram S, Bernardi S, Lauria G. The emergency department in the COVID-19 era. Who are we missing? *Eur J Emerg Med* 2020 Aug;27(4):305-306 [FREE Full text] [doi: [10.1097/MEJ.0000000000000718](https://doi.org/10.1097/MEJ.0000000000000718)] [Medline: [32345851](https://pubmed.ncbi.nlm.nih.gov/32345851/)]
153. Centers for Disease Control (CDC). Guidelines for evaluating surveillance systems. *MMWR Suppl* 1988 May 06;37(5):1-18 [FREE Full text] [Medline: [3131659](https://pubmed.ncbi.nlm.nih.gov/3131659/)]
154. Post LA, Balsen Z, Spano R, Vaca FE. Bolstering gun injury surveillance accuracy using capture-recapture methods. *J Behav Med* 2019 Aug 1;42(4):674-680. [doi: [10.1007/s10865-019-00017-4](https://doi.org/10.1007/s10865-019-00017-4)] [Medline: [31367932](https://pubmed.ncbi.nlm.nih.gov/31367932/)]
155. Piltch-Loeb R, Kraemer J, Lin KW, Stoto MA. Public Health Surveillance for Zika Virus: Data Interpretation and Report Validity. *Am J Public Health* 2018 Oct;108(10):1358-1362. [doi: [10.2105/ajph.2018.304525](https://doi.org/10.2105/ajph.2018.304525)]
156. Gilman M, Green R. The surveillance gap: The harms of extreme privacy and data marginalization. *Faculty Publications* 2018:253-307 [FREE Full text]
157. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: [10.1056/NEJMoa2001316](https://doi.org/10.1056/NEJMoa2001316)] [Medline: [31995857](https://pubmed.ncbi.nlm.nih.gov/31995857/)]
158. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science* 2003 Jun 20;300(5627):1966-1970 [FREE Full text] [doi: [10.1126/science.1086616](https://doi.org/10.1126/science.1086616)] [Medline: [12766207](https://pubmed.ncbi.nlm.nih.gov/12766207/)]
159. Kucharski A, Russell T, Diamond C. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *medRxiv* 2020:e [FREE Full text] [doi: [10.1101/2020.01.31.20019901](https://doi.org/10.1101/2020.01.31.20019901)]
160. Jones JH. Models of Infectious Disease. Stanford Spring Workshop in Formal Demography / Stanford University. 2008. URL: <https://web.stanford.edu/~jhj1/teachingdocs/Jones-Epidemics050308.pdf> [accessed 2020-06-19]
161. Martcheva M. An Introduction to Mathematical Epidemiology. New York, NY: Springer; 2015.
162. Braveman PA, Egerter SA, Woolf SH, Marks JS. When do we know enough to recommend action on the social determinants of health? *Am J Prev Med* 2011 Jan;40(1 Suppl 1):S58-S66. [doi: [10.1016/j.amepre.2010.09.026](https://doi.org/10.1016/j.amepre.2010.09.026)] [Medline: [21146780](https://pubmed.ncbi.nlm.nih.gov/21146780/)]
163. Guidelines: Opening Up America Again. White House / Centers for Disease Control and Prevention. 2020. URL: <https://www.whitehouse.gov/openingamerica/> [accessed 2020-09-21]
164. The COVID Tracking Project. The Atlantic. 2020. URL: <https://covidtracking.com/> [accessed 2020-08-31]
165. Brownstein JS, Freifeld CC, Madoff LC. Digital Disease Detection — Harnessing the Web for Public Health Surveillance. *N Engl J Med* 2009 May 21;360(21):2153-2157. [doi: [10.1056/nejmp0900702](https://doi.org/10.1056/nejmp0900702)]
166. COVID 19 Tracking. GitHub. 2020. URL: <https://github.com/COVID19Tracking/covid-tracking-data> [accessed 2020-08-31]
167. Oehmke J, Oehmke T, Singh L, Post L. Dynamic Panel Estimates of SARS-CoV-2 Infection Rates: Health Surveillance Informs Public Health Policy. *J Med Internet Res* 2020 Sep 09:e [FREE Full text] [doi: [10.2196/20924](https://doi.org/10.2196/20924)] [Medline: [32915762](https://pubmed.ncbi.nlm.nih.gov/32915762/)]
168. Hansen LP, Singleton KJ. Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica* 1982 Sep;50(5):1269. [doi: [10.2307/1911873](https://doi.org/10.2307/1911873)]
169. Hansen LP. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 1982 Jul;50(4):1029. [doi: [10.2307/1912775](https://doi.org/10.2307/1912775)]
170. Roodman D. xtabond2: Stata module to extend xtabond dynamic panel data estimator. Statistical Software Components S435901, Boston College Department of Economics. 2018. URL: <https://ideas.repec.org/c/boc/bocode/s435901.html> [accessed 2020-07-20]
171. Roodman D. How to do Xtabond2: An Introduction to Difference and System GMM in Stata. *The Stata Journal* 2018 Nov 19;9(1):86-136. [doi: [10.1177/1536867x0900900106](https://doi.org/10.1177/1536867x0900900106)]
172. Blundell R, Bond S. GMM Estimation with persistent panel data: an application to production functions. *Econometric Reviews* 2000 Jan;19(3):321-340. [doi: [10.1080/07474930008800475](https://doi.org/10.1080/07474930008800475)]
173. Bond S, Hoeffler A, Temple J. GMM estimation of empirical growth models. Economics Papers 2001-W21, Economics Group, Nuffield College, University of Oxford. 2001. URL: <https://ideas.repec.org/p/nuf/econwp/0121.html> [accessed 2020-07-20]
174. Arellano M. Dynamic Panel Data Estimation Using DPD: A Guide for Users. Great Britain: Institute for Fiscal Studies London; 1988.
175. COVID-19 Dashboard. Johns Hopkins University. 2020. URL: <https://coronavirus.jhu.edu/us-map> [accessed 2020-09-18]

*Edited by G Eysenbach; submitted 29.06.20; peer-reviewed by A Khurshid, E Paskett; comments to author 15.07.20; revised version received 03.09.20; accepted 13.09.20; published 05.10.20.*

*Please cite as:*

Oehmke JF, Moss CB, Singh LN, Oehmke TB, Post LA

Dynamic Panel Surveillance of COVID-19 Transmission in the United States to Inform Health Policy: Observational Statistical Study  
*J Med Internet Res* 2020;22(10):e21955

URL: <https://www.jmir.org/2020/10/e21955>

doi: [10.2196/21955](https://doi.org/10.2196/21955)

PMID: [32924962](https://pubmed.ncbi.nlm.nih.gov/32924962/)

©James Francis Oehmke, Charles B Moss, Lauren Nadya Singh, Theresa Bristol Oehmke, Lori Ann Post. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 05.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study

Nicolas Munsch<sup>1</sup>, MSc; Alistair Martin<sup>1</sup>, DPhil; Stefanie Gruarin<sup>2</sup>, MD, PhD; Jama Nateqi<sup>2,3</sup>, MD; Isselmou Abdarahmane<sup>1</sup>, BSc; Rafael Weingartner-Ortner<sup>1,2</sup>, BSc; Bernhard Knapp<sup>1</sup>, PhD

<sup>1</sup>Data Science Department, Symptoma, Vienna, Austria

<sup>2</sup>Medical Department, Symptoma, Attersee, Austria

<sup>3</sup>Department of Internal Medicine, Paracelsus Medical University, Salzburg, Austria

**Corresponding Author:**

Bernhard Knapp, PhD

Data Science Department

Symptoma

Landstraßer Gürtel 3

Vienna, 1030

Austria

Phone: 43 662458206

Email: [science@symptoma.com](mailto:science@symptoma.com)

## Abstract

**Background:** A large number of web-based COVID-19 symptom checkers and chatbots have been developed; however, anecdotal evidence suggests that their conclusions are highly variable. To our knowledge, no study has evaluated the accuracy of COVID-19 symptom checkers in a statistically rigorous manner.

**Objective:** The aim of this study is to evaluate and compare the diagnostic accuracies of web-based COVID-19 symptom checkers.

**Methods:** We identified 10 web-based COVID-19 symptom checkers, all of which were included in the study. We evaluated the COVID-19 symptom checkers by assessing 50 COVID-19 case reports alongside 410 non-COVID-19 control cases. A bootstrapping method was used to counter the unbalanced sample sizes and obtain confidence intervals (CIs). Results are reported as sensitivity, specificity, F1 score, and Matthews correlation coefficient (MCC).

**Results:** The classification task between COVID-19–positive and COVID-19–negative for “high risk” cases among the 460 test cases yielded (sorted by F1 score): Symptoma (F1=0.92, MCC=0.85), Infermedica (F1=0.80, MCC=0.61), US Centers for Disease Control and Prevention (CDC) (F1=0.71, MCC=0.30), Babylon (F1=0.70, MCC=0.29), Cleveland Clinic (F1=0.40, MCC=0.07), Providence (F1=0.40, MCC=0.05), Apple (F1=0.29, MCC=–0.10), Docyet (F1=0.27, MCC=0.29), Ada (F1=0.24, MCC=0.27) and Your.MD (F1=0.24, MCC=0.27). For “high risk” and “medium risk” combined the performance was: Symptoma (F1=0.91, MCC=0.83) Infermedica (F1=0.80, MCC=0.61), Cleveland Clinic (F1=0.76, MCC=0.47), Providence (F1=0.75, MCC=0.45), Your.MD (F1=0.72, MCC=0.33), CDC (F1=0.71, MCC=0.30), Babylon (F1=0.70, MCC=0.29), Apple (F1=0.70, MCC=0.25), Ada (F1=0.42, MCC=0.03), and Docyet (F1=0.27, MCC=0.29).

**Conclusions:** We found that the number of correctly assessed COVID-19 and control cases varies considerably between symptom checkers, with different symptom checkers showing different strengths with respect to sensitivity and specificity. A good balance between sensitivity and specificity was only achieved by two symptom checkers.

(*J Med Internet Res* 2020;22(10):e21299) doi:[10.2196/21299](https://doi.org/10.2196/21299)

**KEYWORDS**

COVID-19; symptom checkers; benchmark; digital health; symptom; chatbot; accuracy

## Introduction

In the modern world, large numbers of patients initially turn to various web-based sources for self-diagnoses of health concerns

before seeking diagnoses from a trained medical professional. However, web-based sources have inherent problems, such as misinformation, misunderstandings, misleading advertisements, and varying quality [1]. Interactive web sources developed to



provide web-based diagnoses are sometimes referred to as symptom checkers or chatbots [2,3]. Based on a list of entered symptoms and other factors, these symptom checkers return a list of potential diseases.

Web-based symptom checkers have become popular in the context of the novel COVID-19 pandemic, as access to physicians is reduced, concern in the population is high, and large amounts of misinformation are circulating the internet [1]. On COVID-19 symptom checker web pages, users are asked a series of COVID-19-specific questions; upon completion, an association between the answers and COVID-19 is given alongside behavioral recommendations, such as self-isolation.

In this context, COVID-19 symptom checkers are valuable tools for preassessment and screening during this pandemic; they can both ease pressure on clinicians and decrease footfall within hospitals. One example is practicing social distancing by not going to a physician's waiting room when feeling sick. The importance of social distancing has been highlighted in the COVID-19 pandemic [4,5], the 2009 H1N1 influenza pandemic [6], and the 1918-1919 influenza pandemic [7] and is reviewed in [8]. Symptom checkers can also ease pressure on medical telephone hotlines [9,10] by reducing the number of human phone operators needed.

A large number of symptom checkers specific to COVID-19 have been developed. Empirical evidence (eg, a newspaper article [11]) suggests that their conclusions differ, with possible implications for the quality of the symptom assessment. To our knowledge, there are no studies comparing and evaluating COVID-19 symptom checkers.

In this paper, we present a study evaluating 10 different web-based COVID-19 symptom checkers using 50 COVID-19 cases extracted from the literature and 410 non-COVID-19 control cases of patients with other diseases. We found that the classifications of many patient cases by the COVID-19 symptom checkers differ. Therefore the accuracies of symptom checkers also differ.

## Methods

### COVID-19 Symptom Checkers

In April 2020, we conducted a Google search for COVID-19 symptom checkers using the search terms *COVID-19 symptom checker* and *Corona symptom checker*. All ten COVID-19 symptom checkers that we found and that were freely available on the internet between April 3 and 9, 2020, were included in this study (Table 1). Nine checkers were implemented in the English language, while one was in German. These symptom checkers were used in the versions available in this date range, and updates after this date were not considered for analysis.

As a baseline for the performance evaluation of the 10 web-based COVID-19 symptom checkers, we developed two additional simplistic symptom checkers. These two checkers evaluate and weigh the presence of COVID-19 symptom frequencies provided by the World Health Organization (WHO) [12] (see Multimedia Appendix 1) based on vector distance (SF-DIST) and cosine similarity (SF-COS). These approaches can be implemented in a few lines of code (see Multimedia Appendix 2).

**Table 1.** List of web-based COVID-19 symptom checkers included in this study.

Name	Reference
Ada	[13]
Apple	[14]
Babylon	[15]
CDC <sup>a</sup>	[16]
Cleveland Clinic	[17]
Docyet	[18]
Infermedica	[19]
Providence	[20]
Symptoma	[21]
Your.MD	[22]

<sup>a</sup>CDC: US Centers for Disease Control and Prevention.

### Clinical Cases

We used a total of 460 clinical cases to evaluate the performance of the COVID-19 symptom checkers. Each case lists both

symptoms and the correct diagnosis alongside the age and sex of the patient when available. Details of the two case sets used are given below and in Table 2.

**Table 2.** Number of symptoms and age and sex distributions in each case set (N=460).

Characteristic	Case set	
	COVID-19, n=50	Control, n=410
<b>Number of symptoms</b>		
Mean (SD)	8.4 (4.1)	9.8 (4.4)
Median	7	9
<b>Age (years)</b>		
Mean (SD)	45.6 (16.9)	38.6 (22.4)
Median	45	38
<b>Sex, n (%)</b>		
Male	25 (50)	238 (58)
Female	21 (42)	160 (39)
Unknown	4 (8)	12 (2.9)

### COVID-19 Cases

A total of 50 COVID-19 cases were extracted by three trained physicians from the literature in March and April 2020 and are listed in [Multimedia Appendix 3](#). Each case describes one patient's medical situation (ie, symptoms experienced or COVID-19 contacts). The symptoms of each case were extracted separately from the COVID-19 engine construction and evaluation. The physicians entering the symptoms did not know how the engine would react to their symptom lists. To the best of our knowledge, we included all cases available at the time except for severe edge cases (eg, several severe comorbidities causing unrelated symptoms). Changes to the initial symptom lists were not allowed later.

### Control Cases

The COVID-19 case data enabled us to evaluate the sensitivity of the symptom checkers. To evaluate the specificity, 410 control cases from the *British Medical Journal* (BMJ) were also sourced [23,24]. To allow a fair assessment, we only used cases containing at least one of the COVID-19 symptoms reported by the WHO [12] (see [Multimedia Appendix 4](#)). Classifying nonrelevant cases (eg, a fracture) would overestimate the symptom checkers' specificity. Furthermore, these patients would not consult a web-based COVID-19 symptom checker. None of the 410 BMJ cases lists COVID-19 as the diagnosis, as the cases were collected before the COVID-19 outbreak.

### Mapping of Symptoms and Addressing Missing Inputs and Questions

Each of the symptom checkers has a different interface and different question sequences to reach the diagnosis. Therefore, we mapped the exhibited symptoms across our cases to the constrained input allowed by each checker via a synonym table and hierarchy created by a trained physician. For example, if a checker asked for "shortness of breath" but the case description listed "respiratory distress" or "(acute) dyspnea", the symptom was still correctly used for this case and symptom checker.

Not all cases contained answers to all the questions of a checker. In such cases, the answer "I don't know" was chosen; if the "I don't know" answer option did not exist in a checker, "No" was

used. In contrast, if a case contained information that did not fit any of the questions of the checker, this information was not used for this checker.

### Accuracy Evaluation

For statistical analysis, we used the following classification:

- True-positive : COVID-19 case classified as COVID-19
- False-positive: non-COVID-19 case classified as COVID-19
- True-negative: non-COVID-19 case classified as non-COVID-19
- False-negative: COVID-19 case classified as non-COVID-19

For each symptom checker, we calculated the following metrics:

Sensitivity (true-positive rate):



Specificity (true-negative rate):



F1 score (harmonic mean of the precision and recall):



Matthews correlation coefficient (MCC):



### Classification of the Outputs of the Symptom Checkers

Most COVID-19 symptom checkers return human-readable text that contains an association between the entered symptoms and COVID-19. We classified these associations into three different categories: high risk, medium risk, and low risk. Respective examples of a high, medium, and low risk classification are "There is a high risk that COVID-19 is causing your symptoms," "Your symptoms are worrisome and may be related to COVID-19," and "There's nothing at present to suggest that

you have coronavirus (COVID-19). Please practice physical/social distancing.” Our full mapping of text outputs to risk for all symptom checkers and all text outputs is given in [Multimedia Appendix 5](#).

Some symptom checkers only have two possible outputs: COVID-19 risk or no COVID-19 risk. To compare symptom checkers with three and two risk levels, we performed two different analyses: (1) medium risk and high risk were treated as COVID-19–positive (and low risk was treated as COVID-19 as negative), and (2) high risk was treated as COVID-19–positive (and low risk and medium risk were treated as COVID-19–negative).

### Bootstrapping

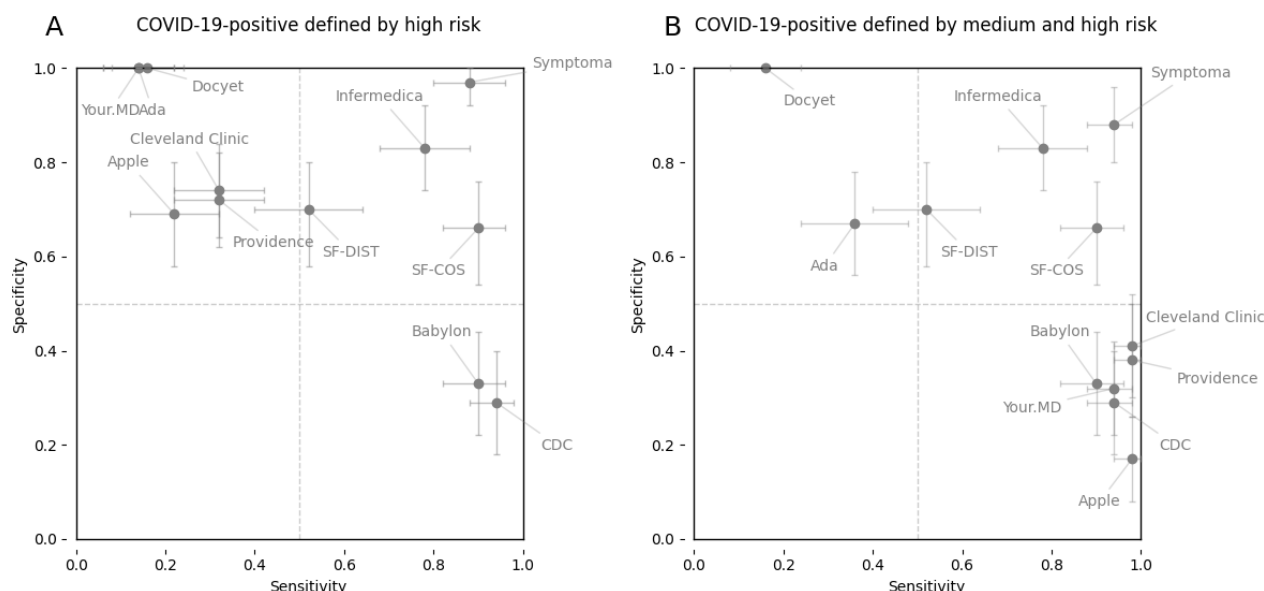
To evaluate the robustness of our statistical measures and account for the unbalanced dataset, we performed bootstrapping

across our cases. A total of 3000 random samples consisting of 50 COVID-19 cases and 50 control cases was created by sampling with replacement from the original set of 50 COVID-19 cases and the 410 control cases.

## Results

To analyze the performance of the 10 web-based symptom checkers, we calculated the sensitivity and the specificity of each symptom checker based on the cases described in the method section. Scatterplots of the sensitivity and specificity to COVID-19 of the different symptom checkers are given in [Figure 1](#), and detailed numerical values are provided in [Multimedia Appendix 6](#) and [Multimedia Appendix 7](#). These symptom checkers fall approximately into four groups: upper left corner, lower right corner, central region, and upper right corner.

**Figure 1.** Sensitivities and specificities of web-based COVID-19 symptom checkers to COVID-19 cases and controls. The means of the 3000 random samples and 90% bootstrap CIs are reported as dots and crosses, respectively. (A) High risk: A COVID-19–positive prediction is defined only by a high risk result returned by a symptom checker. (B) Medium-high risk: A COVID-19–positive prediction is defined by either a medium risk or high risk result returned by a symptom checker. CDC: US Centers for Disease Control and Prevention; SF-COS: symptom frequency based on cosine similarity; SF-DIST: symptom frequency based on vector distance.



Further analysis of the true and false case classifications of these groups shows that the group in the upper left corner is composed of symptom checkers that require the presence of one (or few) highly specific symptoms to classify a case as COVID-19–positive (eg, “intensive contact with a COVID-19–positive person”). In this way, these symptom checkers miss many patients who are positive for COVID-19 who did not exactly report this highly specific symptom. In contrast, these highly specific symptoms are rarely present in non–COVID-19 cases. This results in low sensitivity and high specificity.

The group in the lower right corner is composed of symptom checkers that predict a case as COVID-19–positive based on the presence of one or few COVID-19 associated symptoms (eg, the presence of fever or cough is sufficient to predict a patient to be COVID-19–positive). These checkers classify almost every patient that has a respiratory disorder or viral

infection as COVID-19–positive. As such, they do not miss many patients with COVID-19 but wrongly predict many patients who do not have COVID-19 to be COVID-19–positive. This results in low specificity and high sensitivity.

The group in the more central region is composed of symptom checkers that use a more balanced prediction but exhibit limited success in correctly classifying patients with and without COVID-19.

The group in the upper right corner is composed of symptom checkers that also use a more balanced model to associate symptoms to COVID-19; however, in this case, the classification of patients with and without COVID-19 is more successful.

## Discussion

### Principal Findings

We classified 50 COVID-19 case descriptions from the recent literature as well as 410 non-COVID-19 control cases using 10 different web-based COVID-19 symptom checkers. Only 2/10 symptom checkers showed a reasonably good balance between sensitivity and specificity (Figure 1). Most other checkers were either too sensitive, classifying almost all patients as COVID-19-positive, or too specific, classifying many patients with COVID-19 as COVID-19-negative (Figure 1). For example, our BMJ control cases included a patient suffering from a pulmonary disease who presented with various symptoms, including fever, cough, and shortness of breath, which are the three most frequent symptoms associated with COVID-19. Additional symptoms and risk factors were not considered by most checkers. Namely, loss of appetite, green sputum, and a history of smoking can be used to discern a correct diagnosis of COVID-19-negative.

Furthermore, in terms of F1 score, most of the symptom checkers were outperformed by a simplistic symptom frequency vector approach; the F1 scores were 0.57 and 0.79 for SF-DIST and SF-COS, respectively. Notably, the cosine version showed surprisingly good results, outperforming 8/10 symptom checkers based on the F1 score.

In contrast, it could also be argued that sensitivity is more important for a COVID-19 symptom checker than specificity (ie, numerous false-positive COVID-19 diagnoses are not of concern as long as no COVID-19 infections are missed). However, it is not difficult to create a symptom checker that is 100% sensitive by simply returning every test as COVID-19-positive. While no checker does this 100% of the time, some checkers tend to declare every person who reports any flu-like symptom to be COVID-19-positive. This assesses every patient with allergic asthma (“shortness of breath”), heatstroke (“fever”), or heavy smoker (“cough”) to be COVID-19-positive. Therefore, we believe that a healthy balance between sensitivity and specificity is necessary for a useful checker. However, from the figure in this paper, readers can decide for themselves which balance between sensitivity and specificity is most useful and select the corresponding checker.

An additional aspect is that the developers of the 10 checkers may have had different purposes in mind during development. For example, they may have envisioned the checker to be a self-triage and recommendation tool or a likelihood predictor (as classified in [2]). In our study, we found that most checkers provide a certain likelihood as well as recommendations; therefore, classification is difficult. Therefore, we did not further subgroup the checkers in our analysis.

To our knowledge, this is the first scientific evaluation of web-based COVID-19 symptom checkers; however, there are a number of related studies evaluating symptom checkers. These include a study that evaluated 23 general-purpose symptom checkers based on 45 clinical case descriptions across a wide range of medical conditions and found that the correct diagnosis

was listed among the top 20 results of the checkers in 58% of all cases on average [2]. The aforementioned study design was extended to five additional symptom checkers using ear, nose, and throat (ENT) cases, showing similar results [25]. Other evaluations include a study of symptom checkers used for knee pain cases; based on 527 patients and 26 knee problems, it was found that the physician’s diagnosis was present within the prediction list in 89% of the cases, while the specificity was only 27% [26]. In another study, an analysis of an automated self-assessment triage system for university students prior to an in-person consultation with a physician found that the system’s urgency rating agreed perfectly in only 39% of cases; meanwhile, for the remaining cases, the system tended to be more risk-averse than the physician [27]. Also, the applicability of web-based symptom checkers for 79 persons aged  $\geq 50$  years based on “think-aloud” protocols [28], deep learning algorithms for medical imaging [29], and services for urgent care [3] were evaluated.

The acceptability of the performance of a web-based symptom checker depends on the perspective and use of the results. In the case of COVID-19, a web-based assessment cannot fully replace a polymerase chain reaction (PCR) test, as some people are asymptomatic while others presenting with very specific COVID-19 symptoms may in fact have a very similar but different disease. Regardless, web-based COVID-19 symptom checkers can act as a first triage shield to avoid in-person physician visits or ease pressure on hospitals. Symptom checkers could even replace telephone triage lines in which non-medically trained personnel read a predefined sequence of questions. Although this was not part of this study, the authors believe that COVID-19 symptom checkers (if appropriately maintained and tested) may also be more reliable than the direct use of search engines such as Google or information via social media.

### Strengths and Limitations

The strength of this study lies in the fact that it is based on a large number of real patients’ case descriptions from the literature ( $n=460$ ) and a detailed evaluation of the best performing symptom checker in terms of F1 score (Multimedia Appendix 8). In contrast, a potential weakness of this study lies in its use of real literature-based cases, which may have biased the test set to rather severe cases of COVID-19 because mild and uninteresting cases are usually not found in the literature. We countered this bias by not including extreme edge cases from the literature in our 50 COVID-19 cases. A limitation of our study is that the benchmarking represents a specific point in time (April 2020; see Methods) and underlying algorithms may change. However, this temporal limitation is present in all benchmarking studies as knowledge increases and software is updated. Another bias may be that our control case descriptions do not report a COVID-19 contact, even though, for example, a person with a cold may have had a COVID-19 contact (and did not become infected). Another limitation of this study is the nonstraightforward mapping of the symptom checker outputs to risk levels (Multimedia Appendix 5). The interpretation of the textual output is debatable in some cases. We countered this by allowing three different risk levels and merging them in two different ways (see Figure 1A and Figure 1B). Also, every

symptom checker output was classified by multiple persons until consensus was reached.

## Conclusion

Symptom checkers are being widely used in response to the global COVID-19 pandemic. As such, quality assessment of

these tools is critical. We show that various web-based COVID-19 symptom checkers vary widely in their predictive capabilities, with some performing equivalently to random guessing while others show strength in sensitivity, specificity, or both.

## Acknowledgments

This study received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 830017.

## Conflicts of Interest

All authors are employees of Symptoma GmbH. JN holds shares in Symptoma.

### Multimedia Appendix 1

Symptom frequencies used in Multimedia Appendix 2.

[\[PDF File \(Adobe PDF File\), 28 KB - jmir\\_v22i10e21299\\_app1.pdf\]](#)

### Multimedia Appendix 2

Pseudocode of symptom frequencies based on vector distance (SF-DIST) and cosine similarity (SF-COS).

[\[PDF File \(Adobe PDF File\), 49 KB - jmir\\_v22i10e21299\\_app2.pdf\]](#)

### Multimedia Appendix 3

List of the COVID-19 cases.

[\[PDF File \(Adobe PDF File\), 50 KB - jmir\\_v22i10e21299\\_app3.pdf\]](#)

### Multimedia Appendix 4

List of COVID-19 symptoms according to the World Health Organization.

[\[PDF File \(Adobe PDF File\), 13 KB - jmir\\_v22i10e21299\\_app4.pdf\]](#)

### Multimedia Appendix 5

Mappings between output texts and risk levels of the symptom checkers. All mappings were independently performed by two different persons, and conflicts were resolved by a third person's opinion.

[\[PDF File \(Adobe PDF File\), 33 KB - jmir\\_v22i10e21299\\_app5.pdf\]](#)

### Multimedia Appendix 6

Full table of sensitivities, specificities, accuracies, F1 scores, and Matthews correlation coefficients for all symptom checkers (COVID-19-positive was defined by "high risk" for nonbinary symptom checkers).

[\[PDF File \(Adobe PDF File\), 24 KB - jmir\\_v22i10e21299\\_app6.pdf\]](#)

### Multimedia Appendix 7

Full table of sensitivities, specificities, accuracies, F1 scores, and Matthews correlation coefficients for all symptom checkers (COVID-19-positive was defined by "medium risk or "high risk" for nonbinary symptom checkers).

[\[PDF File \(Adobe PDF File\), 24 KB - jmir\\_v22i10e21299\\_app7.pdf\]](#)

### Multimedia Appendix 8

Constraining symptoms for Symptoma.

[\[PDF File \(Adobe PDF File\), 101 KB - jmir\\_v22i10e21299\\_app8.pdf\]](#)

### Multimedia Appendix 9

Sensitivity vs specificity for all symptom checkers and Symptoma input constraint respectively by each symptom checker.

[\[PDF File \(Adobe PDF File\), 253 KB - jmir\\_v22i10e21299\\_app9.pdf\]](#)

### Multimedia Appendix 10



Full table of sensitivities, specificities, accuracies, F1 scores, and Matthews correlation coefficients for Symptoma constrained by each symptom checker (COVID-19–positive was defined by “high risk” for nonbinary symptom checkers).

[PDF File (Adobe PDF File), 33 KB - [jmir\\_v22i10e21299\\_app10.pdf](#)]

#### Multimedia Appendix 11

Full table of sensitivities, specificities, accuracies, F1 scores, and Matthews correlation coefficients for Symptoma constrained by each symptom checker (COVID-19–positive was defined by “medium risk” or “high risk” for nonbinary symptom checkers).

[PDF File (Adobe PDF File), 33 KB - [jmir\\_v22i10e21299\\_app11.pdf](#)]

#### Multimedia Appendix 12

Pairwise comparisons between all symptom checkers and Symptoma based on the MCC only if the subset of symptoms used by one checker is also used for Symptoma.

[PDF File (Adobe PDF File), 79 KB - [jmir\\_v22i10e21299\\_app12.pdf](#)]

## References

1. Tasnim S, Hossain MM, Mazumder H. Impact of Rumors and Misinformation on COVID-19 in Social Media. *J Prev Med Public Health* 2020 May;53(3):171-174 [FREE Full text] [doi: [10.3961/jpmph.20.094](#)] [Medline: [32498140](#)]
2. Semigran H, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015 Jul 08;351:h3480. [doi: [10.1136/bmj.h3480](#)] [Medline: [26157077](#)]
3. Chambers D, Cantrell A, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and assessment services for urgent care to inform a new digital platform: a systematic review. *Health Services and Delivery Research* 2019;7(29):online. [doi: [10.3310/hsdr07290](#)] [Medline: [31433612](#)]
4. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* 2020 Jun 01:1973-1987 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)31142-9](#)] [Medline: [32497510](#)]
5. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 2020 May 22;368(6493):860-868 [FREE Full text] [doi: [10.1126/science.abb5793](#)] [Medline: [32291278](#)]
6. Lee B, Haidari L, Lee M. Modelling during an emergency: the 2009 H1N1 influenza pandemic. *Clin Microbiol Infect* 2013 Nov;19(11):1014-1022 [FREE Full text] [doi: [10.1111/1469-0691.12284](#)] [Medline: [23800220](#)]
7. Caley P, Philp DJ, McCracken K. Quantifying social distancing arising from pandemic influenza. *J R Soc Interface* 2008 Jun 06;5(23):631-639 [FREE Full text] [doi: [10.1098/rsif.2007.1197](#)] [Medline: [17916550](#)]
8. Ahmed F, Zviedrite N, Uzicanin A. Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review. *BMC Public Health* 2018 Apr 18;18(1):518 [FREE Full text] [doi: [10.1186/s12889-018-5446-1](#)] [Medline: [29669545](#)]
9. Kristal R, Rowell M, Kress M, Keeley C, Jackson H, Piwnica-Worms K, et al. A Phone Call Away: New York's Hotline And Public Health In The Rapidly Changing COVID-19 Pandemic. *Health Aff (Millwood)* 2020 Aug 01;39(8):1431-1436. [doi: [10.1377/hlthaff.2020.00902](#)] [Medline: [32525707](#)]
10. Judson T, Odisho A, Neinstein A, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. *J Am Med Inform Assoc* 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: [10.1093/jamia/ocaa051](#)] [Medline: [32267928](#)]
11. Ross C. I asked eight chatbots whether I had Covid-19. The answers ranged from ‘low’ risk to ‘start home isolation’. *STAT*. 2020 Mar 23. URL: <https://www.statnews.com/2020/03/23/coronavirus-i-asked-eight-chatbots-whether-i-had-covid-19/> [accessed 2020-09-30]
12. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). World Health Organization. 2020 Feb. URL: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf> [accessed 2020-09-30]
13. COVID-19 screener. ada. URL: <https://ada.com/covid-19-screener/> [accessed 2020-04-09]
14. COVID-19 Screening Tool. Apple. URL: <https://www.apple.com/covid19> [accessed 2020-04-09]
15. COVID-19 Symptom Checker. Babylon. URL: <https://www.babylonhealth.com/ask-babylon-chat> [accessed 2020-04-09]
16. Symptoms of Coronavirus. US Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html> [accessed 2020-04-09]
17. Coronavirus Self-Checker. Cleveland Clinic. URL: <http://covid19chat.clevelandclinic.org/> [accessed 2020-04-09]
18. Corona information. Webpage in German. Docyet. URL: <https://corona.docyet.com/client/index.html> [accessed 2020-04-09]
19. Symptomate. Infermedica. URL: <https://symptomate.com/covid19/checkup/en/> [accessed 2020-04-09]
20. Coronavirus assessment tool. Providence. URL: <https://coronavirus.providence.org/> [accessed 2020-04-09]
21. COVID-19 Chatbot Test. COVID-19. URL: <https://www.symptoma.com/covid-19> [accessed 2020-04-09]

22. Your.MD. URL: <https://webapp.your.md/login> [accessed 2020-04-09]
23. BMJ Best Practice. URL: <https://bestpractice.bmj.com/info/> [accessed 2020-04-09]
24. BMJ Case Reports. BMJ Journals. URL: <https://casereports.bmj.com> [accessed 2020-09-30]
25. Nateqi J, Lin S, Krobath H, Gruarin S, Lutz T, Dvorak T, et al. From symptom to diagnosis-symptom checkers re-evaluated: Are symptom checkers finally sufficient and accurate to use? An update from the ENT perspective.. HNO 2019 May 16;67(5):334-342. [doi: [10.1007/s00106-019-0666-y](https://doi.org/10.1007/s00106-019-0666-y)] [Medline: [30993374](https://pubmed.ncbi.nlm.nih.gov/30993374/)]
26. Bisson LJ, Komm JT, Bernas GA, Fineberg MS, Marzo JM, Rauh MA, et al. Accuracy of a computer-based diagnostic program for ambulatory patients with knee pain. Am J Sports Med 2014 Oct;42(10):2371-2376. [doi: [10.1177/0363546514541654](https://doi.org/10.1177/0363546514541654)] [Medline: [25073597](https://pubmed.ncbi.nlm.nih.gov/25073597/)]
27. Poote AE, French DP, Dale J, Powell J. A study of automated self-assessment in a primary care student health centre setting. J Telemed Telecare 2014 Mar 18;20(3):123-127. [doi: [10.1177/1357633x14529246](https://doi.org/10.1177/1357633x14529246)]
28. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: [10.2196/jmir.2924](https://doi.org/10.2196/jmir.2924)] [Medline: [24434479](https://pubmed.ncbi.nlm.nih.gov/24434479/)]
29. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020 Mar 25;368:m689. [doi: [10.1136/bmj.m689](https://doi.org/10.1136/bmj.m689)] [Medline: [32213531](https://pubmed.ncbi.nlm.nih.gov/32213531/)]

## Abbreviations

**BMJ:** British Medical Journal

**ENT:** ear, nose, and throat

**MCC:** Matthews correlation coefficient

**SF-COS:** symptom frequency based on cosine similarity

**SF-DIST:** symptom frequency based on vector distance

**WHO:** World Health Organization

*Edited by T Rashid Soron, G Eysenbach; submitted 15.06.20; peer-reviewed by T Judson, E Bellei; comments to author 11.07.20; revised version received 27.07.20; accepted 14.09.20; published 06.10.20.*

*Please cite as:*

Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, Knapp B

Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study

J Med Internet Res 2020;22(10):e21299

URL: <http://www.jmir.org/2020/10/e21299/>

doi: [10.2196/21299](https://doi.org/10.2196/21299)

PMID: [33001828](https://pubmed.ncbi.nlm.nih.gov/33001828/)

©Nicolas Munsch, Alistair Martin, Stefanie Gruarin, Jama Nateqi, Isselmou Abdarahmane, Rafael Weingartner-Ortner, Bernhard Knapp. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Clinical Predictive Models for COVID-19: Systematic Study

Patrick Schwab<sup>1</sup>, PhD; August DuMont Schütte<sup>2</sup>, MSc; Benedikt Dietz<sup>2</sup>, MSc; Stefan Bauer<sup>3</sup>, PhD

<sup>1</sup>F Hoffmann-La Roche Ltd, Basel, Switzerland

<sup>2</sup>Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland

<sup>3</sup>Max Planck Institute for Intelligent Systems, Tübingen, Germany

**Corresponding Author:**

Patrick Schwab, PhD

F Hoffmann-La Roche Ltd

Grenzacherstrasse 124

Basel, 4070

Switzerland

Phone: 41 793093361

Email: [patrick.schwab@icloud.com](mailto:patrick.schwab@icloud.com)

## Abstract

**Background:** COVID-19 is a rapidly emerging respiratory disease caused by SARS-CoV-2. Due to the rapid human-to-human transmission of SARS-CoV-2, many health care systems are at risk of exceeding their health care capacities, in particular in terms of SARS-CoV-2 tests, hospital and intensive care unit (ICU) beds, and mechanical ventilators. Predictive algorithms could potentially ease the strain on health care systems by identifying those who are most likely to receive a positive SARS-CoV-2 test, be hospitalized, or admitted to the ICU.

**Objective:** The aim of this study is to develop, study, and evaluate clinical predictive models that estimate, using machine learning and based on routinely collected clinical data, which patients are likely to receive a positive SARS-CoV-2 test or require hospitalization or intensive care.

**Methods:** Using a systematic approach to model development and optimization, we trained and compared various types of machine learning models, including logistic regression, neural networks, support vector machines, random forests, and gradient boosting. To evaluate the developed models, we performed a retrospective evaluation on demographic, clinical, and blood analysis data from a cohort of 5644 patients. In addition, we determined which clinical features were predictive to what degree for each of the aforementioned clinical tasks using causal explanations.

**Results:** Our experimental results indicate that our predictive models identified patients that test positive for SARS-CoV-2 a priori at a sensitivity of 75% (95% CI 67%-81%) and a specificity of 49% (95% CI 46%-51%), patients who are SARS-CoV-2 positive that require hospitalization with 0.92 area under the receiver operator characteristic curve (AUC; 95% CI 0.81-0.98), and patients who are SARS-CoV-2 positive that require critical care with 0.98 AUC (95% CI 0.95-1.00).

**Conclusions:** Our results indicate that predictive models trained on routinely collected clinical data could be used to predict clinical pathways for COVID-19 and, therefore, help inform care and prioritize resources.

(*J Med Internet Res* 2020;22(10):e21439) doi:[10.2196/21439](https://doi.org/10.2196/21439)

**KEYWORDS**

SARS-CoV-2; COVID-19; machine learning; clinical prediction; prediction; infectious disease; clinical data; testing; hospitalization; intensive care

## Introduction

COVID-19 was first discovered in December 2019 in China and has since rapidly spread to over 200 countries [1]. The COVID-19 pandemic has challenged health care systems worldwide, as a high peak capacity for testing and hospitalization is necessary to diagnose and treat affected patients, particularly if the spread of SARS-CoV-2 is not

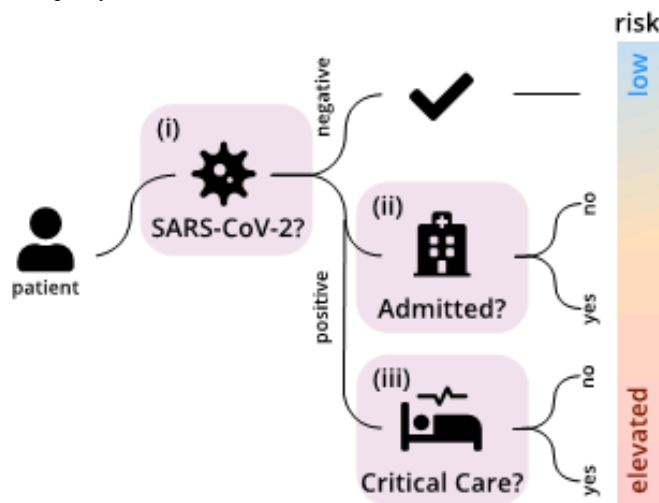
mitigated. To avoid exceeding the available health care capacities, many countries have adopted social distancing policies, imposed travel restrictions, and postponed nonessential care and surgeries to reduce peak demand on their health care systems [2-4].

The adoption of clinical predictive models that accurately predict who is likely to require testing, hospitalization, and intensive care from routinely collected clinical data could potentially

further reduce peak demand by ensuring resources are prioritized to those individuals with the highest risk (Figure 1). For example, a clinical predictive model that accurately identifies patients that are likely to test positive for SARS-CoV-2 a priori could help prioritize limited SARS-CoV-2 testing capacity.

However, developing accurate clinical prediction models for SARS-CoV-2 is difficult as relationships between clinical data, hospitalization, and intensive care unit (ICU) admission have not yet been established conclusively due to the recent emergence of SARS-CoV-2.

**Figure 1.** We study the use of predictive models (light purple) to estimate whether patients are likely (i) to be SARS-CoV-2 positive and whether SARS-CoV-2 positive patients are likely (ii) to be admitted to the hospital and (iii) to require critical care based on clinical, demographic, and blood analysis data. Accurate clinical predictive models stratify patients according to individual risk and, in this manner, help prioritize health care resources such as testing, hospital, and critical care capacity.



In this systematic study, we develop and evaluate clinical predictive models that use routinely collected clinical data to identify patients that are likely to receive a positive SARS-CoV-2 test, patients who are SARS-CoV-2 positive that are likely to require hospitalization, and patients who are SARS-CoV-2 positive that are likely to require intensive care. Using the developed predictive models, we additionally determined which clinical features are most predictive for each of the aforementioned clinical tasks. Our results indicate that predictive models could be used to predict clinical pathways for patients with COVID-19. Such predictive models may be of significant utility for health care systems, as preserving health care capacity has been linked to successfully combating SARS-CoV-2 [5,6].

Concretely, this paper contains the following contributions:

- We developed and systematically studied predictive models for estimating the likelihoods of a positive SARS-CoV-2 test in patients presenting at hospitals, hospital admission in patients who are SARS-CoV-2 positive, and critical care admission in patients who are SARS-CoV-2 positive.
- We validated the performance of the developed clinical predictive models in a retrospective evaluation using real-world data from a cohort of 5644 patients.
- We determined and quantified the predictive power of routinely collected clinical, demographic, and blood analysis data for the aforementioned clinical prediction tasks.

## Methods

### Problem Setting

In the present setting, we are given 106 routine clinical, laboratory, and demographic measurements, or features,  $x_i \in x$

for presenting patients (see Multimedia Appendix 1 for full list). Features may be discrete or continuous, and some features may be missing as not all tests are necessarily performed on all patients. The clinical predictive tasks consist of using the routine clinical features  $x_i$  to predict, for a newly presenting patient, the likelihood  $\hat{y}_{SARS-CoV-2}$  of receiving a positive SARS-CoV-2 test result, the likelihood  $\hat{y}_{admission}$  of requiring hospital admission, and the likelihood  $\hat{y}_{ICU}$  of requiring intensive care. In addition, we are given a development data set consisting of  $N$  patients, their corresponding observed routine clinical features  $x_i$ , SARS-CoV-2 test results  $\hat{y}_{SARS-CoV-2} \in \{0,1\}$ , hospital admissions  $\hat{y}_{admission} \in \{0,1\}$ , and ICU admissions  $\hat{y}_{ICU} \in \{0,1\}$ , where 1 indicates the presence of an outcome. Using this development data set, our goal is to derive clinical predictive models  $\hat{y}_{SARS-CoV-2}$ ,  $\hat{y}_{admission}$ , and  $\hat{y}_{ICU}$  for the aforementioned tasks, respectively, to inform care and help prioritize scarce health care resources.



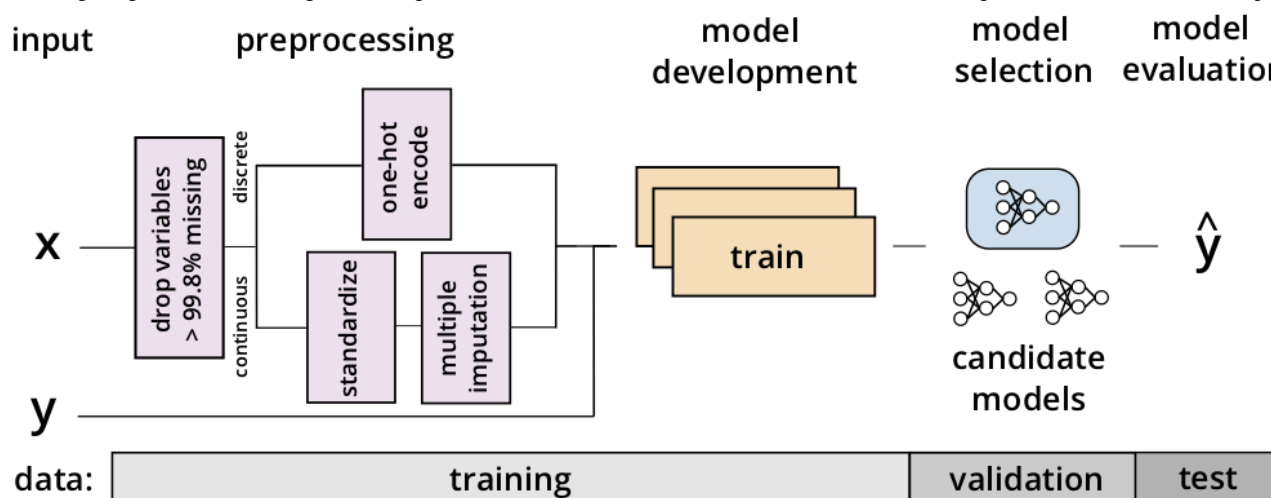
### Methodology

To derive the clinical predictive models  $\hat{y}_{SARS-CoV-2}$ ,  $\hat{y}_{admission}$ , and  $\hat{y}_{ICU}$  from the given development data set, we set up a systematic model development, validation, and evaluation pipeline (Figure 2). To evaluate the generalization ability of the developed clinical predictive models and to rule out overfitting to patients in the evaluation cohort, the development data is initially split into independent and stratified training, validation, and test folds without any patient overlap. Concretely, the multistage pipeline consists of preprocessing, model development, model selection, and model evaluation stages. For preprocessing and model

development, only the training fold was used, and only the validation and test folds of the development data were used for

model selection and model evaluation, respectively. We outline the pipeline stages in detail in the following paragraphs.

**Figure 2.** The presented multistage machine learning pipeline consists of preprocessing (light purple) the input data  $x$ , developing multiple candidate models using the given data set (orange), selecting the best candidate model for evaluation (blue), and evaluating the selected best model's outputs  $\hat{y}$ .



## Preprocessing

In the preprocessing stage, we first dropped all input features that were missing for more than 99.8% of all training set patients to ensure we had a minimal amount of data for each feature. This removed a total of 9 features from the original 106 routine clinical, laboratory, and demographic features. We then transformed all discrete features for each patient into their one-hot encoded representation with one out of  $p$  indicator variables set to 1 to indicate the discrete value for this patient, and all others set to 0 with  $p$  being the number of unique values for the discrete feature. We defined those features as discrete that have fewer than 6 unique values across all patients in the training fold. For discrete features, missing features were counted as a separate category in the one-hot representation. Next, we standardized all continuous features to have zero mean and unit standard deviation across the training fold data. Last, we performed multiple imputation by chained equations (MICE) to impute all missing values of every continuous feature from the respective other features in an iterative fashion [7]. We additionally added a missing indicator that indicates 1 if the feature was imputed by MICE and 0 if it was originally present to preserve missingness information in the data after imputation. After the preprocessing stage, continuous input features are standardized and fully imputed, and discrete input features are one-hot encoded. All preprocessing operations were derived only from the training fold and naïvely applied without adjustment to validation and test folds to avoid information leakage.

## Model Development

In the model development stage, we trained candidate clinical predictive models  $\square$ ,  $\square$ , and  $\square$  using supervised learning on the training fold of the preprocessed data. To derive the models from the preprocessed training fold data, we optimized various types of predictive models and performed a hyperparameter search with  $m$  runs for each of them. The model development process yielded  $m$  candidate models with different

hyperparameter choices and predictive performances for each model category.

## Model Selection

To select the best model among the set of candidate models, we evaluated their predictive performance against the held-out validation fold that had not been used for model development. We chose the top candidate model by ranking all models by their evaluated predictive performance in terms of the validation set area under the receiver operator characteristic curve (AUC). The model selection stage using the independent validation fold enabled us to optimize hyperparameters without using test fold data.

## Model Evaluation

In the model evaluation stage, we evaluated the selected best clinical predictive model against the held-out test fold that had not been used for training or model selection to estimate the expected generalization error of the models on previously unseen data. Using this approach, every selected best model from the model selection stage was evaluated exactly once against the test fold.

Using the presented standardized model development, selection, and evaluation pipeline, we compared various types of clinical predictive models in the same test setting with exactly the same amount of hyperparameter optimization and input features against the same test fold. This process enables us to systematically study the expected generalization ability, predictive performance, and influential features of clinical predictive models for predicting SARS-CoV-2 test results, hospital admission for patients who are SARS-CoV-2 positive, and ICU admission for patients who are SARS-CoV-2 positive.

## Experiments

We conducted retrospective experiments to evaluate the predictive performance of a number of clinical predictive models on each of the presented clinical prediction tasks using the standardized development, validation, and evaluation pipeline.



Concretely, our experiments aimed to answer the following questions:

- What is the expected predictive performance of the various clinical predictive models in predicting SARS-CoV-2 test results for presenting patients, hospital admission for patients who are SARS-CoV-2 positive, and ICU admission for patients who are SARS-CoV-2 positive?
- Which clinical, demographic, and blood analysis features were most important for the best encountered predictive models for each clinical prediction task?

### Data Set and Study Cohort

We used anonymized data from a cohort of 5644 patients seen at the Hospital Israelita Albert Einstein in São Paulo, Brazil in the early months of 2020. Exact data collection dates are unknown. The data set is available at [8]. Over the data collection time frame, the rate of patients who were SARS-CoV-2 positive at the hospital was around 10%, of which around 6.5% and 2.5% required hospitalization and critical care, respectively (Table 1). Notably, younger patients were

underrepresented in the SARS-CoV-2 positive group relative to the general patient population, which may have been caused by the reportedly more severe disease progression in older patients [9]. Information on patient sex was not included in our data set. Sex has been reported to be associated with COVID-19 outcomes with men reportedly being at higher risk for severe outcomes, and models including sex as a covariate may, therefore, achieve superior predictive performance [10]. We randomly split the entire available patient cohort into training ( $n=2822$ , 50%), validation ( $n=1129$ , 20%), and test folds ( $n=1693$ , 30%) within strata of patient age, SARS-CoV-2 test result, hospital admission status, and ICU admission status. We performed the stratification by randomly shuffling the entire set of available patients and then assigning a proportional number of patients within the same strata of patient age, SARS-CoV-2 test result, hospital admission status, and ICU admission status to each fold, resulting in three separate folds of the desired target fold sizes that had balanced proportions of the stratification covariates (Table 1). We used the implementation of the described stratification procedure provided in [11] (StratifiedShuffleSplit, package version 0.22.2).

**Table 1.** Training, validation, and test fold statistics for all patients and patients who are SARS-CoV-2 positive.

Property	Training	Validation	Test
<b>All patients</b>			
Patients ( $N=5644$ ), $n$ (%)	2822 (50)	1129 (20)	1693 (30)
SARS-CoV-2 (%)	9.85	9.92	9.92
Admission (%)	1.42	1.33	1.42
ICU <sup>a</sup> (%)	1.59	1.68	1.59
Age (20-quantiles) <sup>b</sup>	9.0 (1.0, 17.0)	9.0 (1.0, 18.0)	9.0 (2.0, 17.0)
<b>Patients who are SARS-CoV-2 positive</b>			
Patients ( $n=558$ ), $n$ (%)	279 (50)	112 (20)	167 (30)
SARS-CoV-2 (%)	100	100	100
Admission (%)	6.45	6.25	6.59
ICU (%)	2.87	2.68	2.99
Age (20-quantiles) <sup>b</sup>	10.0 (4.0, 17.0)	11.5 (4.5, 18.5)	10.0 (4.0, 17.5)

<sup>a</sup>ICU: intensive care unit.

<sup>b</sup>Patient ages are specified in 20-quantiles to maintain patient privacy (10% and 90% percentiles in parentheses).

### Models

Using the presented systematic evaluation methodology, we trained five different model types: logistic regression (LR), neural network (NN), random forest (RF), support vector machine (SVM), and gradient boosting (XGB) [12]. The NN was a multilayer perceptron consisting of  $L$  hidden layers with  $N$  hidden units each followed by a nonlinear activation function (rectified linear unit [13], scaled exponential linear unit [14], or exponential linear unit [15]) and batch normalization [16], and was trained using the Adam optimizer [17] for up to 300 epochs with an early stopping patience of 12 epochs on the validation set loss.

### Hyperparameters

We followed an unbiased, systematic approach to hyperparameter selection and optimization. For each type of clinical predictive model, we performed a maximum of 30 hyperparameter optimization runs with hyperparameters chosen from predefined ranges (Table 2). The performance of each hyperparameter optimization run was evaluated against the validation cohort. After computing the validation set performance, we selected the best candidate predictive model across the 30 hyperparameter optimization runs by AUC for further evaluation against the test set.

**Table 2.** Hyperparameter ranges used for hyperparameter optimization of logistic regression, neural network, random forest, support vector machine, and gradient boosting models for all tasks.

Model and hyperparameter	Range/choices <sup>a</sup>
<b>Logistic regression</b>	
Regularization strength $C$	0.01, 0.1, 1.0, 10.0
<b>Neural network</b>	
Number of hidden units $N$	16, 32, 64, 128
Number of hidden layers $L$	1, 2, 3
Activation $a$	ReLU <sup>b</sup> [13], SELU <sup>c</sup> [14], ELU <sup>d</sup> [15]
Batch size $B$	16, 32, 64, 128
L2 regularization $\lambda_2$	0.0, 0.00001, 0.0001
Learning rate $\alpha$	0.003, 0.03
Dropout percentage $p$	(0%-25%)
<b>Random forest</b>	
Tree depth $D$	3, 4, 5
Number of Trees $T$	32, 64, 128, 256
<b>Support vector machine</b>	
Regularization strength $C$	0.01, 0.1, 1.0, 10.0
Kernel $k$	polynomial, radial basis function, sigmoid
Polynomial degree $d$	3, 5, 7
<b>Gradient boosting</b>	
Subsample ratio $r$	0.25, 0.5, 0.75, 1.0
Max <sup>e</sup> tree depth $T$	2, 3, 4, 5, 6, 7, 8
Min <sup>f</sup> partition loss $\gamma$	0.0, 0.1, 1.0, 10.0
Learning rate $\alpha$	0.003, 0.03, 0.3, 0.5
L1 regularization $\lambda_1$	1.0, 0.1, 0.001, 0.0
L2 regularization $\lambda_2$	1.0, 0.1, 0.001, 0.0
Num <sup>g</sup> boosting rounds $B$	5, 10, 15, 20

<sup>a</sup>Parentheses indicate continuous ranges within the indicated limits sampled uniformly. Comma-delimited lists indicate discrete choices with equal selection probability.

<sup>b</sup>ReLU: rectified linear unit.

<sup>c</sup>SELU: scaled exponential linear unit.

<sup>d</sup>ELU: exponential linear unit.

<sup>e</sup>Max: maximum.

<sup>f</sup>Min: minimum.

<sup>g</sup>Num: number.

## Predictive Performance

To assess the predictive performance of each of the developed clinical predictive models, we evaluated their performance in terms AUC, area under the precision recall curve (AUPR), sensitivity, specificity, and specificity at greater than 95% sensitivity (Spec@95%Sens) on the held-out test set cohorts for each task (Table 1). After model development and hyperparameter optimization, we evaluated each model type exactly once against the test set to calculate the final performance metrics. Operating thresholds for each model were

the operating points on the receiver operator characteristic curve closest to the top left coordinate as calculated for the validation cohort. We chose a variety of complementary evaluation metrics to give a comprehensive picture of the expected performance of each clinical predictive model on the evaluated tasks. For each of the performance metrics, we additionally computed 95% CIs using bootstrap resampling with 100 bootstrap samples on the test set cohort to quantify the uncertainty of our analysis results. We also assessed whether differences between clinical predictive models were statistically significant at significance

level  $\alpha=.05$  using pairwise  $t$  tests with the respective best models for each task as measured by AUC.

### Importance of Test Types

To quantify the importance of specific clinical, demographic, and blood analysis features on each of the predicted outcomes, we used causal explanation (CXPlain) models [18]. CXPlain provides standardized relative feature importance attributions for any predictive model by computing the marginal contribution of each input feature toward the predictive performance of a model [19] and is, therefore, particularly well-suited for assessing feature importance in our diverse set of models. We used the test fold's ground truth labels to compute the exact marginal contribution of each input feature without any estimation uncertainty.

## Results

### Predictive Performance

In terms of predictive performance (Table 3), we found that the overall best identified models by AUC were XGB for predicting SARS-CoV-2 test results, RF for predicting hospital admissions for patients who are SARS-CoV-2 positive, and SVM for predicting ICU admission for patients who are SARS-CoV-2 positive with AUCs of 0.66 (95% CI 0.63-0.70), 0.92 (95% CI 0.81-0.98), and 0.98 (95% CI 0.95-1.00), respectively. Notably, we found that predicting positive SARS-CoV-2 results from routinely collected clinical measurements was a considerably more difficult task for clinical predictive models than predicting hospitalization and ICU admission. Nonetheless, the best encountered clinical predictive model for predicting SARS-CoV-2 test results (XGB) achieved a respectable sensitivity of 75% (95% CI 67%-81%) and specificity of 49% (95% CI 46%-51%). After fixing the operating threshold of the

model to meet a sensitivity level of at least 95% (Spec@95%Sens), the best XGB model for predicting SARS-CoV-2 test results would achieve a specificity of 23% (95% CI 7%-32%). We additionally found that the differences in predictive performance between the best XGB model for predicting SARS-CoV-2 test results and the other predictive models was significant at a prespecified significance level of  $\alpha=.05$  ( $t$  test) for all but the AUPR metric, where NN achieved a significantly better AUPR of 0.22, and the difference to SVM was not significant at the prespecified significance level. On the task of predicting hospital admissions for patients who are SARS-CoV-2 positive, the best encountered RF model achieved a sensitivity of 55% (95% CI 19%-85%), a high specificity of 96% (95% CI 92%-98%), and a Spec@95%Sens of 34% (95% CI 29%-97%). Owing to the lower sample size due to the smaller cohort of patients who are SARS-CoV-2 positive, the performance results for predicting hospital admission generally had wider uncertainty bounds but were nonetheless significantly better for RF than the other predictive models at the prespecified significance level of  $\alpha=.05$  ( $t$  test) for most performance metrics, with the exception of AUC, where XGB achieved an AUC of 0.91, and AUPR, where LR achieved an AUPR of 0.44. On the task of predicting ICU admission for patients who are SARS-CoV-2 positive, SVM had a sensitivity of 80% (95% CI 36%-100%), a specificity of 96% (95% CI 92%-98%), and a Spec@95% Sens of 95% (95% CI 91%-100%). Due to the small percentage of about 3% of patients who were SARS-CoV-2 positive that were admitted to the ICU (Table 1), uncertainty bounds were wider than for the models predicting hospital admissions, and the results of the best encountered SVM were found to be not significantly better than LR and RF in terms of AUC, LR, and NN in terms of sensitivity, and NN in terms of Spec@95%Sens at the prespecified significance level of  $\alpha=.05$  ( $t$  test).

**Table 3.** Comparison of LR, NN, RF, SVM, and XGB models in terms of AUC, AUPR, sensitivity, specificity, and Spec@95% Sens for predicting SARS-CoV-2 test results, hospital admission for patients who are SARS-CoV-2 positive, and intensive care unit admission for patients who are SARS-CoV-2 positive on the test set cohort.

Model	AUC <sup>a</sup> (95% CI) <sup>b</sup>	AUPR <sup>c</sup> (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Spec@95% Sens <sup>d</sup> (95% CI)
<b>SARS-CoV-2 test results</b>					
XGB <sup>e</sup>	0.66 <sup>f</sup> (0.63-0.70)	0.21 (0.15-0.28)	0.75 (0.67-0.81)	0.49 (0.46-0.51)	0.23 (0.07-0.32)
RF <sup>g</sup>	0.65 (0.62-0.69) <sup>h</sup>	0.19 (0.14-0.24) <sup>h</sup>	0.69 (0.61-0.74)	0.54 (0.46-0.57) <sup>h</sup>	0.19 (0.10-0.25) <sup>h</sup>
NN <sup>i</sup>	0.62 (0.57-0.65) <sup>h</sup>	0.22 (0.15-0.28) <sup>h</sup>	0.60 (0.52-0.67) <sup>h</sup>	0.55 (0.46-0.58) <sup>h</sup>	0.17 (0.14-0.28) <sup>h</sup>
LR <sup>j</sup>	0.61 (0.57-0.65) <sup>h</sup>	0.17 (0.13-0.24) <sup>h</sup>	0.58 (0.51-0.65) <sup>h</sup>	0.55 (0.46-0.57) <sup>h</sup>	0.19 (0.16-0.25) <sup>h</sup>
SVM <sup>k</sup>	0.61 (0.57-0.65) <sup>h</sup>	0.21 (0.15-0.27)	0.57 (0.51-0.64) <sup>h</sup>	0.59 (0.56-0.61) <sup>h</sup>	0.14 (0.06-0.16) <sup>h</sup>
<b>Hospital admissions for patients who are SARS-CoV-2 positive</b>					
RF	0.92 (0.81-0.98)	0.43 (0.19-0.81)	0.55 (0.19-0.85)	0.96 (0.92-0.98)	0.34 (0.29-0.97)
XGB	0.91 (0.80-0.98)	0.52 (0.28-0.84) <sup>h</sup>	0.64 (0.43-0.95) <sup>h</sup>	0.94 (0.90-0.97) <sup>h</sup>	0.00 (0.00-0.94) <sup>h</sup>
LR	0.88 (0.70-0.98) <sup>h</sup>	0.44 (0.18-0.83)	0.82 (0.52-1.00) <sup>h</sup>	0.85 (0.79-0.90) <sup>h</sup>	0.13 (0.08-0.93) <sup>h</sup>
NN	0.85 (0.68-0.97) <sup>h</sup>	0.31 (0.13-0.66) <sup>h</sup>	0.64 (0.33-1.00) <sup>h</sup>	0.95 (0.91-0.97) <sup>h</sup>	0.11 (0.06-0.93) <sup>h</sup>
SVM	0.85 (0.70-0.98) <sup>h</sup>	0.35 (0.17-0.77) <sup>h</sup>	0.64 (0.30-1.00) <sup>h</sup>	0.95 (0.91-0.97) <sup>h</sup>	0.21 (0.15-0.96) <sup>h</sup>
<b>Critical care admissions for patients who are SARS-CoV-2 positive</b>					
SVM	0.98 (0.95-1.00)	0.53 (0.14-1.00)	0.80 (0.36-1.00)	0.96 (0.92-0.98)	0.95 (0.91-1.00)
LR	0.98 (0.93-1.00)	0.67 (0.09-1.00) <sup>h</sup>	0.80 (0.29-1.00)	0.93 (0.89-0.96)	0.91 (0.87-1.00) <sup>h</sup>
NN	0.97 (0.94-0.99) <sup>h</sup>	0.35 (0.10-0.88) <sup>h</sup>	0.80 (0.36-1.00)	0.95 (0.91-0.99) <sup>h</sup>	0.94 (0.90-0.99)
RF	0.97 (0.92-1.00)	0.56 (0.13-1.00) <sup>h</sup>	0.60 (0.15-1.00) <sup>h</sup>	0.98 (0.96-1.00) <sup>h</sup>	0.90 (0.86-1.00) <sup>h</sup>
XGB	0.67 (0.53-0.98) <sup>h</sup>	0.29 (0.01-0.68) <sup>h</sup>	0.40 (0.00-1.00) <sup>h</sup>	0.94 (0.91-0.97) <sup>h</sup>	0.00 (0.00-0.96) <sup>h</sup>

<sup>a</sup>AUC: area under the receiver operator characteristic curve.<sup>b</sup>95% CIs obtained via bootstrap resampling with 100 samples.<sup>c</sup>AUPR: area under the precision recall curve.<sup>d</sup>Spec@95% Sens: specificity at greater than 95% sensitivity.<sup>e</sup>XGB: gradient boosting.<sup>f</sup>Italics represent the best results.<sup>g</sup>RF: random forest.<sup>h</sup>Significant at  $P < .05$  ( $t$  test) to the model with the highest predictive performance in terms of AUC.<sup>i</sup>NN: neural network.<sup>j</sup>LR: logistic regression.<sup>k</sup>SVM: support vector machine.

## Feature Importance

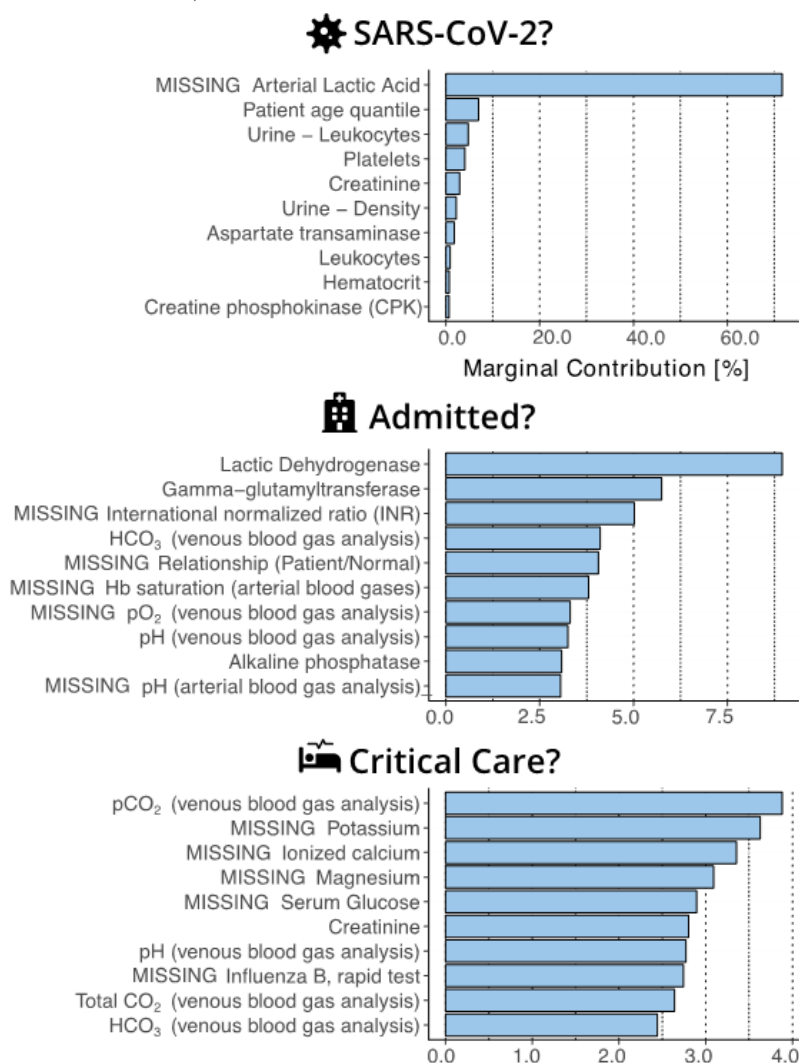
In terms of feature importance, we found that importance scores were distributed highly unequally, relatively uniform, and highly uniform for the best models encountered for predicting SARS-CoV-2 test results, for predicting hospital admissions for patients who are SARS-CoV-2 positive, and for predicting ICU admission, respectively (Figure 3). Most notably, we found that 71.7% of the importance for the best XGB model for predicting SARS-CoV-2 test results was assigned to the missing indicator corresponding to the arterial lactic acid measurement (ie, much of the marginal predictive performance gain of the XGB model was attributed to whether or not the arterial lactic acid test had been ordered). Beyond arterial lactic acid being missing, age, leukocyte count, platelet count [20], and creatinine

[21] were implied to be associated with a positive SARS-CoV-2 test result by the best encountered predictive model, which further substantiates recent independent reports of those factors being potentially associated with SARS-CoV-2 [20-24]. Similar to the best encountered XGB model for predicting SARS-CoV-2 test results, the top encountered predictive models for hospital admission and ICU admission for patients who are SARS-CoV-2 positive assigned a considerable degree of importance to missingness patterns associated with a number of measurements. A possible explanation for missingness appearing as a top predictor across the different tasks is that decisions on whether or not to order a certain test to be performed for a given patient were influenced by patient characteristics that were not captured in the set of clinical measurements that were available to the

predictive models. In the case of the missingness of the lactic acid test being predictive of SARS-CoV-2 test results, the importance could stem either from clinicians judging patients to be more likely to have COVID-19 due to their clinical presentation and, therefore, ordering a lactic acid test to account for potential lactic acidosis due to COVID-19–induced reduced oxygenation levels or from patients that clinicians see as at risk for lactic acidosis being likely to have their symptoms caused by an underlying SARS-CoV-2 infection. A controlled setting with standardized testing guidelines would be required to determine which confounding factors are behind the predictive power of the missingness patterns that have been implied to be associated with COVID-19 by the predictive models. Beyond

missingness patterns, top predictors for predicting hospital admission were lactate dehydrogenase [25]; gamma-glutamyl transferase, which through abnormal liver function has been reported to be implicated in COVID-19 severity [26]; and  $HCO_3$  [27]. For predicting ICU admission in patients who are SARS-CoV-2 positive,  $pCO_2$ , creatinine [21], and pH [23] were top predictors. Blood pH, and in particular respiratory alkalosis, has been reported to be associated with severe COVID-19 [28]. We note that several factors that were not included in our study have recently been reported to be implicated in COVID-19 outcomes, such as the number of ICU beds available at a hospital [29], patients' racial and ethnic backgrounds [30], and several pre-existing conditions [31].

**Figure 3.** A comparison of the top 10 features ranked by relative feature importance scores for the best-encountered model for predicting SARS-CoV-2 test results (gradient boosting, top), hospital admissions (random forest, middle), and critical care admission for patients who are SARS-CoV-2 positive (support vector machine, bottom), respectively. The bar length corresponds to the relative marginal importance (in %) of the displayed features toward the predictive performance of the respective model. Feature names that include “MISSING” indicate that the given marginal contribution refers to the importance of the presence of that feature's absence, not the feature itself.



## Discussion

### Principal Findings

We presented a systematic study of predictive models that predict SARS-CoV-2 test results, hospital admission for patients who are SARS-CoV-2 positive, and ICU admission for patients

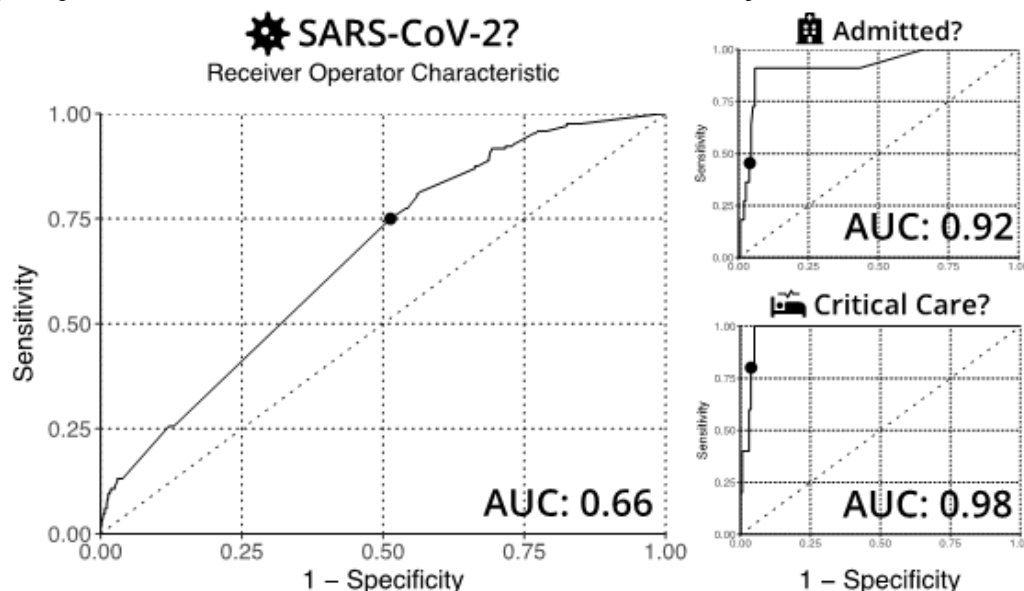
who are SARS-CoV-2 positive using routinely collected clinical measurements. Models that predict SARS-CoV-2 test results could help prioritize scarce testing capacity by identifying those individuals that are more likely to receive a positive result. Similarly, predictive models that predict which patient who is SARS-CoV-2 positive would be most likely to require hospital and critical care beds could help better use existing hospital



capacity by prioritizing those patients that have the highest risk of deterioration. Facilitating the efficient use of scarce health care resources is particularly important in dealing with

SARS-CoV-2, as its rapid transmission significantly increases demand for health care services worldwide (Figure 4).

**Figure 4.** Receiver operator characteristic curves for the best-encountered model for predicting SARS-CoV-2 test results (gradient boosting, left), hospital admissions for patients who are SARS-CoV-2 positive (random forest, top right), and critical care admissions for patients who are SARS-CoV-2 positive (support vector machine, bottom right). Numbers in the bottom right of each subgraph show the respective model's AUC. Solid dots on the curves indicate operating thresholds selected on the validation fold. AUC: Area under the receiver operator characteristic curve.



## Limitations

The main limitation of this study is that its experimental evaluation was based on data collected from a single study site, and its results may, therefore, not generalize to settings with significantly different patient populations, admission criteria, patterns of missingness, and testing guidelines. Operationally, to ensure robustness, it is important to check for any significant deviations in terms of patterns observed in the training cohort when attempting to transfer predictive models trained in one context to another (ie, when transferring a model to another hospital). In case any significant deviations are detected, fine-tuning the predictive model for the new target context is strongly advised. In addition, we did not have access to mortality data for the analyzed cohort, and we were, therefore, not able to correlate our predicted individual risk scores with patient mortality, which is another related prediction task that may be of clinical importance. Future studies should include a broader set of clinical measurements and outcomes, cohorts from multiple distinct geographical sites, and under varying patterns of missingness to determine the robustness of the clinical predictive models to these confounding factors. Finally, we believe that the inclusion of data from other modalities such as genomic profiling and medical imaging, and data on comorbidities, symptoms, and treatment histories could potentially further improve predictive performance of clinical predictive models across the presented prediction tasks.

## Comparison With Prior Work

A substantial body of work is dedicated to the study, validation, and implementation of predictive models for clinical tasks. Clinical predictive models have, for example, been used to predict risk of septic shock [32,33], risk of heart failure [34],

readmission following heart failure [35-37], false alarms in critical care [38], risk scores [39], outcomes [40] and mortality in pneumonia [41,42], and mortality risk in critical care [43-45]. Predicting clinical outcomes for individual patients is difficult because many confounding factors may influence patient outcomes, and collecting and accounting for these factors in an unbiased way remains an open challenge in clinical practice [46]. Systematic studies such as this paper enable medical practitioners to better understand, assess, and potentially overcome these issues by systematically evaluating generalization ability, expected predictive performance, and influential predictors of various clinical predictive models. Beyond the need for systematic evaluation, missingness [47-50], noise [51,52], multivariate input data [38,53-55], and the need for interpretability [18,56-58] have been highlighted as particularly important considerations in health care settings. In this study, we build on recent methodological advances to develop and systematically study clinical predictive models that may aid in prioritizing health care resources [59] for COVID-19 and, thereby, help prevent a potential overextension of health care system capacity.

## Clinical Predictive Models for COVID-19

Several clinical predictive models have recently been proposed for COVID-19, for example, for predicting potential COVID-19 diagnoses using data from emergency care admission exams [60] and chest imaging data [61-66], for predicting COVID-19-related mortality from clinical risk factors [67,68], for predicting which patients will develop acute respiratory distress syndrome from patients' clinical characteristics [69], for predicting critical illness in patients with COVID-19 [70,71], and for predicting progression risk in patients with COVID-19 pneumonia [72]. Siordia [73] presented a review of

epidemiology and clinical features associated with COVID-19, and Wynants et al [74] performed a critical review that assessed limitations and risk of bias in diagnostic and prognostic models for COVID-19. In addition, Wang et al [23] performed a cohort study for clinical and laboratory predictors of COVID-19–related in-hospital mortality that identified baseline neutrophil count, age, and several other clinical features as top predictors of mortality. Beyond prediction, Ienca and Vayena [75] have argued for the responsible use of data in tackling the challenges posed by SARS-CoV-2.

Owing to the recent emergence of SARS-CoV-2, there currently exists, to the best of our knowledge, no prior systematic study on clinical predictive models that predict the likelihood of a positive SARS-CoV-2 test and hospital and ICU admission from clinical, demographic, and blood analysis data that accounts for the missingness that is characteristic for the clinical setting. We additionally assessed the influence of various clinical, demographic, and blood analysis measurements on the predictions of the developed clinical predictive models.

## Conclusions

We present a systematic study in which we developed and evaluated clinical predictive models for COVID-19 that estimate

the likelihood of a positive SARS-CoV-2 test in patients presenting at hospitals and the likelihood of hospital admission and ICU admission in patients who are SARS-CoV-2 positive. We evaluated our developed clinical predictive models in a retrospective evaluation using a cohort of 5644 hospital patients seen in São Paulo, Brazil. In addition, we determined the clinical, demographic, and blood analysis measurements that were most important for accurately predicting SARS-CoV-2 status, hospital admissions, and ICU admissions. Our experimental results indicate that clinical predictive models may in the future potentially be used to inform care and help prioritize scarce health care resources by assigning personalized risk scores for individual patients using routinely collected clinical, demographic, and blood analysis data. Furthermore, our findings on the importance of routine clinical measurements toward predicting clinical pathways for patients increases our understanding of the interrelations of individual risk profiles and outcomes in SARS-CoV-2. Based on our study's results, we conclude that health care systems should explore the use of predictive models that assess individual COVID-19 risk to improve health care resource prioritization and inform patient care.

## Acknowledgments

The anonymized data used in this paper were generously contributed by patients at Hospital Israelita Albert Einstein in São Paulo, Brazil and are freely available at [8].

## Conflicts of Interest

PS is an employee and shareholder of F Hoffmann-La Roche Ltd.

## Multimedia Appendix 1

Demographic, clinical, and blood analysis data used as features by our systematic model development and evaluation pipeline, and their respective value ranges in the data set.

[DOCX File, 41 KB - [jmir\\_v22i10e21439\\_app1.docx](https://www.jmir.org/2020/10/e21439_app1.docx)]

## References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. 2020. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> [accessed 2020-05-01]
2. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: [10.1126/science.aba9757](https://doi.org/10.1126/science.aba9757)] [Medline: [32144116](https://pubmed.ncbi.nlm.nih.gov/32144116/)]
3. Jernigan DB, CDC COVID-19 Response Team. Update: public health response to the coronavirus disease 2019 outbreak - United States, February 24, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Feb 28;69(8):216-219. [doi: [10.15585/mmwr.mm6908e1](https://doi.org/10.15585/mmwr.mm6908e1)] [Medline: [32106216](https://pubmed.ncbi.nlm.nih.gov/32106216/)]
4. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int J Infect Dis* 2020 Apr;93:211-216 [FREE Full text] [doi: [10.1016/j.ijid.2020.02.058](https://doi.org/10.1016/j.ijid.2020.02.058)] [Medline: [32145465](https://pubmed.ncbi.nlm.nih.gov/32145465/)]
5. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. *JAMA* 2020 Apr 14;323(14):1341-1342. [doi: [10.1001/jama.2020.3151](https://doi.org/10.1001/jama.2020.3151)] [Medline: [32125371](https://pubmed.ncbi.nlm.nih.gov/32125371/)]
6. Lee VJ, Chiew CJ, Khong WX. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. *J Travel Med* 2020 May 18;27(3):taaa039 [FREE Full text] [doi: [10.1093/jtm/taaa039](https://doi.org/10.1093/jtm/taaa039)] [Medline: [32167146](https://pubmed.ncbi.nlm.nih.gov/32167146/)]
7. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011 Feb 20;30(4):377-399. [doi: [10.1002/sim.4067](https://doi.org/10.1002/sim.4067)] [Medline: [21225900](https://pubmed.ncbi.nlm.nih.gov/21225900/)]
8. Einstein Data4u. Diagnosis of COVID-19 and its clinical spectrum AI and Data Science supporting clinical decisions (from 28th Mar to 3st Apr). Kaggle. URL: <https://www.kaggle.com/einsteindata4u/covid19> [accessed 2020-04-01]

9. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet* 2020 Apr;395(10231):1225-1228. [doi: [10.1016/s0140-6736\(20\)30627-9](https://doi.org/10.1016/s0140-6736(20)30627-9)]
10. Klein SL, Dhakal S, Ursin RL, Deshpande S, Sandberg K, Mauvais-Jarvis F. Biological sex impacts COVID-19 outcomes. *PLoS Pathog* 2020 Jun;16(6):e1008570 [FREE Full text] [doi: [10.1371/journal.ppat.1008570](https://doi.org/10.1371/journal.ppat.1008570)] [Medline: [32569293](https://pubmed.ncbi.nlm.nih.gov/32569293/)]
11. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Machine Learning Res* 2011 Oct;12:2825-2830. [doi: [10.1145/2786984.2786995](https://doi.org/10.1145/2786984.2786995)]
12. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August, 2016; San Francisco, California. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
13. Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. 2010 Presented at: International Conference on Machine Learning; June 21-24, 2010; Haifa, Israel.
14. Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-normalizing neural networks. 2017 Presented at: Neural Information Processing Systems 2017; Dec 2017; Long Beach, CA.
15. Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). *arXiv* 2015 [FREE Full text]
16. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015 Presented at: 32nd International Conference on International Conference on Machine Learning; 2015; Lille, France.
17. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2015 Presented at: International Conference on Learning Representations; 2015; San Diego, California.
18. Schwab P, Karlen W. CXPlain: causal explanations for model interpretation under uncertainty. 2019 Presented at: Neural Information Processing Systems 2019; Vancouver, BC.
19. Schwab P, Miladinovic D, Karlen W. Granger-causal attentive mixtures of experts: learning important features with neural networks. *AAAI* 2019 Jul 17;33:4846-4853. [doi: [10.1609/aaai.v33i01.33014846](https://doi.org/10.1609/aaai.v33i01.33014846)]
20. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis. *Clin Chim Acta* 2020 Jul;506:145-148 [FREE Full text] [doi: [10.1016/j.cca.2020.03.022](https://doi.org/10.1016/j.cca.2020.03.022)] [Medline: [32178975](https://pubmed.ncbi.nlm.nih.gov/32178975/)]
21. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* 2020 May;97(5):829-838. [doi: [10.1016/j.kint.2020.03.005](https://doi.org/10.1016/j.kint.2020.03.005)] [Medline: [32247631](https://pubmed.ncbi.nlm.nih.gov/32247631/)]
22. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun* 2020 May;109:102433. [doi: [10.1016/j.jaut.2020.102433](https://doi.org/10.1016/j.jaut.2020.102433)] [Medline: [32113704](https://pubmed.ncbi.nlm.nih.gov/32113704/)]
23. Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, et al. Clinical and laboratory predictors of in-hospital mortality in 305 patients with COVID-19: a cohort study in Wuhan, China. *SSRN J* 2020 Mar 05:1. [doi: [10.2139/ssrn.3546115](https://doi.org/10.2139/ssrn.3546115)]
24. CDC COVID-19 Response Team. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) - United States, February 12-March 16, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Mar 27;69(12):343-346. [doi: [10.15585/mmwr.mm6912e2](https://doi.org/10.15585/mmwr.mm6912e2)] [Medline: [32214079](https://pubmed.ncbi.nlm.nih.gov/32214079/)]
25. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020 Mar;395(10229):1054-1062. [doi: [10.1016/s0140-6736\(20\)30566-3](https://doi.org/10.1016/s0140-6736(20)30566-3)]
26. Fan Z, Chen L, Li J, Cheng X, Yang J, Tian C, et al. Clinical features of COVID-19-related liver functional abnormality. *Clin Gastroenterol Hepatol* 2020 Jun;18(7):1561-1566 [FREE Full text] [doi: [10.1016/j.cgh.2020.04.002](https://doi.org/10.1016/j.cgh.2020.04.002)] [Medline: [32283325](https://pubmed.ncbi.nlm.nih.gov/32283325/)]
27. Dondorp AM, Hayat M, Aryal D, Beane A, Schultz MJ. Respiratory support in COVID-19 patients, with a focus on resource-limited settings. *Am J Trop Med Hyg* 2020 Jun;102(6):1191-1197. [doi: [10.4269/ajtmh.20-0283](https://doi.org/10.4269/ajtmh.20-0283)] [Medline: [32319424](https://pubmed.ncbi.nlm.nih.gov/32319424/)]
28. Ashraf MA, Shokouhi N, Shirali E, Davari-tanha F, Memar O, Kamalipour A, et al. COVID-19 in Iran, a comprehensive investigation from exposure to treatment outcomes. *medRxiv* 2020 Apr 24:1. [doi: [10.1101/2020.04.20.20072421](https://doi.org/10.1101/2020.04.20.20072421)]
29. Gupta S, Hayek SS, Wang W, Chan L, Mathews KS, Melamed ML, STOP-COVID Investigators. Factors associated with death in critically ill patients with coronavirus disease 2019 in the US. *JAMA Intern Med* 2020 Jul 15:e203596. [doi: [10.1001/jamainternmed.2020.3596](https://doi.org/10.1001/jamainternmed.2020.3596)] [Medline: [32667668](https://pubmed.ncbi.nlm.nih.gov/32667668/)]
30. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. *JAMA* 2020 Jun 23;323(24):2466-2467. [doi: [10.1001/jama.2020.8598](https://doi.org/10.1001/jama.2020.8598)] [Medline: [32391864](https://pubmed.ncbi.nlm.nih.gov/32391864/)]
31. Atkins JL, Masoli JAH, Delgado J, Pilling LC, Kuo CL, Kuchel GA, et al. Preexisting comorbidities predicting COVID-19 and mortality in the UK Biobank community cohort. *J Gerontol A Biol Sci Med Sci* 2020 Jul 20:glaa183. [doi: [10.1093/gerona/glaa183](https://doi.org/10.1093/gerona/glaa183)] [Medline: [32687551](https://pubmed.ncbi.nlm.nih.gov/32687551/)]
32. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS One* 2017;12(4):e0174708 [FREE Full text] [doi: [10.1371/journal.pone.0174708](https://doi.org/10.1371/journal.pone.0174708)] [Medline: [28384212](https://pubmed.ncbi.nlm.nih.gov/28384212/)]
33. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015 Aug 05;7(299):299ra122. [doi: [10.1126/scitranslmed.aab3719](https://doi.org/10.1126/scitranslmed.aab3719)] [Medline: [26246167](https://pubmed.ncbi.nlm.nih.gov/26246167/)]

34. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. *Med Care* 2010;48:S106-S113. [doi: [10.1097/mlr.0b013e3181de9e17](https://doi.org/10.1097/mlr.0b013e3181de9e17)]
35. Frizzell JD, Liang L, Schulte PJ, Yancy CW, Heidenreich PA, Hernandez AF, et al. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA Cardiol* 2017 Feb 01;2(2):204-209. [doi: [10.1001/jamacardio.2016.3956](https://doi.org/10.1001/jamacardio.2016.3956)] [Medline: [27784047](https://pubmed.ncbi.nlm.nih.gov/27784047/)]
36. Shameer K, Johnson KW, Yahi A, Miotto R, Li LI, Ricks D, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using Mount Sinai heart failure cohort. *Pac Symp Biocomput* 2017;22:276-287 [FREE Full text] [doi: [10.1142/9789813207813\\_0027](https://doi.org/10.1142/9789813207813_0027)] [Medline: [27896982](https://pubmed.ncbi.nlm.nih.gov/27896982/)]
37. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak* 2018 Jun 22;18(1):44 [FREE Full text] [doi: [10.1186/s12911-018-0620-z](https://doi.org/10.1186/s12911-018-0620-z)] [Medline: [29929496](https://pubmed.ncbi.nlm.nih.gov/29929496/)]
38. Schwab P, Keller E, Muroi C, Mack DJ, Strässle C, Karlen W. Not to cry wolf: distantly supervised multitask learning in critical care. 2018 Presented at: Thirty-fifth International Conference on Machine Learning; 2018; Stockholm, Sweden.
39. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. 2015 Presented at: KDD '15: The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 2015; Sydney, NSW, Australia. [doi: [10.1145/2783258.2788613](https://doi.org/10.1145/2783258.2788613)]
40. Visweswaran S, Cooper GF. Patient-specific models for predicting the outcomes of patients with community acquired pneumonia. *AMIA Annu Symp Proc* 2005:759-763. [Medline: [16779142](https://pubmed.ncbi.nlm.nih.gov/16779142/)]
41. Cooper GF, Aliferis CF, Ambrosino R, Aronis J, Buchanan BG, Caruana R, et al. An evaluation of machine-learning methods for predicting pneumonia mortality. *Artif Intelligence Med* 1997 Feb;9(2):107-138. [doi: [10.1016/s0933-3657\(96\)00367-3](https://doi.org/10.1016/s0933-3657(96)00367-3)]
42. Wu C, Rosenfeld R, Clermont G. Using data-driven rules to predict mortality in severe community acquired pneumonia. *PLoS One* 2014;9(4):e89053 [FREE Full text] [doi: [10.1371/journal.pone.0089053](https://doi.org/10.1371/journal.pone.0089053)] [Medline: [24699007](https://pubmed.ncbi.nlm.nih.gov/24699007/)]
43. Clermont G, Angus DC, DiRusso SM, Griffin M, Linde-Zwirble WT. Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models. *Crit Care Med* 2001 Feb;29(2):291-296. [doi: [10.1097/00003246-200102000-00012](https://doi.org/10.1097/00003246-200102000-00012)] [Medline: [11246308](https://pubmed.ncbi.nlm.nih.gov/11246308/)]
44. Ghassemi M, Naumann T, Doshi-Velez F, Brimmer N, Joshi R, Rumshisky AA, et al. Unfolding physiological state: mortality modelling in intensive care units. 2014 Presented at: KDD '14: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 2014; New York, NY. [doi: [10.1145/2623330.2623742](https://doi.org/10.1145/2623330.2623742)]
45. Johnson AEW, Pollard TJ, Mark RG. Reproducibility in critical care: a mortality prediction case study. 2017 Presented at: 2nd Machine Learning for Healthcare Conference; August 18-19, 2017; Toronto, ON.
46. Obermeyer Z, Emanuel EJ. Predicting the future — big data, machine learning, and clinical medicine. *N Engl J Med* 2016 Sep 29;375(13):1216-1219. [doi: [10.1056/nejmp1606181](https://doi.org/10.1056/nejmp1606181)]
47. Faris PD, Ghali WA, Brant R, Norris CM, Galbraith P, Knudtson ML. Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses. *J Clin Epidemiol* 2002 Feb;55(2):184-191. [doi: [10.1016/s0895-4356\(01\)00433-4](https://doi.org/10.1016/s0895-4356(01)00433-4)]
48. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for handling missing data in electronic health record derived data. *EGEMS (Wash DC)* 2013;1(3):1035. [doi: [10.13063/2327-9214.1035](https://doi.org/10.13063/2327-9214.1035)] [Medline: [25848578](https://pubmed.ncbi.nlm.nih.gov/25848578/)]
49. Lipton ZC, Kale DC, Wetzel R. Modeling missing data in clinical time series with RNNs. 2016 Presented at: Machine Learning for Healthcare Conference; August 19-20, 2016; Los Angeles, CA.
50. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. *Sci Rep* 2018 Apr 17;8(1):6085. [doi: [10.1038/s41598-018-24271-9](https://doi.org/10.1038/s41598-018-24271-9)] [Medline: [29666385](https://pubmed.ncbi.nlm.nih.gov/29666385/)]
51. Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PLoS One* 2013;8(6):e66341 [FREE Full text] [doi: [10.1371/journal.pone.0066341](https://doi.org/10.1371/journal.pone.0066341)] [Medline: [23826094](https://pubmed.ncbi.nlm.nih.gov/23826094/)]
52. Schwab P, Scebbba GC, Zhang J, Delai M, Karlen W. Beat by beat: classifying cardiac arrhythmias with recurrent neural networks. *Computing Cardiol* 2017;1. [doi: [10.22489/cinc.2017.363-223](https://doi.org/10.22489/cinc.2017.363-223)]
53. Schwab P, Karlen W. PhoneMD: learning to diagnose Parkinson's disease from smartphone data. *AAAI* 2019 Jul 17;33:1118-1125. [doi: [10.1609/aaai.v33i01.33011118](https://doi.org/10.1609/aaai.v33i01.33011118)]
54. Schwab P, Karlen W. A deep learning approach to diagnosing multiple sclerosis from smartphone data. *IEEE J Biomed Health Inform* 2019 Jul 17:1-1. [doi: [10.1109/jbhi.2020.3021143](https://doi.org/10.1109/jbhi.2020.3021143)]
55. Ghassemi M, Pimentel MAF, Naumann T, Brennan T, Clifton DA, Szolovits P, et al. A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data. *Proc Conf AAAI Artif Intell* 2015 Jan;2015:446-453 [FREE Full text] [Medline: [27182460](https://pubmed.ncbi.nlm.nih.gov/27182460/)]
56. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W. RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism. 2016 Presented at: Neural Information Processing Systems; 2016; Barcelona, Spain.
57. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *arXiv* 2017 Feb 28 [FREE Full text]



58. Ross AS, Hughes MC, Doshi-Velez F. Right for the right reasons: training differentiable models by constraining their explanations. *Proc Twenty-Sixth Int Joint Conference Artificial Intelligence 2017*:2662-2670. [doi: [10.24963/ijcai.2017/371](https://doi.org/10.24963/ijcai.2017/371)]
59. Chen JH, Asch SM. Machine learning and prediction in medicine — beyond the peak of inflated expectations. *N Engl J Med* 2017 Jun 29;376(26):2507-2509. [doi: [10.1056/nejmp1702071](https://doi.org/10.1056/nejmp1702071)]
60. de Moraes Batista A, Miraglia JL, Donato THR, Filho ADPC. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. *medRxiv* 2020 Apr 14:1. [doi: [10.1101/2020.04.04.20052092](https://doi.org/10.1101/2020.04.04.20052092)]
61. Wang L, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. *arXiv* 2020 Mar 22 (e-print ahead of publication) [[FREE Full text](#)]
62. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks. *arXiv* 2020 Mar 24 (e-print ahead of publication) [[FREE Full text](#)]
63. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology* 2020 Aug;296(2):E65-E71. [doi: [10.1148/radiol.2020200905](https://doi.org/10.1148/radiol.2020200905)] [Medline: [32191588](https://pubmed.ncbi.nlm.nih.gov/32191588/)]
64. Castiglioni I, Ippolito D, Interlenghi M, Monti CB, Salvatore C, Schiaffino S, et al. Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: a first experience from Lombardy, Italy. *medRxiv* 2020 Apr 10:1. [doi: [10.1101/2020.04.08.20040907](https://doi.org/10.1101/2020.04.08.20040907)]
65. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *medRxiv* 2020 Apr 24:1. [doi: [10.1101/2020.02.14.20023028](https://doi.org/10.1101/2020.02.14.20023028)]
66. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, et al. Deep learning system to screen coronavirus disease 2019 pneumonia. *arXiv* 2020 (e-print ahead of publication) [[FREE Full text](#)]
67. Sarkar J, Chakrabarti P. A machine learning model reveals older age and delayed hospitalization as predictors of mortality in patients with COVID-19. *medRxiv* 2020 Mar 30:1. [doi: [10.1101/2020.03.25.20043331](https://doi.org/10.1101/2020.03.25.20043331)]
68. Yan L, Zhang HT, Goncalves J, Xiao Y, Wang M, Guo Y, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. *medRxiv* 2020 Mar 17:1. [doi: [10.1101/2020.02.27.20028027](https://doi.org/10.1101/2020.02.27.20028027)]
69. Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, et al. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Comput Materials Continua* 2020;63(1):537-551. [doi: [10.32604/cmc.2020.010691](https://doi.org/10.32604/cmc.2020.010691)]
70. Wu G, Yang P, Xie Y, Woodruff HC, Rao X, Guiot J, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. *Eur Respir J* 2020 Aug;56(2):2001104. [doi: [10.1183/13993003.01104-2020](https://doi.org/10.1183/13993003.01104-2020)] [Medline: [32616597](https://pubmed.ncbi.nlm.nih.gov/32616597/)]
71. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med* 2020 Aug 01;180(8):1081-1089 [[FREE Full text](#)] [doi: [10.1001/jamainternmed.2020.2033](https://doi.org/10.1001/jamainternmed.2020.2033)] [Medline: [32396163](https://pubmed.ncbi.nlm.nih.gov/32396163/)]
72. Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, et al. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. *Clin Infect Dis* 2020 Sep 12;71(6):1393-1399 [[FREE Full text](#)] [doi: [10.1093/cid/ciaa414](https://doi.org/10.1093/cid/ciaa414)] [Medline: [32271369](https://pubmed.ncbi.nlm.nih.gov/32271369/)]
73. Siordia JA. Epidemiology and clinical features of COVID-19: a review of current literature. *J Clin Virol* 2020 Jun;127:104357 [[FREE Full text](#)] [doi: [10.1016/j.jcv.2020.104357](https://doi.org/10.1016/j.jcv.2020.104357)] [Medline: [32305884](https://pubmed.ncbi.nlm.nih.gov/32305884/)]
74. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020 Apr 07;369:m1328. [doi: [10.1136/bmj.m1328](https://doi.org/10.1136/bmj.m1328)] [Medline: [32265220](https://pubmed.ncbi.nlm.nih.gov/32265220/)]
75. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat Med* 2020 Apr;26(4):463-464 [[FREE Full text](#)] [doi: [10.1038/s41591-020-0832-5](https://doi.org/10.1038/s41591-020-0832-5)] [Medline: [32284619](https://pubmed.ncbi.nlm.nih.gov/32284619/)]

## Abbreviations

**AUC:** area under the receiver operator characteristic curve  
**AUPR:** area under the precision recall curve  
**CXPlain:** causal explanation  
**ICU:** intensive care unit  
**LR:** logistic regression  
**MICE:** multiple imputation by chained equations  
**NN:** neural network  
**RF:** random forest  
**Spec@95%Sens:** specificity at greater than 95% sensitivity  
**SVM:** support vector machine  
**XGB:** gradient boosting



*Edited by G Eysenbach; submitted 15.06.20; peer-reviewed by A Davoudi, T Aslanidis; comments to author 07.08.20; revised version received 30.08.20; accepted 14.09.20; published 06.10.20.*

*Please cite as:*

Schwab P, DuMont Schütte A, Dietz B, Bauer S

Clinical Predictive Models for COVID-19: Systematic Study

J Med Internet Res 2020;22(10):e21439

URL: <http://www.jmir.org/2020/10/e21439/>

doi: [10.2196/21439](https://doi.org/10.2196/21439)

PMID: [32976111](https://pubmed.ncbi.nlm.nih.gov/32976111/)

©Patrick Schwab, August DuMont Schütte, Benedikt Dietz, Stefan Bauer. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# System-Wide Accelerated Implementation of Telemedicine in Response to COVID-19: Mixed Methods Evaluation

Diego Garcia-Huidobro<sup>1</sup>, MD, PhD; Solange Rivera<sup>1</sup>, MD, MSc; Sebastián Valderrama Chang<sup>2,3</sup>, MD; Paula Bravo<sup>3</sup>, BSc; Daniel Capurro<sup>2,4,5</sup>, MD, PhD

<sup>1</sup>Department of Family Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile

<sup>2</sup>Department of Internal Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile

<sup>3</sup>Office of Innovation and Digital Transformation, UC Christus Health Network, Santiago, Chile

<sup>4</sup>School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

<sup>5</sup>Centre for Digital Transformation of Health, University of Melbourne, Melbourne, Australia

**Corresponding Author:**

Diego Garcia-Huidobro, MD, PhD

Department of Family Medicine

Pontificia Universidad Catolica de Chile

Vicuna Mackenna 4686

Macul

Santiago,

Chile

Phone: 56 930233407

Email: [dngarcih@uc.cl](mailto:dngarcih@uc.cl)

## Abstract

**Background:** As the COVID-19 pandemic disrupted medical practice, telemedicine emerged as an alternative to outpatient visits. However, it is not known how patients and physicians responded to an accelerated implementation of this model of medical care.

**Objective:** The aim of this study is to report the system-wide accelerated implementation of telemedicine, compare patient satisfaction between telemedicine and in-person visits, and report provider perceptions.

**Methods:** This study was conducted at the UC Christus Health Network, a large private academic health network in Santiago, Chile. The satisfaction of patients receiving telemedicine care in March and April 2020 was compared to those receiving in-person care during the same period (concurrent control group) as well as in March and April 2019 (retrospective control group). Patient satisfaction with in-person care was measured using the Net Promoter Score (NPS) survey. Patient satisfaction with telemedicine was assessed with an online survey assessing similar domains. Providers rated their satisfaction and responded to open-ended questions assessing challenges, strategies used to address challenges, the diagnostic process, treatment, and the patient-provider relationship.

**Results:** A total of 3962 patients receiving telemedicine, 1187 patients from the concurrent control group, and 1848 patients from the retrospective control group completed the surveys. Satisfaction was very high with both telemedicine and in-person services. Overall, 263 physicians from over 41 specialties responded to the survey. During telemedicine visits, most providers felt their clinical skills were challenged (61.8%). Female providers felt more challenged than male providers (70.7% versus 50.9%,  $P=.002$ ). Surgeons, obstetricians, and gynecologists felt their clinical skills were challenged the least, compared to providers from nonsurgical specialties ( $P<.001$ ). Challenges related to the delivery modality, diagnostic process, and patient-provider relationship differed by provider specialty ( $P=.046$ ,  $P<.001$ , and  $P=.02$ , respectively).

**Conclusions:** Telemedicine implemented in response to the COVID-19 pandemic produced high patient and provider satisfaction. Specialty groups perceived the impact of this new mode of clinical practice differently.

(*J Med Internet Res* 2020;22(10):e22146) doi:[10.2196/22146](https://doi.org/10.2196/22146)

**KEYWORDS**

telemedicine; telehealth; virtual medicine; health services evaluation; COVID-19

## Introduction

COVID-19 is the largest pandemic of the current century. It has generated transformative changes in health care, including the prioritization of emergency departments and intensive care units [1]. In addition, with physical distancing and isolation as major strategies to contain the spread of the disease, there are fewer patients receiving outpatient clinical care [2]. Although reducing in-person encounters is needed to contain the transmission of SARS-CoV-2, medical care is still needed for acute and chronic conditions.

In this context, telemedicine has emerged as a necessary clinical innovation to provide patient care [3]. This delivery modality has demonstrated clinical effectiveness, optimization in the delivery of health services, usability, and high patient and provider satisfaction [4,5]. Although appealing, several factors may hinder the sustainable use of telemedicine. First, providers prefer in-person encounters to virtual ones [6]. Potential concerns include uncertainty in the clinical encounter, difficulties performing a physical examination, and concerns with the patient-provider relationship [7]. These difficulties may reduce a clinician's motivation to deliver virtual services. In addition, health insurance agencies have not always recognized this type of clinical service to be reimbursable, limiting the scaling up of telemedicine. Even though telemedicine has existed for several years and has had several implementation barriers, the COVID-19 pandemic has quickly transformed the practice of ambulatory medicine [8,9] and the willingness of health care organizations to implement this mode of delivery. The effect of rapidly deploying new models of care is still incompletely understood.

In this context, the UC Christus Health Network in Santiago, Chile, implemented telemedicine services two weeks after the first patient with COVID-19 was diagnosed in the country. The purpose of this paper is to report the system-wide accelerated implementation of telemedicine, compare patient satisfaction with telemedicine and in-person visits, and report provider perceptions regarding the implementation of virtual visits.

## Methods

### Design

This study used a convergent parallel mixed methods evaluation. In this study design, quantitative and qualitative methods have similar relevance, the data are collected in the same phase of the research project, the analysis of the data is independent, and the findings are combined and interpreted together [10]. We chose this design to better understand participants' experiences. Quantitative data consisted of in-person and telemedicine visits, patient demographics, satisfaction, and challenges identified by clinicians. Data from patients receiving telemedicine care were compared to that of patients receiving in-person care between March 1 and April 30, 2019 (retrospective control group), and March 1 and April 30, 2020 (concurrent control group). Qualitative data included physicians' responses to open-ended questions asking about challenges related to using telemedicine and mechanisms to address those challenges when delivering virtual health services. This study was approved by the

Institutional Review Board of the Pontificia Universidad Católica de Chile.

### Setting

Chile's health system combines public and private health providers. Public health services are paid for by public health insurance. Private health providers are paid per service delivered, and health insurance pays some or all of the fees charged by the health care providers.

This study was conducted at the UC Christus Health Network, a large private academic health network in Santiago, Chile. This health network has two hospitals with nearly 500 beds in total and 6 primary and secondary care clinics, with more than 800,000 patient visits every year. UC Christus has been nationally recognized for its innovation and patient service with the "Best Place to Innovate," "Praxis Xperience Index," and "ProCalidad" Awards.

### Telemedicine and its Implementation

UC Christus Health started implementing telemedicine on March 17, 2020, about two weeks after the first patient with COVID-19 was diagnosed in Chile. Initially, it started with generalist specialties, including internal medicine and family medicine, which offered services and guidance for patients with respiratory symptoms. Telemedicine was subsequently implemented as a strategy to provide care across the entire health care system. Currently, more than 720 providers from 61 clinical specialties are providing patient care through telemedicine.

To access these encounters, patients schedule appointments online or over the phone. When scheduling the appointment, patients must accept the terms of telemedicine, including that providers cannot provide certain services virtually, such as ordering restricted prescriptions or providing official health certificates. Once the appointment is scheduled, patients receive a confirmation email and are invited to pay for their visit. The cost of this service is about US \$50, representing about two-thirds of the cost for in-person visits. Once the visit is paid for, patients and providers receive a Zoom link (Zoom Video Communications Inc) for their telemedicine visit, and they connect at the scheduled time. An administrative team supports patient-provider connections. Providers have access to the patient's electronic medical record and use a separate platform to order prescriptions, laboratory exams, images, and procedures, which are usually written manually during in-person visits. These electronic orders are emailed automatically to patients from the medication and laboratory platform.

### Outcomes and Data

#### Overview

Primary outcomes were patient and provider satisfaction along with providers' challenges in implementing telemedicine. This study used four databases, including the following: (1) demographics of outpatient and telemedicine visits, (2) patient satisfaction with in-person visits, (3) patient satisfaction with telemedicine visits, and (4) provider satisfaction with telemedicine visits. Demographic data were extracted from patients' electronic medical records for the complete months of March and April of 2019 and 2020, generating three different

groups: 2019 in-person care (retrospective control group), 2020 in-person care (concurrent control group), and 2020 virtual care (telemedicine group). Satisfaction data from patients and providers were stored in independent databases.

### **Patient Demographics**

The number of visits and physician's specialty, as well as patient's sex, date of birth, insurance type, and address for all visits were extracted from electronic medical records. In this study, physician specialty was classified as one of the following seven groups: generalist specialties (family medicine, general internal medicine, geriatrics, pediatrics), pediatric subspecialties (including genetics and infant neurology), internal medicine subspecialties (including neurology), surgical subspecialties (including anesthesia, orthopedics, urology), obstetrics and gynecology, psychiatry, and ophthalmology, dermatology, and otolaryngology.

### **Patient Satisfaction With In-Person Visits**

Patient satisfaction is usually evaluated using the Net Promoter Score (NPS) survey (Cronbach  $\alpha=.96$ ) (FBA Consulting, unpublished material, 2020) [11]. This measure assesses satisfaction across several dimensions including access, payment process, infrastructure, and provider's services, as well as general satisfaction, using a 5-point Likert scale (1=very bad, 5=very good). Data were collected through an online survey sent via SMS text messaging after the visit.

### **Patient Satisfaction With Telemedicine Visits**

An 8-item questionnaire was developed to assess patient satisfaction with telemedicine services using a 1-7 scoring system. Like the survey completed by patients receiving in-person services, survey domains included questions assessing satisfaction with access, payment process, web portal (infrastructure), and provider's services, as well as general satisfaction (Cronbach  $\alpha=.86$ ). This evaluation was sent via email after the encounter.

### **Physician Satisfaction With Telemedicine**

Using a 19-item anonymous survey, clinicians were asked about their demographics, telemedicine experience, and general satisfaction. Open-ended questions assessed challenges in care delivery, diagnostic process, treatment, and the patient-provider relationship, as well as strategies used to overcome these barriers.

### **Data Analyses**

Following the procedures of the selected mixed methods design, each type of data was analyzed independently.

### **Quantitative Data**

Patient and provider demographic information was summarized using descriptive statistics. Trends for in-person and

telemedicine visits were plotted and compared to the null hypothesis of no trend using the chi-square test for trend. Patient characteristics of those receiving telemedicine services were compared to the demographics of patients from the retrospective and concurrent control groups using the chi-square test. As satisfaction data did not have a normal distribution, the Wilcoxon-Mann-Whitney test was used, after converting in-person and telemedicine scoring systems to a 0-100 scale. Responses regarding considering telemedicine as challenging and counts of qualitative responses were modeled after conducting bivariate analysis with logistic regressions that included the physician's specialty category as an independent variable, and gender, age, years of clinical experience, and telemedicine experience as covariables. Adjusted odds ratios, predicted probabilities, and 95% confidence intervals for the selected outcomes were estimated for each specialty category. All analyses were conducted in STATA (Version 14; StataCorp) and resulting *P* values  $<.05$  were considered statistically significant.

### **Qualitative Data**

Responses to open-ended questions were coded by two independent researchers following the procedures of content analysis [12]. Using a combination of inductive and deductive approaches, physicians' perceptions were grouped into emerging categories and subcategories organized according to the domains of the guiding questions. To ensure coding reliability, a random sample of 20% of the clinician surveys ( $n=53$ ) were double-coded, resulting in 96.4% coding agreement. Responses in each category were counted and grouped by domain. Quotes were selected to represent the participant's opinions and these were translated to English.

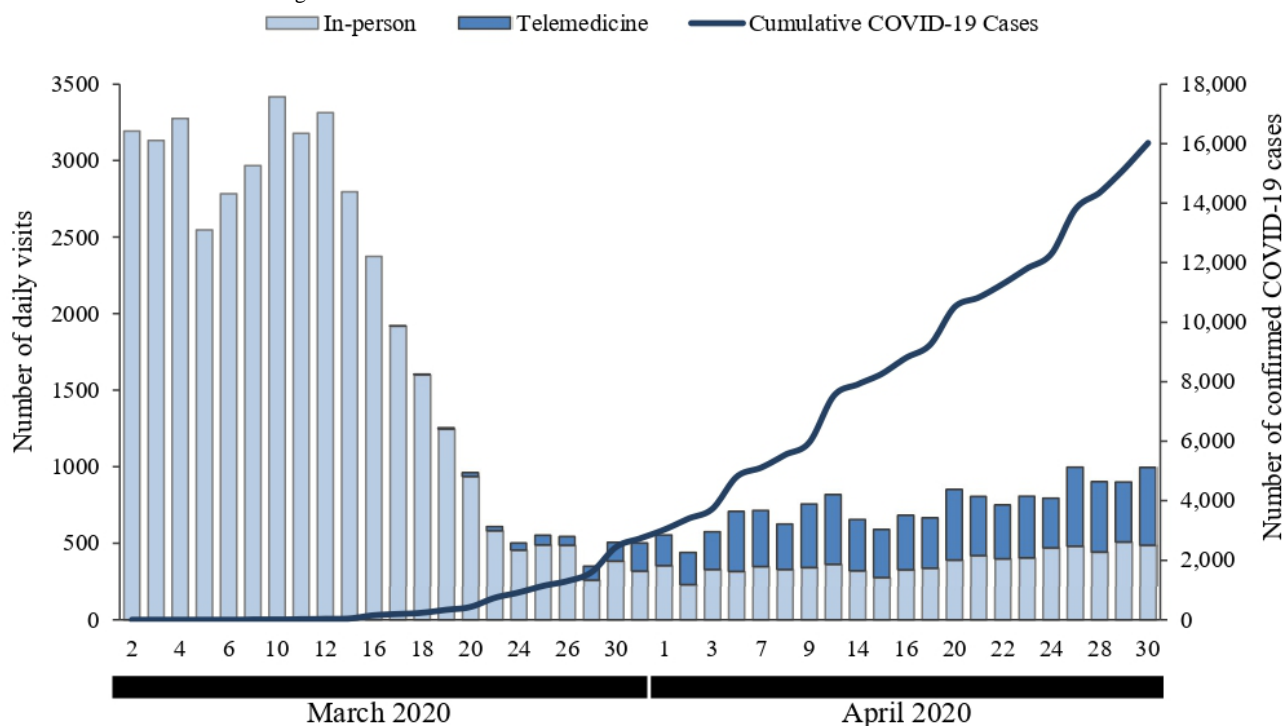
Mixed methods integration was conducted at the finding interpretation and reporting phases. To interpret both sources of data jointly, researchers gathered to discuss quantitative and qualitative results together. Integration at the reporting level occurred through a continuous narrative approach, in which the mixed data are presented in a single report, but in different sections [13].

## **Results**

### **Change in Clinical Practice**

In 2019 and 2020 before COVID-19 arrived in Chile, there were an average of 3039 and 3163 daily outpatient visits in the UC Christus Health Network, respectively. In-person visits were reduced to an average of 384.7 daily visits 3 weeks after the first patient with COVID-19 was confirmed in Chile, representing an 87.9% reduction in outpatient visits (Figure 1,  $P<.001$ ). Since telemedicine services began, the number of daily visits has increased, with the UC Christus Health Network delivering up to 509 virtual visits each day ( $P<.001$ ).

**Figure 1.** Rates of in-person and telemedicine visits at UC Christus Health and cumulative COVID-19 cases from March 1 to April 30, 2020, as well as relevant dates of COVID-19 management in Chile and UC Christus Health.



#### Relevant dates:

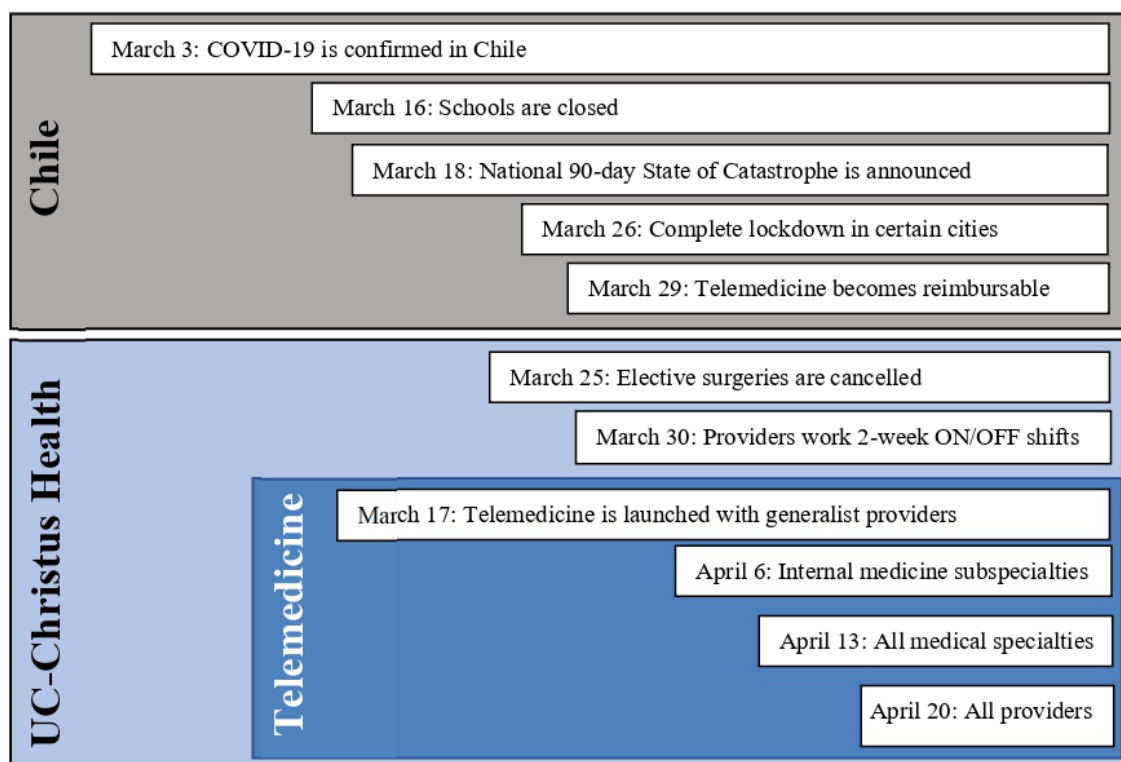
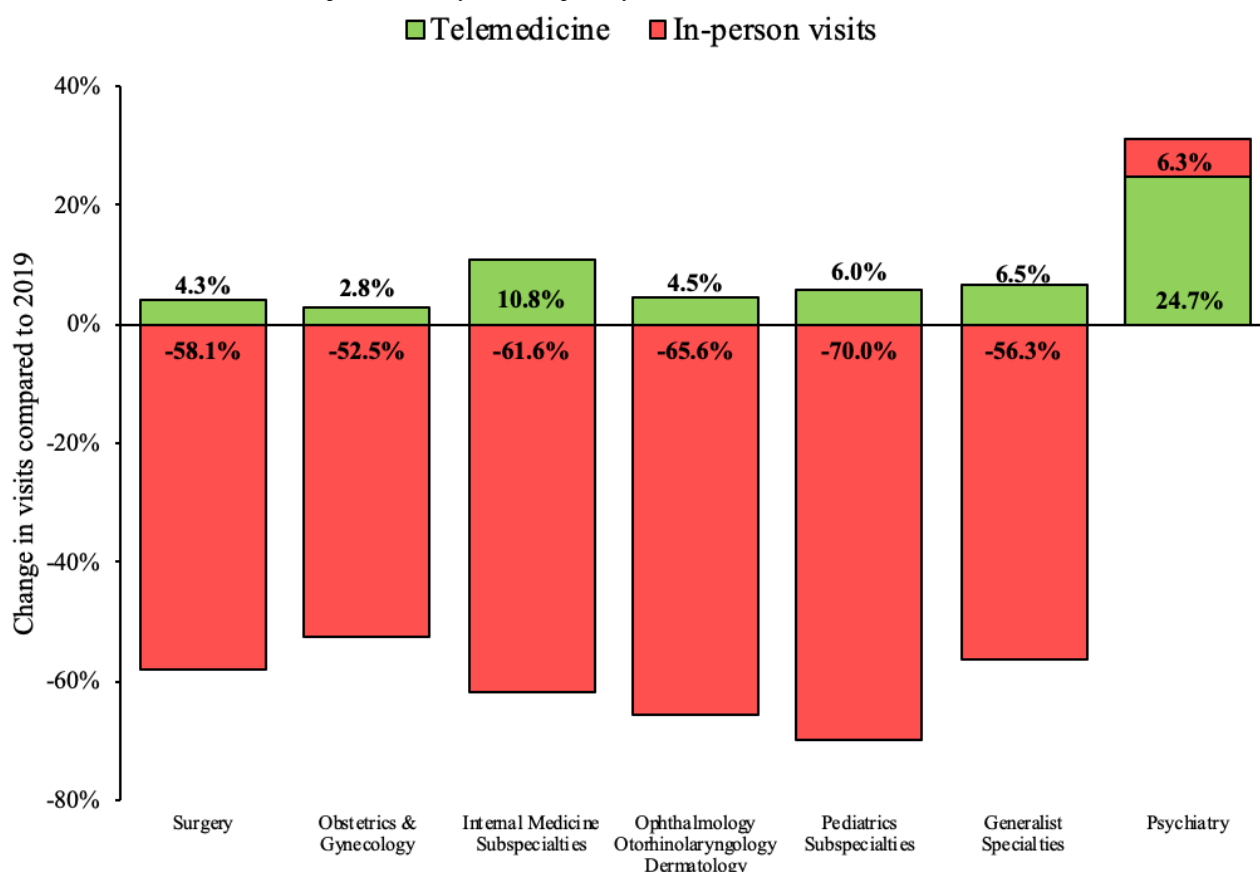


Figure 2 displays telemedicine and in-person activity across medical specialties during 2020 compared to 2019. All specialties except psychiatry decreased the number of in-person visits. Pediatrics subspecialties, along with ophthalmology,

otorhinolaryngology, and dermatology had the greatest reduction of in-person visits, experiencing between 61.6% to 70.0% of the demand of the previous year ( $P<.001$ ).



**Figure 2.** Rates of telemedicine and in-person visits by medical specialty.

Physicians from all specialty groups were able to implement virtual visits. Even though psychiatrists have been able to provide up to 24.7% of their 2019 volume of clinical encounters through virtual care, most medical groups have only delivered between 2.8% and 10.8% of visits compared to their 2019 baseline ( $P<.001$ ).

### Patient Demographics

Table 1 compares the demographic characteristics of the telemedicine, concurrent control, and retrospective control groups. Patients accessing telemedicine were more likely to be female and have private insurance, and less likely to be children or older adults, or residents of the Santiago Metropolitan Region compared to both in-person control groups ( $P<.01$  for all comparisons).

**Table 1.** Demographic characteristics of patients receiving care through telemedicine compared to patients from concurrent and retrospective control groups.

Characteristics	Telemedicine group (n=8592), n (%)	Concurrent control group (n=51,290), n (%)	P value	Retrospective control group (n=127,669), n (%)	P value
Gender (female), n (%)	5376 (62.6)	31,436 (61.4)	.024	77,845 (61.0)	.003
<b>Age category (years), n (%)</b>	N/A <sup>a</sup>	N/A	<.001	N/A	<.001
0-18	1749 (20.4)	11,855 (23.1)	N/A	34,357 (26.9)	N/A
19-64	5725 (66.6)	31,607 (61.6)	N/A	73,632 (57.7)	N/A
≥65	1118 (13.0)	7828 (15.3)	N/A	19,680 (15.4)	N/A
Private health insurance, n (%)	7542 (87.8)	21,319 (41.6)	<.001	56,495 (44.3)	<.001
Resident of the Santiago Metropolitan Region, n (%)	6529 (76.0)	42,833 (83.5)	<.001	111,822 (87.6)	<.001

<sup>a</sup>N/A: not applicable.

### Patient Satisfaction

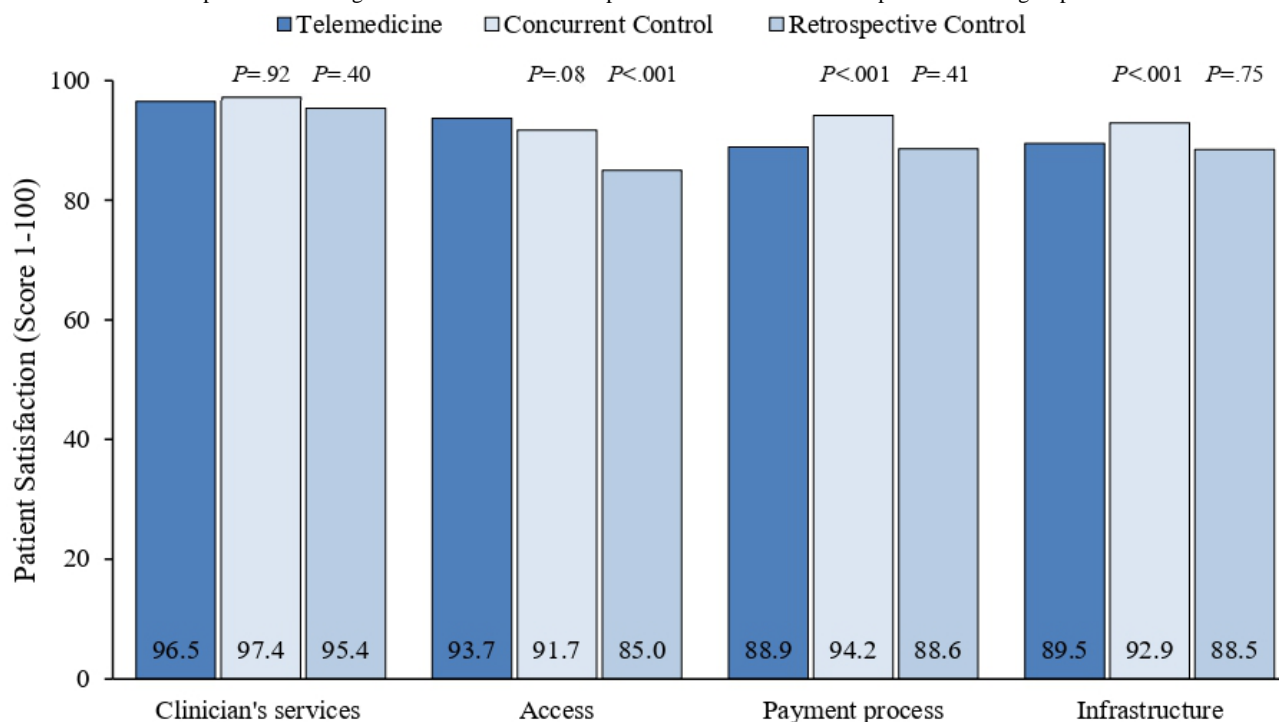
During March and April of 2019 and 2020, 1848 and 1187 patients responded to the satisfaction questionnaire after the

in-person visits, respectively. In addition, 3962 patients responded to the telemedicine satisfaction assessment. Satisfaction was very high with both telemedicine and in-person services (Figure 3). Patients receiving telemedicine services

reported similar satisfaction with clinician's services compared to both control groups. Patients using telemedicine reported similar access to providers compared to in-person visits in 2020 ( $P=.08$ ), but greater access compared to in-person visits in 2019 (8.7% increase,  $P<.001$ ). Patients using telemedicine care reported less satisfaction with the payment process (5.3%

reduction,  $P<.001$ ) and infrastructure (web portal, 3.4% reduction,  $P<.001$ ) compared to patients receiving in-person visits concurrently. Ratings in these domains were not statistically different compared to patients receiving in-person visits in 2019.

**Figure 3.** Satisfaction of patients receiving care via telemedicine compared to concurrent and retrospective control groups.

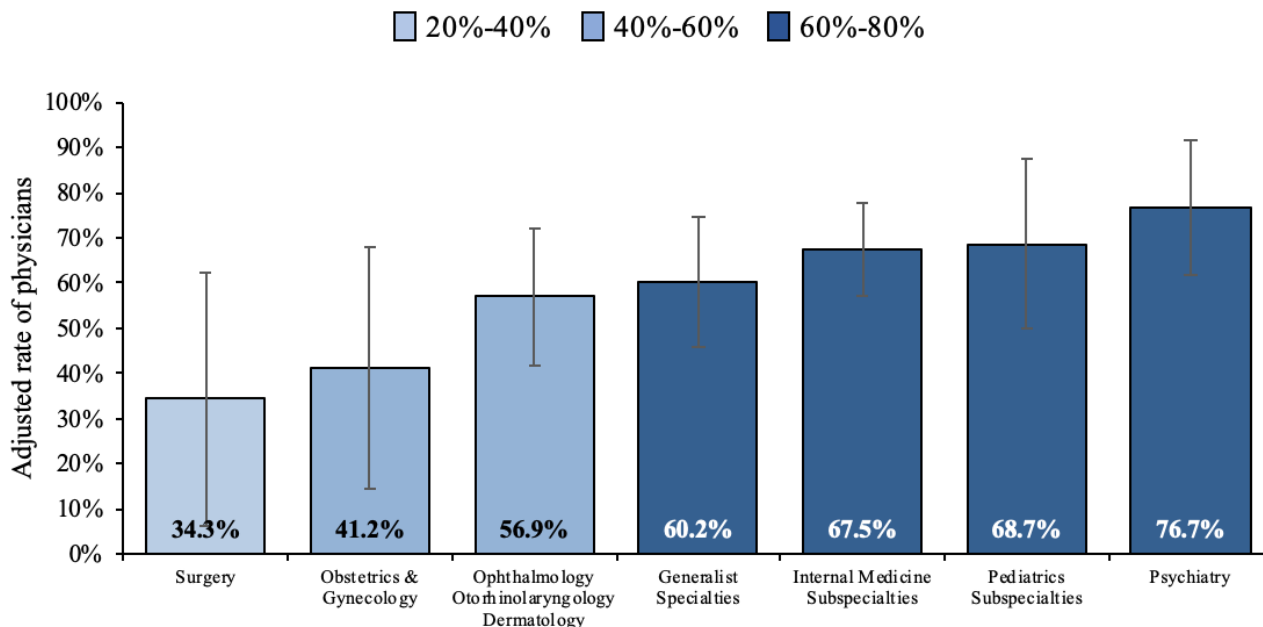


### Physician Satisfaction

A total of 263 clinicians responded to the satisfaction assessment (36.5%). They were mostly female ( $n=155$ , 58.9%), with an average age of 44 years (SD 10.9), and an average of 16.8 years (SD 11.4) of clinical experience. Most providers reported limited experience with telemedicine, with 147 clinicians (61%) reporting 10 or fewer virtual patients encounters at the time they responded to the questionnaire. There were no differences in gender ( $P=.37$ ), age ( $P=.10$ ), or specialty category ( $P=.32$ ) among survey respondents and nonrespondents ( $n=721$  providers). Overall, 244 providers were satisfied or very satisfied with telemedicine (92.8%), and most would recommend this service to friends or family members (94.2%).

### Experiencing Challenges When Conducting Telemedicine Visits

When providing telemedicine care, most physicians felt their clinical skills challenged somewhat or a lot (61.8%). Female providers felt more challenged than male providers (70.7% versus 50.9%,  $P=.002$ ). Surgeons, obstetricians, and gynecologists felt that their clinical skills were challenged the least, compared to providers from other medical specialties ( $P=.02$ ). These associations persisted after adjusting for covariates in a logistic regression model (Figure 4). There were no statistically significant differences in feeling that clinical skills were challenged according to the provider's age, years of clinical experience, or the number of patients seen virtually.

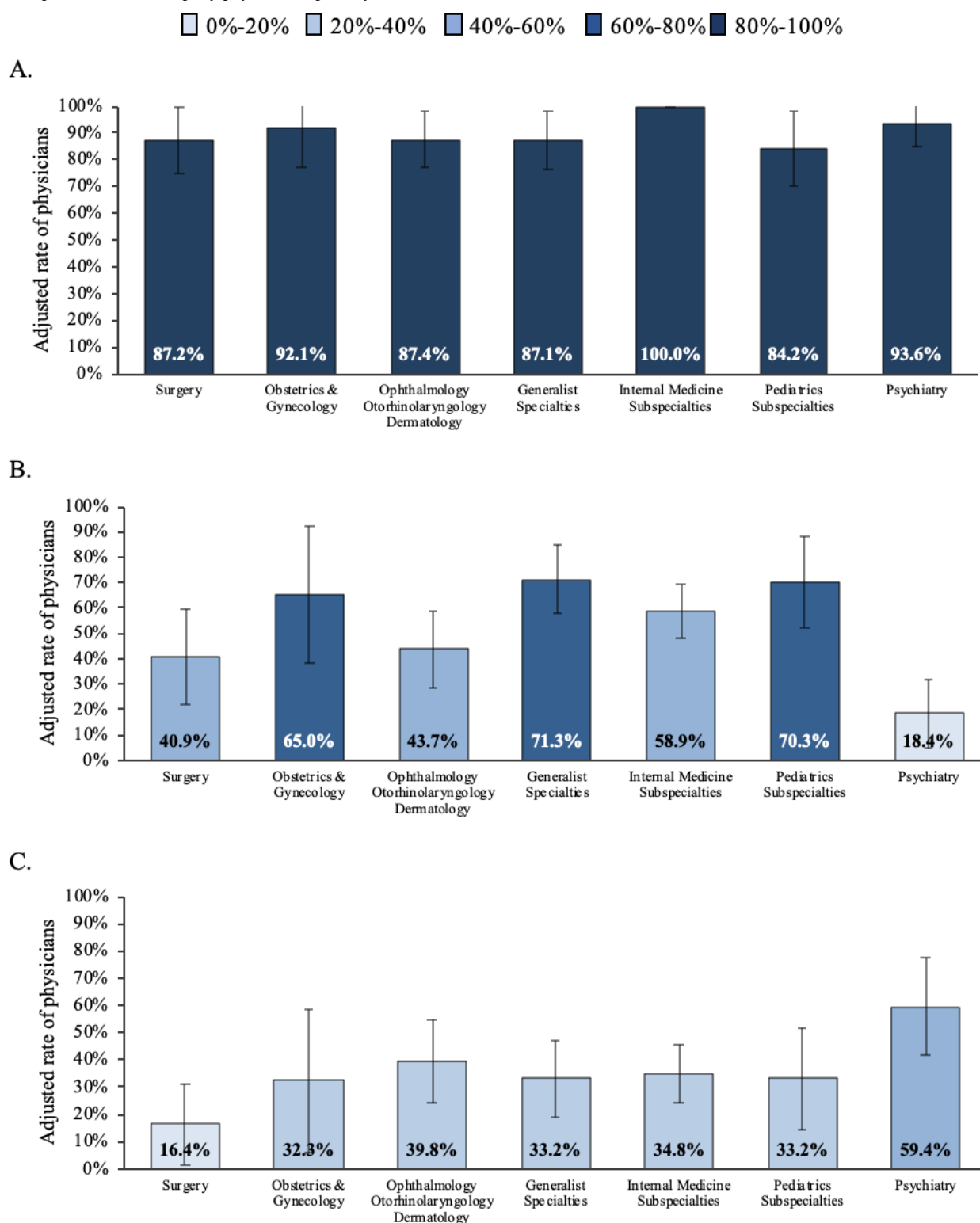
**Figure 4.** Adjusted rate and 95% CIs of physicians perceiving their clinical skills being challenged when using telemedicine, by physician's specialty.

### Specific Challenges With Telemedicine and Mechanisms of Addressing Them

Table 1 in [Multimedia Appendix 1](#) summarizes the categories and subcategories of challenges experienced with telemedicine delivery, the diagnostic process, and the patient-provider relationship. Most challenges were related to the delivery mode, specifically accessing the platforms used, dealing with patient scheduling, and accessing the resources required for telemedicine. There were no challenges specifically related to patient treatment. Several strategies were implemented by physicians to address these barriers (Table 2 in [Multimedia Appendix 1](#)). Strategies most commonly used were directly contacting the patient, and requesting support from the help center, peers, or patients.

[Figure 5](#) represents the rates of physicians experiencing any challenge with the delivery modality, diagnostic process, and patient-provider relationship according to their specialty category. Providers from all specialties experienced challenges with the delivery system, especially providers from internal medicine subspecialties and psychiatry ([Figure 5A](#),  $P=.046$ ). There were differences by specialty when reporting challenges related to a patient's diagnostic process. Most physicians who experienced challenges were from primary care and pediatrics subspecialties ([Figure 5B](#),  $P<.001$ ). Challenges in the patient-provider relationship also differed by specialty group. About 60% of psychiatrists reported experiencing challenges in this domain compared to 16.4% of surgeons ([Figure 5C](#),  $P=.02$ ).

**Figure 5.** Adjusted rate and 95% CIs of providers reporting challenges with (A) the modality of service, (B) the patient's diagnostic process, and (C) the patient-provider relationship, by physician's specialty.



## Discussion

### Main Results

The COVID-19 pandemic has radically transformed clinical care and our organization was forced to quickly deploy telemedicine services without the standard change management and training strategies that are usually required for successful digital transformations in health care [14]. Despite this, we have

seen a quick uptake and the practice has been widely accepted in a short period of time [15]. According to Kotter's change management process, the first step toward successful change is the sense of urgency to change [16]; it is undeniable that the COVID-19 pandemic has generated an enormous sense of urgency to adapt to this reality. In a few weeks, we were able to include all clinical specialties in the telemedicine project. Moreover, the Chilean National Health Fund (FONASA), which

had delayed including clinical specialties in funding schemes for years, established insurance coverage for telemedical services three weeks after the first COVID-19 case in Chile [17]. The COVID-19 pandemic created the perfect environment for providers and health care organizations to bring down long-lasting barriers to the implementation of telemedicine and other digital innovations [18].

### Comparison With Prior Work

The fear of becoming infected with COVID-19 in health care settings drove many patients away from in-person care [19]. Specialties that focus on children or require medical equipment (eg, otorhinolaryngology or ophthalmology) experienced the highest reduction of in-person visits. This is consistent with recent studies that identified a lack of access to laryngoscopes as a downside of telemedicine visits [20]. Even though telemedicine has provided physicians with an option to deliver outpatient services, the recovery of outpatient clinical activity levels toward 2019 levels differed between specialty groups. Psychiatrists and internal medicine subspecialists have been able to implement the highest number of virtual visits. This different level of adoption could be attributed to patient or clinician characteristics and could be further explored in future studies. In addition, differences were observed in the demographics of patients accessing telemedicine services compared to in-person visits. To continue the growth of this modality of care among all medical groups, and reduce disparities in health care access, further innovations should be developed to provide access to the patient and provider groups that have unequally benefited from this clinical transformation. For example, physicians could use patient-controlled devices to supply equipment normally available in offices [21], and marketing can be focused on groups with less access to care. It is noteworthy that telemedicine increased access to specialty care that previously could have been limited by geographic barriers [22].

Overall, both clinicians and patients were highly satisfied with the implemented telemedicine services. This is partially comparable to the findings from a previous systematic review that synthesized 17 studies assessing patient and physician perceptions of telemedicine versus in-person consultations in different medical specialties [23]. The systematic review found that most studies reported no significant differences in patient-related outcomes such as convenience, format of the consultation, or rapport. In addition, the same systematic review found that overall, physicians showed reduced satisfaction with telemedicine visits regarding communication (patient-provider relationship), and physical examination or diagnostic assessment. These findings are compatible with the results of our study, as some of the concerns identified by providers related to the impact of telemedicine on diagnostic assessment and the patient-provider relationship, as well as complications associated with the implemented delivery modality. Even though the system was far from perfect (eg, using three different platforms during a virtual encounter), patients and providers valued telemedicine as an innovation that mitigated the risk of infection associated with in-person visits. Reducing the risk of contagion has been identified as a likely facilitator to adopting telemedicine [24].

In parallel with a high overall satisfaction with telemedicine, we observed that more than 60% of all clinicians felt that their clinical skills were challenged during virtual visits. This is similar to a previous study that reported an increase in overall workload, mental effort, and psychological stress during virtual visits [25]. We also observed differences across specialties in the perception of telemedicine consultations and how telemedicine can challenge their clinical skills; those in technical specialties found telemedicine less challenging than those in relationship-focused specialties [26,27]. This difference in the physicians' orientation toward patient care might also reflect differences in adjusting to a new service delivery process.

When analyzing the types of challenges faced by clinicians and the strategies used to address them, it is noticeable that most of the challenges faced were due to the delivery mode, the platform used, and patient scheduling. This finding was identified by embedding a qualitative assessment to the quantitative survey and is consistent with a Cochrane systematic review that found that technical difficulties experienced by clinicians during telemedicine services produced high dropout rates [28]. As telemedicine was quickly deployed as a system-wide innovation in response to COVID-19, there was no time to produce a user-friendly platform, but this will be fundamental to maintaining clinician engagement with virtual services as the pandemic evolves as well as once the pandemic is over. In addition, for telemedicine to overcome all identified challenges, further training is needed among providers. Although medical schools have embraced online training [29], medical students should also be trained in telemedicine [30]. The presented system-wide implementation required that all physicians, regardless of their specialty, were knowledgeable about virtual care. As providers become more experienced conducting telemedicine encounters, it is likely that these barriers will be reduced over time and their experience should be considered as valuable input for future curriculum design.

In the patient-clinician relationship domain there were striking differences between specialty groups. This is especially relevant since psychiatrists experienced the greatest relative increase in telemedicine consultations. We did not explicitly seek to study the causes of this difference, but it is probably a consequence of the differing types of clinician-patient relationships between specialties [26,27]. This is relevant since mental health is one of the areas that has seen the greatest increase in uptake in the past years [31] and it has been a major concern during the COVID-19 pandemic [32]. Mental health telemedicine services are, on average, equally effective when compared to face-to-face consultations [4,28].

### Limitations

Although the implementation of virtual care in response to COVID-19 has been reported in different settings [24,33-38], this is the first report of a system-wide accelerated implementation of telemedicine services using quantitative and qualitative methods. This large-scale implementation allowed us to identify and compare visits and challenges experienced by different specialty groups. However, as any other evaluation, this paper has limitations that are important to acknowledge. First, although there were no differences in the demographics



and medical characteristics of providers who did and did not respond to the satisfaction survey, selection bias is always present in voluntary surveys [39]. This limitation also affects patient satisfaction ratings. In this group, because the available data consisted of anonymous ratings, we could not assess potential responder bias. Second, because of the rapid deployment of telemedicine services, we developed a patient satisfaction questionnaire to assess patient satisfaction with telemedicine similar to the NPS but could have used validated measures [40] and integrated qualitative components, as we did in the developed questionnaire for providers. Finally, COVID-19 not only affected health care provider priorities but also changed patient priorities. For example, during the pandemic, patients may be less likely to seek care for chronic conditions, and this

could affect patient satisfaction. Future system-wide evaluations of telemedicine should be conducted once the current pandemic is controlled. Despite these limitations, this report highlights the challenges related to implementing telemedicine experienced by multiple groups of physicians and the mechanisms they used to address these challenges during the COVID-19 pandemic.

## Conclusions

Telemedicine produced high satisfaction among patients and providers. Although this modality of clinical care was rapidly deployed in response to the COVID-19 pandemic, there was high heterogeneity in its implementation across medical specialties. These differences need to be considered in future implementations of telemedicine when the current medical context is addressed.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary tables.

[DOCX File, 20 KB - [jmir\\_v22i10e22146\\_app1.docx](#)]

## References

1. Adalja AA, Toner E, Inglesby TV. Priorities for the US Health Community Responding to COVID-19. *JAMA* 2020 Apr 14;323(14):1343-1344. [doi: [10.1001/jama.2020.3413](#)] [Medline: [32125355](#)]
2. Mehrotra A, Ray K, Brockmeyer D, Barnett M, Bender J. Rapidly Converting to "Virtual Practices": Outpatient Care in the Era of Covid-19. *NEJM Catal Innov Care Deliv* 2020 Apr 01:1. [doi: [10.1056/CAT.20.0091](#)]
3. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. *N Engl J Med* 2020 Apr 30;382(18):1679-1681. [doi: [10.1056/nejmp2003539](#)]
4. Klaassen B, van Beijnum B, Hermens H. Usability in telemedicine systems-A literature survey. *Int J Med Inform* 2016 Sep;93:57-69. [doi: [10.1016/j.ijmedinf.2016.06.004](#)] [Medline: [27435948](#)]
5. Ekeland A, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. *Int J Med Inform* 2010 Nov;79(11):736-771. [doi: [10.1016/j.ijmedinf.2010.08.006](#)] [Medline: [20884286](#)]
6. Andrees V, Klein T, Augustin M, Otten M. Live interactive teledermatology compared to in-person care - a systematic review. *J Eur Acad Dermatol Venereol* 2020 Apr;34(4):733-745. [doi: [10.1111/jdv.16070](#)] [Medline: [31715035](#)]
7. Kozikowski A, Shotwell J, Wool E, Slaboda JC, Abrashkin KA, Rhodes K, et al. Care Team Perspectives and Acceptance of Telehealth in Scaling a Home-Based Primary Care Program: Qualitative Study. *JMIR Aging* 2019 Jun 02;2(1):e12415 [FREE Full text] [doi: [10.2196/12415](#)] [Medline: [31518266](#)]
8. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. *J Am Med Inform Assoc* 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: [10.1093/jamia/ocaa072](#)] [Medline: [32324855](#)]
9. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. *J Am Med Inform Assoc* 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: [10.1093/jamia/ocaa067](#)] [Medline: [32311034](#)]
10. Cresswell JW, Plano-Clark VL. Designing and conducting mixed methods research. Thousand Oaks, CA, USA: Sage; 2017.
11. Reichheld FF. The one number you need to grow. *Harv Bus Rev* 2003 Dec;81(12):46-54. [Medline: [14712543](#)]
12. Weber RP. Basic Content Analysis. Sage: London, UK; 1990.
13. Fetter MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. *Health Serv Res* 2013 Dec 23;48(6 Pt 2):2134-2156 [FREE Full text] [doi: [10.1111/1475-6773.12117](#)] [Medline: [24279835](#)]
14. Gagnon M, Ghandour EK, Talla PK, Simonyan D, Godin G, Labrecque M, et al. Electronic health record acceptance by physicians: testing an integrated theoretical model. *J Biomed Inform* 2014 Apr;48:17-27 [FREE Full text] [doi: [10.1016/j.jbi.2013.10.010](#)] [Medline: [24184678](#)]
15. Singh A, Hess T. How Chief Digital Officers promote the digital transformation of their companies. *MIS Quarterly Executive* 2017;16(1):1 [FREE Full text] [doi: [10.4324/9780429286797-9](#)]

16. Kotter J. Leading change: Why transformation efforts fail. *Harvard Business Review*. 1995. URL: <https://hbr.org/1995/05/leading-change-why-transformation-efforts-fail-2> [accessed 2020-09-18]
17. Fondo Nacional de Salud. FONASA facilitates remote visits to about 40 providers: general medicine, medical specialties, dietitian, psychology, speech therapy. URL: <https://www.fonasa.cl/sites/fonasa/noticia/Noticia-atencion-remota> [accessed 2020-07-20]
18. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. *N Engl J Med* 2020 Jun 04;382(23):e82. [doi: [10.1056/nejmp2005835](https://doi.org/10.1056/nejmp2005835)]
19. Wong LE, Hawkins JE, Murrell KL. Where are all the patients? Addressing Covid-19 fear to encourage sick patients to seek emergency care. *N Engl J Med Catalyst* 2020;1. [doi: [10.1056/CAT.20.0193](https://doi.org/10.1056/CAT.20.0193)]
20. Strohl MP, Dwyer CD, Ma Y, Rosen CA, Schneider SL, Young VN. Implementation of Telemedicine in a Laryngology Practice During the COVID-19 Pandemic: Lessons Learned, Experiences Shared. *J Voice* 2020 Jun 23;1 [FREE Full text] [doi: [10.1016/j.jvoice.2020.06.017](https://doi.org/10.1016/j.jvoice.2020.06.017)] [Medline: [32778359](https://pubmed.ncbi.nlm.nih.gov/32778359/)]
21. Bitterman N. Design of medical devices--a home perspective. *Eur J Intern Med* 2011 Feb;22(1):39-42. [doi: [10.1016/j.ejim.2010.09.017](https://doi.org/10.1016/j.ejim.2010.09.017)] [Medline: [21238891](https://pubmed.ncbi.nlm.nih.gov/21238891/)]
22. Khairat S, Haithcoat T, Liu S, Zaman T, Edson B, Gianforcaro R, et al. Advancing health equity and access using telemedicine: a geospatial assessment. *J Am Med Inform Assoc* 2019 Aug 01;26(8-9):796-805 [FREE Full text] [doi: [10.1093/jamia/ocz108](https://doi.org/10.1093/jamia/ocz108)] [Medline: [31340022](https://pubmed.ncbi.nlm.nih.gov/31340022/)]
23. Zandbelt L, de Kanter FEC, Ubbink D. E-consulting in a medical specialist setting: Medicine of the future? *Patient Educ Couns* 2016 May;99(5):689-705. [doi: [10.1016/j.pec.2015.11.005](https://doi.org/10.1016/j.pec.2015.11.005)] [Medline: [26597384](https://pubmed.ncbi.nlm.nih.gov/26597384/)]
24. Bloem BR, Dorsey ER, Okun MS. The Coronavirus Disease 2019 Crisis as Catalyst for Telemedicine for Chronic Neurological Disorders. *JAMA Neurol* 2020 Aug 01;77(8):927-928. [doi: [10.1001/jamaneurol.2020.1452](https://doi.org/10.1001/jamaneurol.2020.1452)] [Medline: [32329796](https://pubmed.ncbi.nlm.nih.gov/32329796/)]
25. Krousel-Wood MA, Re RN, Abdoh A, Bradford D, Kleit A, Chambers R, et al. Patient and physician satisfaction in a clinical study of telemedicine in a hypertensive patient population. *J Telemed Telecare* 2001 Jun 22;7(4):206-211. [doi: [10.1258/1357633011936417](https://doi.org/10.1258/1357633011936417)] [Medline: [11506755](https://pubmed.ncbi.nlm.nih.gov/11506755/)]
26. Chan C, Ahmad W. Differences in physician attitudes towards patient-centredness: across four medical specialties. *Int J Clin Pract* 2012 Jan;66(1):16-20. [doi: [10.1111/j.1742-1241.2011.02831.x](https://doi.org/10.1111/j.1742-1241.2011.02831.x)] [Medline: [22171901](https://pubmed.ncbi.nlm.nih.gov/22171901/)]
27. Borges N, Osmon W. Personality and Medical Specialty Choice: Technique Orientation versus People Orientation. *Journal of Vocational Behavior* 2001 Feb;58(1):22-35. [doi: [10.1006/jvbe.2000.1761](https://doi.org/10.1006/jvbe.2000.1761)]
28. Flodgren G, Rachas A, Farmer A, Inzitari M, Shepperd S. Interactive telemedicine effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews* 2015;2015(9):e. [doi: [10.1002/14651858.cd002098.pub2](https://doi.org/10.1002/14651858.cd002098.pub2)]
29. Rose S. Medical Student Education in the Time of COVID-19. *JAMA* 2020 Jun 02;323(21):2131-2132. [doi: [10.1001/jama.2020.5227](https://doi.org/10.1001/jama.2020.5227)] [Medline: [32232420](https://pubmed.ncbi.nlm.nih.gov/32232420/)]
30. Waseh S, Dicker A. Telemedicine Training in Undergraduate Medical Education: Mixed-Methods Review. *JMIR Med Educ* 2019 Apr 08;5(1):e12515 [FREE Full text] [doi: [10.2196/12515](https://doi.org/10.2196/12515)] [Medline: [30958269](https://pubmed.ncbi.nlm.nih.gov/30958269/)]
31. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry* 2017 Feb 26;16(1):90-99 [FREE Full text] [doi: [10.1002/wps.20388](https://doi.org/10.1002/wps.20388)] [Medline: [28127925](https://pubmed.ncbi.nlm.nih.gov/28127925/)]
32. Holmes E, O'Connor R, Perry V, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *Lancet Psychiatry* 2020 Jun;7(6):547-560 [FREE Full text] [doi: [10.1016/S2215-0366\(20\)30168-1](https://doi.org/10.1016/S2215-0366(20)30168-1)] [Medline: [32304649](https://pubmed.ncbi.nlm.nih.gov/32304649/)]
33. Boehm K, Ziewers S, Brandt MP, Sparwasser P, Haack M, Willems F, et al. Telemedicine Online Visits in Urology During the COVID-19 Pandemic-Potential, Risk Factors, and Patients' Perspective. *Eur Urol* 2020 Jul;78(1):16-20 [FREE Full text] [doi: [10.1016/j.eururo.2020.04.055](https://doi.org/10.1016/j.eururo.2020.04.055)] [Medline: [32362498](https://pubmed.ncbi.nlm.nih.gov/32362498/)]
34. Elkaddoum R, Haddad F, Eid R, Kourie H. Telemedicine for cancer patients during COVID-19 pandemic: between threats and opportunities. *Future Oncol* 2020 Jun;16(18):1225-1227 [FREE Full text] [doi: [10.2217/fon-2020-0324](https://doi.org/10.2217/fon-2020-0324)] [Medline: [32356460](https://pubmed.ncbi.nlm.nih.gov/32356460/)]
35. Trinidad J, Kroshinsky D, Kaffenberger BH, Rojek NW. Telemedicine for inpatient dermatology consultations in response to the COVID-19 pandemic. *J Am Acad Dermatol* 2020 Jul;83(1):e69-e71 [FREE Full text] [doi: [10.1016/j.jaad.2020.04.096](https://doi.org/10.1016/j.jaad.2020.04.096)] [Medline: [32339708](https://pubmed.ncbi.nlm.nih.gov/32339708/)]
36. Wong TY, Bandello F. Academic Ophthalmology during and after the COVID-19 Pandemic. *Ophthalmology* 2020 Aug;127(8):e51-e52 [FREE Full text] [doi: [10.1016/j.ophtha.2020.04.029](https://doi.org/10.1016/j.ophtha.2020.04.029)] [Medline: [32359842](https://pubmed.ncbi.nlm.nih.gov/32359842/)]
37. Del Hoyo J, Aguas M. Implementing Telemedicine in Inflammatory Bowel Disease: is COVID-19 the definitive trigger? *Gastroenterol Hepatol* 2020 May 15:S0210 [FREE Full text] [doi: [10.1016/j.gastrohep.2020.05.002](https://doi.org/10.1016/j.gastrohep.2020.05.002)] [Medline: [32419714](https://pubmed.ncbi.nlm.nih.gov/32419714/)]
38. Turrentine M, Ramirez M, Monga M, Gandhi M, Swaim L, Tyer-Viola L, et al. Rapid Deployment of a Drive-Through Prenatal Care Model in Response to the Coronavirus Disease 2019 (COVID-19) Pandemic. *Obstet Gynecol* 2020 Jul;136(1):29-32 [FREE Full text] [doi: [10.1097/AOG.0000000000003923](https://doi.org/10.1097/AOG.0000000000003923)] [Medline: [32332322](https://pubmed.ncbi.nlm.nih.gov/32332322/)]

39. Cheung K, Ten Klooster PM, Smit C, de Vries H, Pieterse M. The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health* 2017 Mar 23;17(1):276 [FREE Full text] [doi: [10.1186/s12889-017-4189-8](https://doi.org/10.1186/s12889-017-4189-8)] [Medline: [28330465](https://pubmed.ncbi.nlm.nih.gov/28330465/)]
40. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, IDEATel Consortium. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. *J Am Med Inform Assoc* 2006;13(6):660-667 [FREE Full text] [doi: [10.1197/jamia.M2146](https://doi.org/10.1197/jamia.M2146)] [Medline: [16929036](https://pubmed.ncbi.nlm.nih.gov/16929036/)]

## Abbreviations

**NPS:** Net Promoter Score

*Edited by G Eysenbach; submitted 04.07.20; peer-reviewed by J Clemensen, J Li, I Mircheva; comments to author 13.07.20; revised version received 20.07.20; accepted 17.08.20; published 06.10.20.*

*Please cite as:*

Garcia-Huidobro D, Rivera S, Valderrama Chang S, Bravo P, Capurro D

System-Wide Accelerated Implementation of Telemedicine in Response to COVID-19: Mixed Methods Evaluation

*J Med Internet Res* 2020;22(10):e22146

URL: <https://www.jmir.org/2020/10/e22146>

doi: [10.2196/22146](https://doi.org/10.2196/22146)

PMID: [32903195](https://pubmed.ncbi.nlm.nih.gov/32903195/)

©Diego Garcia-Huidobro, Solange Rivera, Sebastián Valderrama Chang, Paula Bravo, Daniel Capurro. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Topics, Delivery Modes, and Social-Epistemological Dimensions of Web-Based Information for Patients Undergoing Renal Transplant and Living Donors During the COVID-19 Pandemic: Content Analysis

Charlotte W van Klaveren<sup>1</sup>, MSc; Peter G M de Jong<sup>2</sup>, PhD; Renée A Hendriks<sup>2</sup>, MSc; Franka Luk<sup>1</sup>, PhD; Aiko P J de Vries<sup>1</sup>, MD, PhD; Paul J M van der Boog<sup>1</sup>, MD, PhD; Marlies E J Reinders<sup>1,3</sup>, MD, Prof Dr

<sup>1</sup>Division of Nephrology and Transplant Center, Department of Internal Medicine, Leiden University Medical Center, Leiden, Netherlands

<sup>2</sup>Center for Innovation in Medical Education, Leiden University Medical Center, Leiden, Netherlands

<sup>3</sup>Division of Nephrology and Transplantation, Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, Netherlands

**Corresponding Author:**

Charlotte W van Klaveren, MSc  
Division of Nephrology and Transplant Center  
Department of Internal Medicine  
Leiden University Medical Center  
Albinusdreef 2  
Leiden, 2333 ZA  
Netherlands  
Phone: 31 (0)71 526512  
Email: [c.w.van\\_klaveren@lumc.nl](mailto:c.w.van_klaveren@lumc.nl)

## Abstract

**Background:** The COVID-19 pandemic has markedly affected renal transplant care. During this time of social distancing, limited in-person visits, and uncertainty, patients and donors are relying more than ever on telemedicine and web-based information. Several factors can influence patients' understanding of web-based information, such as delivery modes (instruction, interaction, and assessment) and social-epistemological dimensions (choices in interactive knowledge building).

**Objective:** The aim of this study was to systemically evaluate the content, delivery modes, and social-epistemological dimensions of web-based information on COVID-19 and renal transplantation at time of the pandemic.

**Methods:** Multiple keyword combinations were used to retrieve websites on COVID-19 and renal transplantation using the search engines Google.com and Google.nl. From 14 different websites, 30 webpages were examined to determine their organizational sources, topics, delivery modes, and social-epistemological dimensions.

**Results:** The variety of topics and delivery modes was limited. A total of 13 different delivery modes were encountered, of which 8 (62%) were instructional and 5 (38%) were interactional; no assessment delivery modes were observed. No website offered all available delivery modes. The majority of delivery modes (8/13, 62%) focused on individual and passive learning, whereas group learning and active construction of knowledge were rarely encountered.

**Conclusions:** By taking interactive knowledge transfer into account, the educational quality of eHealth for transplant care could increase, especially in times of crisis when rapid knowledge transfer is needed.

(*J Med Internet Res* 2020;22(10):e22068) doi:[10.2196/22068](https://doi.org/10.2196/22068)

**KEYWORDS**

web-based information; internet; websites; patient education; COVID-19; renal transplantation; eHealth; constructivism; social-epistemological dimensions; teaching modes; health communication

## Introduction

The COVID-19 pandemic is affecting transplant activity worldwide [1-3]. In addition to its impact on donation and transplantation numbers, the pandemic has markedly affected outpatient transplant care [1,2]. Additionally, patients experience uncertainty and fear; therefore, they rely more heavily on information provision [4]. For patients undergoing renal transplant, this uncertainty includes the fear of contracting COVID-19, the postponement of transplant procedures, screening strategies for COVID-19, mental health problems, and the risk of using immunosuppressive therapy during the pandemic [1,2]. With current measures limiting in-person visits between patients and health care providers, it has become vitally important to use alternative means of communication to inform patients undergoing renal transplant and living donors. Telephone and video communication have been generally adopted as means to continue outpatient care during the crisis [5,6]. Web-based information is used in the provision of eHealth to meet patients' information needs when they are facing uncertainty and fear [7].

Multiple eHealth interventions aimed at disease control and increasing knowledge were already available for renal patients before the outbreak [8]. The majority of patients with renal disease search the internet on a regular basis to obtain additional information on their disease and its treatment [9]. In general, patients' health-related internet use is associated with increasing health literacy [10]. Health literacy can be best described as an individual's well-considered health decisions and goal setting by accessing and understanding health-related information [11]. Patients' health literacy is positively associated with their self-management [12], motivation [13], risk perception [14], participation in health decision making [15,16], etc. However, despite the high number of patients with renal disease who use the internet as an additional information source, health literacy among renal patients and transplant recipients is still limited. A systematic review showed that in the United Kingdom, 25% of patients with renal disease not on dialysis, 27% of patients with renal disease on dialysis, and 14% of transplant recipients have limited health literacy [17].

Several factors could influence patients' health literacy, such as their understanding of web-based health information [18]. As knowledge transfer is an essential element in initiating behavioral change, the effective use of delivery modes could play a role in understanding information about COVID-19 and achieving this change [19]. Despite the increased provision of web-based information and eHealth interventions relating to COVID-19, it is not known how web-based information has been provided during the pandemic [20].

Information can be conveyed through instruction, interaction, and assessment [21]. Various delivery modes can be used for this, such as text messages, discussions, and quizzes. Some of these modes focus on the passive transfer of factual information to the receiver (called objectivistic modes), whereas others focus on knowledge construction, information processing, hands-on interaction with the content, and problem-solving (constructivist modes) [22,23]. Constructivist learning promotes more active

processing and personalization of information compared to receiving it passively. This results in deeper understanding and embedding of newly acquired knowledge [24-26]. In addition to this so-called epistemological dimension, delivery modes can also be categorized socially as individual or group learning [27]. From the literature, it can be assumed that people who learn both individually and collaboratively can construct knowledge better than learners who only learn individually [28].

Web-based information was rapidly developed by health services, academic centers, and patient associations to meet people's information needs related to COVID-19 and transplantation. However, if patients' understanding is to be optimized, the educational quality of this web-based information should be taken into account. The aim of this study is to provide a systematic overview of the source organizations, topics discussed, available delivery modes, and corresponding social-epistemological dimensions of websites for patients undergoing renal transplant and living donors on the topic of COVID-19. The practical implications derived from this study can be used to increase the quality of web-based information for transplant care, especially in times of crisis when rapid transfer of knowledge is needed.

## Methods

Websites on renal transplantation and COVID-19 were systemically identified using the most frequently consulted search engines for two countries [29]: Google.com [30] (for the United States, in North America), and Google.nl [31] (for the Netherlands, in Europe). The researcher's internet settings were deleted to obtain the cleanest and most objective search results possible. Additionally, the researcher's Google accounts, prerendering, and location sharing were turned off. The location setting was changed manually to the United States to minimize the influence of the researcher's location (the Netherlands) when using the search engine Google.com. Websites were searched from a laptop in the Netherlands in the month of March 2020.

To obtain a clearer picture of the available web-based source organizations, delivery modes, and social epistemological dimensions, general search engine queries related to renal transplantation and COVID-19 were used to obtain relevant websites. Because the English keywords *renal transplantation* and *COVID-19* have multiple synonyms, and patients might use any combination of those, a non-research-based selection of 14 keyword combinations was used to identify potentially relevant websites in English containing information on both renal transplantation and COVID-19 (Multimedia Appendix 1, columns A and C). In the Dutch search, a total of 7 keyword combinations were used (Multimedia Appendix 1, columns B and C).

Previous studies have demonstrated that searchers rarely read beyond the first 10 search results [32]. Therefore, in this study, websites were included if they were in the top 10 search results, including websites that used paid advertisements to top the search results. Front page news and suggested YouTube videos were not included. All possible English and Dutch keyword combinations were entered respectively in both Google search



engines. The first 10 recovered websites for each search were included for detailed examination.

In this study, we were interested in publicly accessible and available websites on the World Wide Web that patients and donors could access on an ad hoc basis. Therefore, for each search, websites were excluded if the website was no longer available, the website was under construction, or the website's content was behind a paywall. Websites were also excluded if the selected webpage was redundant (eg, press releases, news articles, blogs, academic journals, or webpages to sell products), or the webpage did not provide information on COVID-19 and renal transplantation. These websites were excluded because we were interested in websites which had as goal to objectively inform patients or living donors. Additional exclusion criteria were the website was a duplicate with a previous hit or the written language was other than English or Dutch, depending on the language of the keyword combination.

Potentially relevant webpages were obtained by using the same keyword combination in the website search function that was used for the website inclusion. The exclusion criteria for webpage selection per website were identical to those described for the website selection. Duplicated webpages obtained with different combinations of keywords were all noted. During data analysis, overlapping webpages were included only once. A total of 30 webpages, 15 in English and 15 in Dutch, were included for detailed examination.

Three randomly selected websites were individually examined by two authors. The results of both authors were discussed and calibrated until agreement on coding was reached. Then, the first author re-examined these three websites and examined the remaining websites. A calibration diary was maintained during examination, and another author was consulted when uncertainty arose.

Data analysis was performed in the month of April 2020. All websites were classified based on source organization, namely professional nonprofit organizations, such as hospitals; support groups, such as patient associations; governments, such as the Ministry of Health; individual practice, such as personal websites; or commercial organizations, such as independent dieticians. The website was labelled as "other" if the organizational source of a website did not match any of these categories. Additionally, English language websites were classified based on their generic top level domain (eg, .com, .org), which was not applicable to Dutch websites.

Each webpage was classified based on available topics and delivery modes. Immediately available content on COVID-19 and renal transplantation was categorized thematically. The available delivery modes on each webpage were classified into

instruction, interaction, or assessment based on the studies of Toven-Lindsey et al (2015) [21] and Hendriks et al (2019) [33]. Delivery modes that were not predetermined were categorized individually by two authors, followed by discussion and calibration until agreement was reached.

After data collection was completed, the Teaching Approach Framework described in 2006 by Arbaugh and Benbunan-Fich [27] was used to categorize the identified delivery modes into social-epistemological dimensions: objectivist-individual; objectivist-group; constructivist-individual; and constructivist-group. Previously implemented categorizations by Toven-Lindsey et al (2015) [21] and Hendriks et al (2019) [33] were taken into account. However, in contrast to these studies, links to external web-based resources were categorized as objectivist-individual instead of constructivist-individual because external links available on websites transmit knowledge and do not actively build knowledge as designed for massive open online courses (MOOCs), which was the context of the prior studies. Newly found delivery modes were categorized into a social-epistemological dimension individually by two authors and discussed and calibrated until concurrence was reached.

Descriptive statistics were used to analyze the variety of organizational sources, content topics, delivery modes, and social-epistemological dimensions within and between websites.

## Results

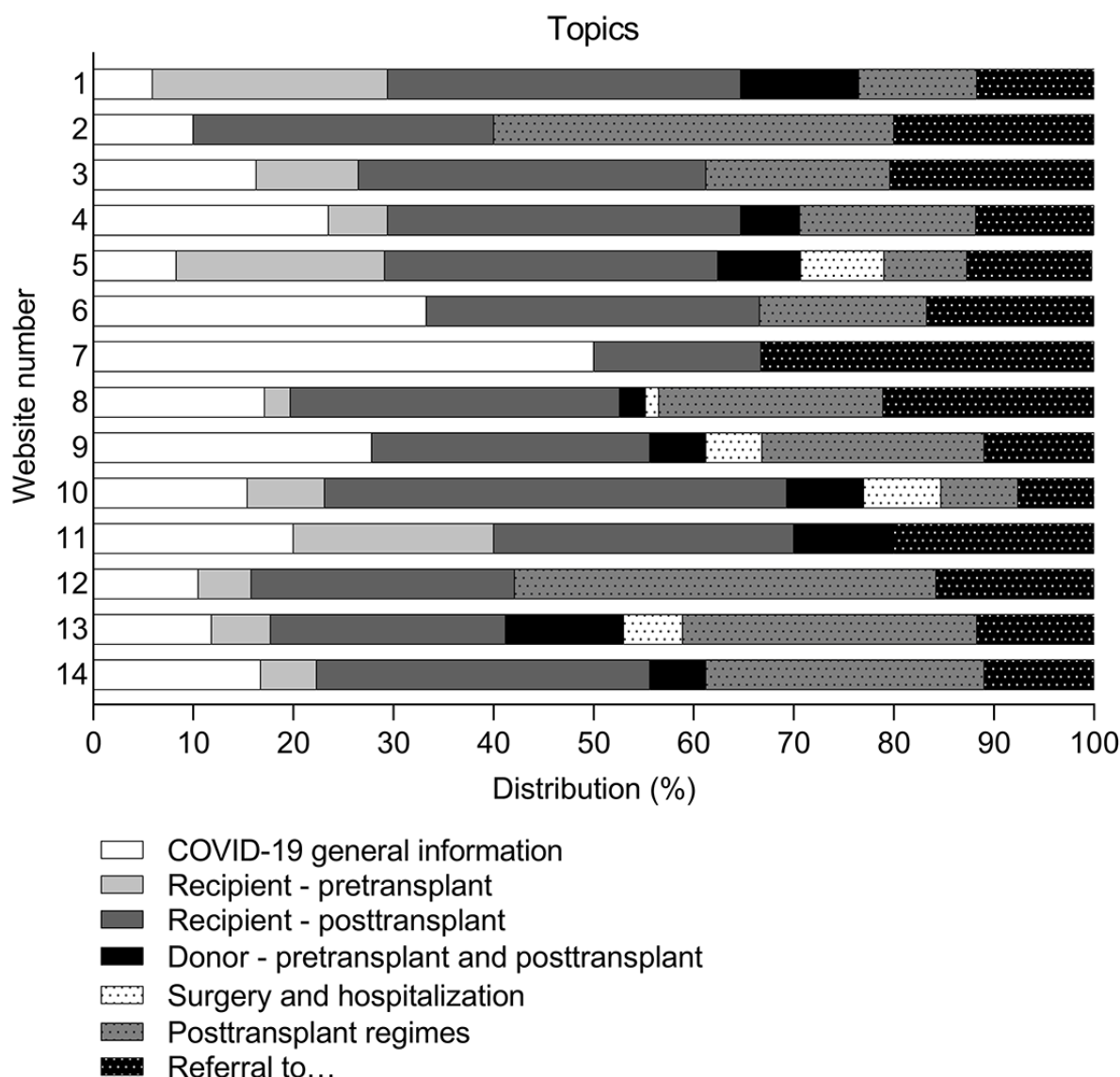
### Organizational Sources of Websites

In total, 14 websites (7 English and 7 Dutch) were analyzed. The source organization of 8 of the 14 websites (57%) was a support group, 4 websites (29%) had a professional nonprofit organizational source, 1 website (7%) was for an individual practice, and 1 website (7%), WikiKids [34], did not match any of the given categories and was therefore labelled as "other." All English websites had an organizational generic top level domain (.org).

### Topics Discussed on the Websites

The topics discussed on all 30 included webpages were analyzed and covered by 7 main themes: COVID-19 general information, recipient–pretransplant, recipient–posttransplant, donor–pretransplant and posttransplant, surgery and hospitalization, posttransplant regimens, and referral to.... Of these 7 main themes, 3 (43%) were discussed on all 14 websites: COVID-19 general information, recipient–posttransplant, and referral to... (Figure 1). The minimum number of different main themes available per website was 3, and the maximum number was 7. A total of 4/14 websites (29%) discussed all main themes.

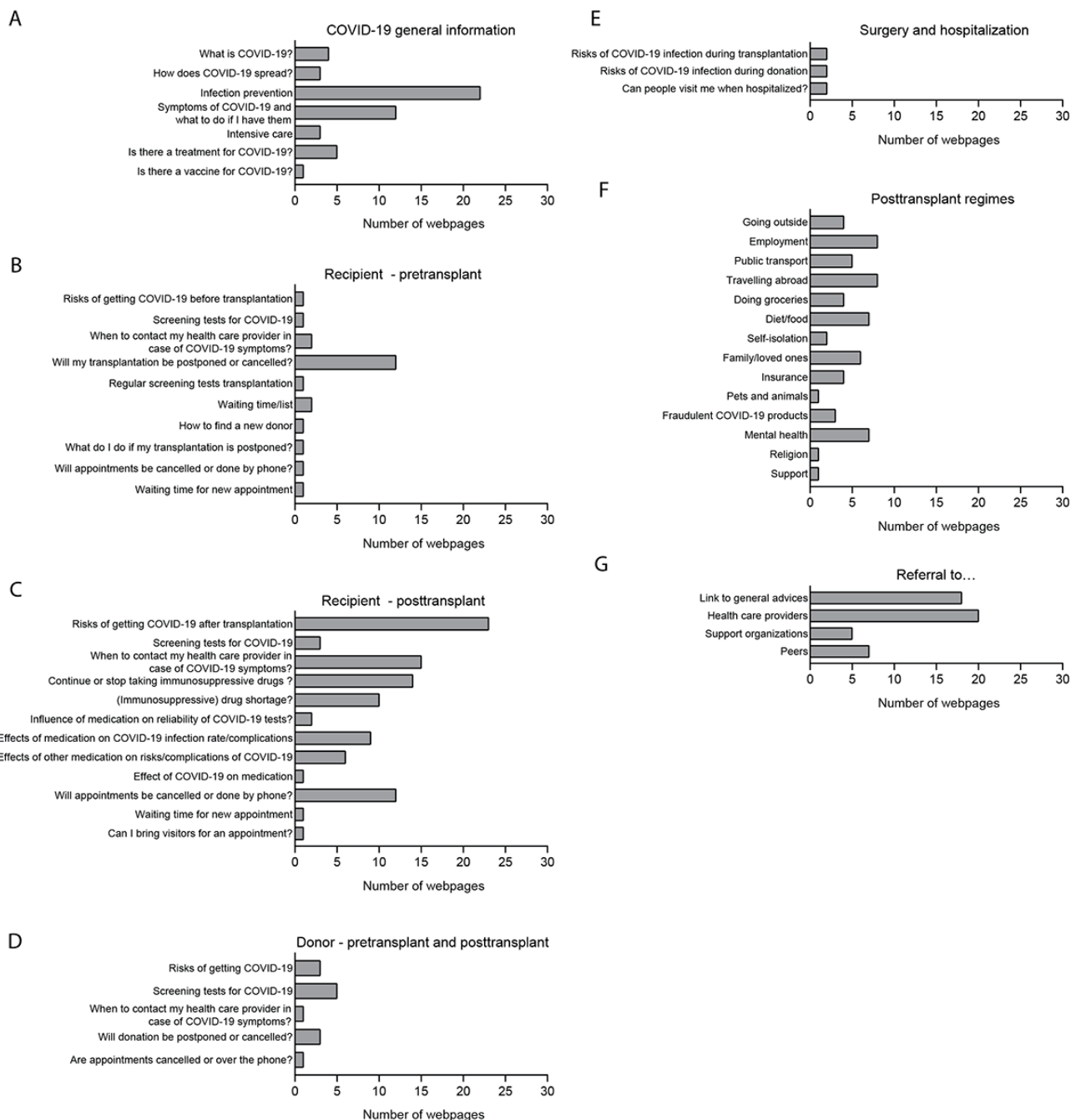
**Figure 1.** Distribution of different topics discussed on the included websites on COVID-19 for patients undergoing renal transplant and living donors, categorized in 7 main themes.



Within the 7 main themes, a total of 56 different topics were discussed (Figure 2). Most topics were related to issues on posttransplant care for recipients, such as risks of contracting COVID-19 after transplant (23/30 webpages, 77%); when to contact health care providers in case of COVID-19 symptoms after transplant (15/30 webpages, 50%); and whether renal transplant recipients should continue or cease taking immunosuppressive therapy (14/30 webpages, 47%). In total, 14 different topics related to regimens were found on the 30

webpages, such as employment (8 webpages, 27%), travelling abroad (8 webpages, 27%), mental health (7 webpages, 23%), and diet (7 webpages, 23%). General information on COVID-19 infection prevention was discussed on 22/30 webpages (73%). Referrals to health care providers were encountered on 20 of the 30 webpages (67%), and 18 webpages (60%) offered links to general advice. Topics related to surgery and hospitalization and living donor information were found on 6/30 webpages (20%) and 13/30 webpages (43%), respectively.

**Figure 2.** Numbers of webpages discussing content topics regarding COVID-19 for patients undergoing renal transplant and living donors. The webpages per included website that discussed the content topics were categorized in 7 main themes: (A) COVID-19 general information, (B) recipient–pretransplant, (C) recipient–posttransplant, (D) donor–pretransplant and posttransplant, (E) surgery and hospitalization, (F) posttransplant regimens, and (G) referral to....



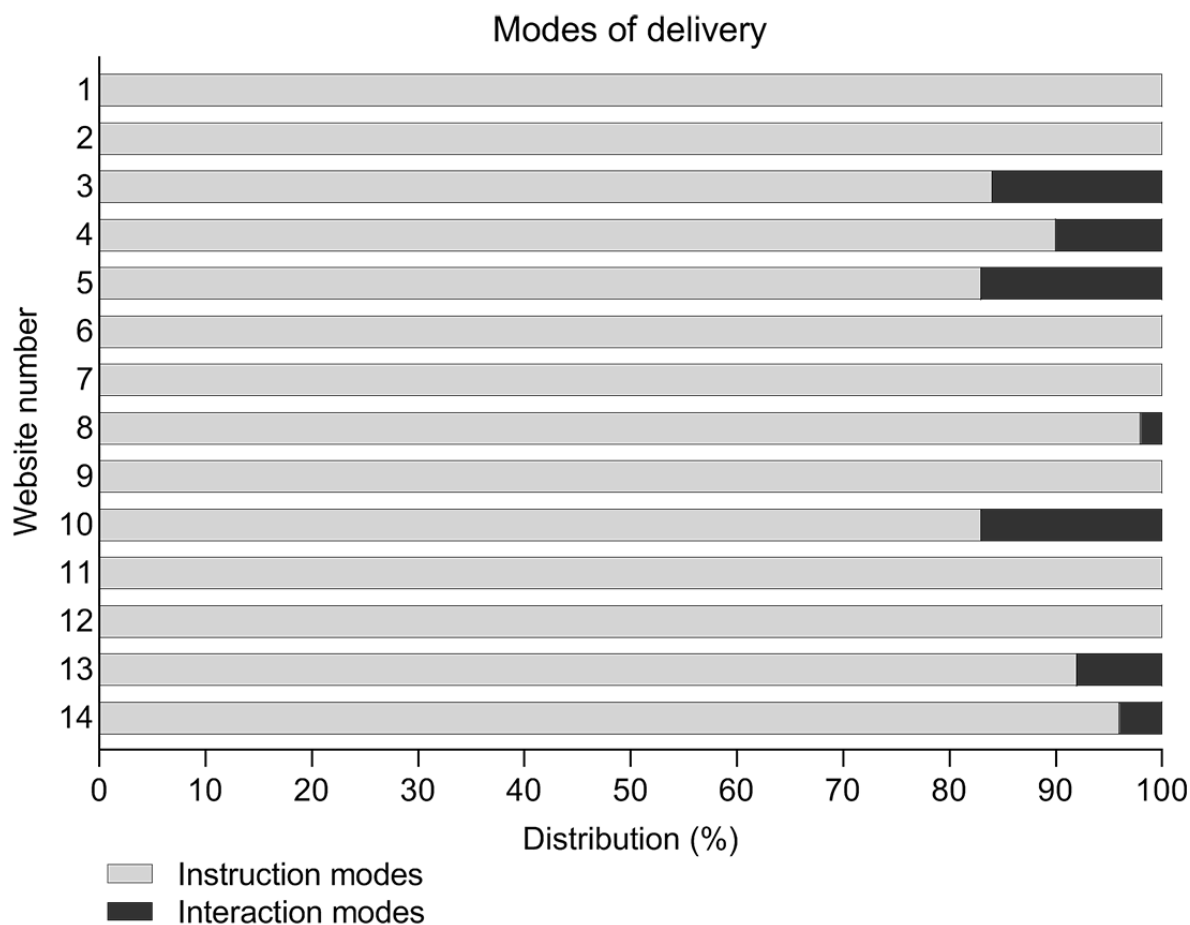
### Delivery Modes on the Websites

A total of 13 different delivery modes were encountered (Table 1). Of 13 these modes, 8 (62%) were instructional, mainly links to external web-based resources and text. Of these 8 instruction modes, 3 (38%) were not predetermined by Toven-Lindsey et al [21] and Hendriks et al [33]: text-to-speech function, instruction video, and documentary. Of the 5 different

interaction modes found, 4 (80%) were not predetermined: question submission form, survey, webinar, and one-on-one chat. Webinars and discussion boards for dialogue were the most commonly offered modes (5 times each). All examined websites offered instructional modes, and 7/14 websites (50%) offered interaction modes (Figure 3). Assessment modes were not observed on any of the included websites.

**Table 1.** Numbers of available delivery modes on the 14 included websites with a total of 30 webpages (N=270), n (%). No assessment modes were encountered.

Delivery mode	Value
<b>Instruction modes</b>	
Text	30 (11.1)
Link to external web-based resource	210 (77.8)
Video of instructor talking to camera	1 (0.4)
Illustration or simulation	9 (3.3)
Digital textbook	2 (0.7)
Documentary	1 (0.4)
Instruction video	2 (0.7)
Text-to-speech	1 (0.4)
<b>Interaction modes</b>	
Discussion board for dialogue	5 (1.9)
One-on-one chat	2 (0.7)
Question submission form	1 (0.4)
Survey	1 (0.4)
Webinar	5 (1.9)

**Figure 3.** Distribution of delivery modes on each website.

The different delivery modes (n=13) were offered 270 times in total, consisting mainly of links to external resources (210, 77.8%) and text (30, 11.1%) (Table 1). Limited variation of available delivery modes was observed between all websites:

256/270 (94.8%) of all available modes were instructional. Text was the only delivery mode that was encountered on all websites (Table 2). Both instruction and interaction modes were offered by 7 of the 14 websites (50%). However, none of the websites

offered all available delivery modes. The minimum number of different modes available per website was 1 (text only), and the maximum number was 7.

**Table 2.** Presence of delivery modes per included website per search engine, classified as instruction or interaction mode.

Search engine	Google.nl							Google.com						
Website number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Webpages, n	1	1	7	1	3	1	1	8	1	1	2	1	1	1
<b>Instruction modes, n</b>														
Text	1	1	7	1	3	1	1	8	1	1	2	1	1	1
Link to external web-based resources	4	1	29	7	2	— <sup>a</sup>	35	73	7	4	4	18	11	15
Video of instructor talking to camera	—	—	—	1	—	—	—	—	—	—	—	—	—	—
Illustration or simulation	—	—	—	—	—	—	2	—	3	—	—	—	—	4
Digital textbook	—	—	—	—	—	—	—	—	—	—	—	—	—	2
Documentary	—	—	1	—	—	—	—	—	—	—	—	—	—	—
Instruction video	—	—	1	—	—	—	—	1	—	—	—	—	—	—
Text-to-speech	—	1	—	—	—	—	—	—	—	—	—	—	—	—
<b>Interaction modes</b>														
Discussion board for dialogue	—	—	5	—	—	—	—	—	—	—	—	—	—	—
One-on-one chat	—	—	—	—	—	—	—	2	—	—	—	—	—	—
Question submission form	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Survey	—	—	1	—	—	—	—	—	—	—	—	—	—	—
Webinar	—	—	1	1	1	—	—	—	—	1	—	—	1	—

<sup>a</sup>—: not applicable.

## Social-Epistemological Dimensions

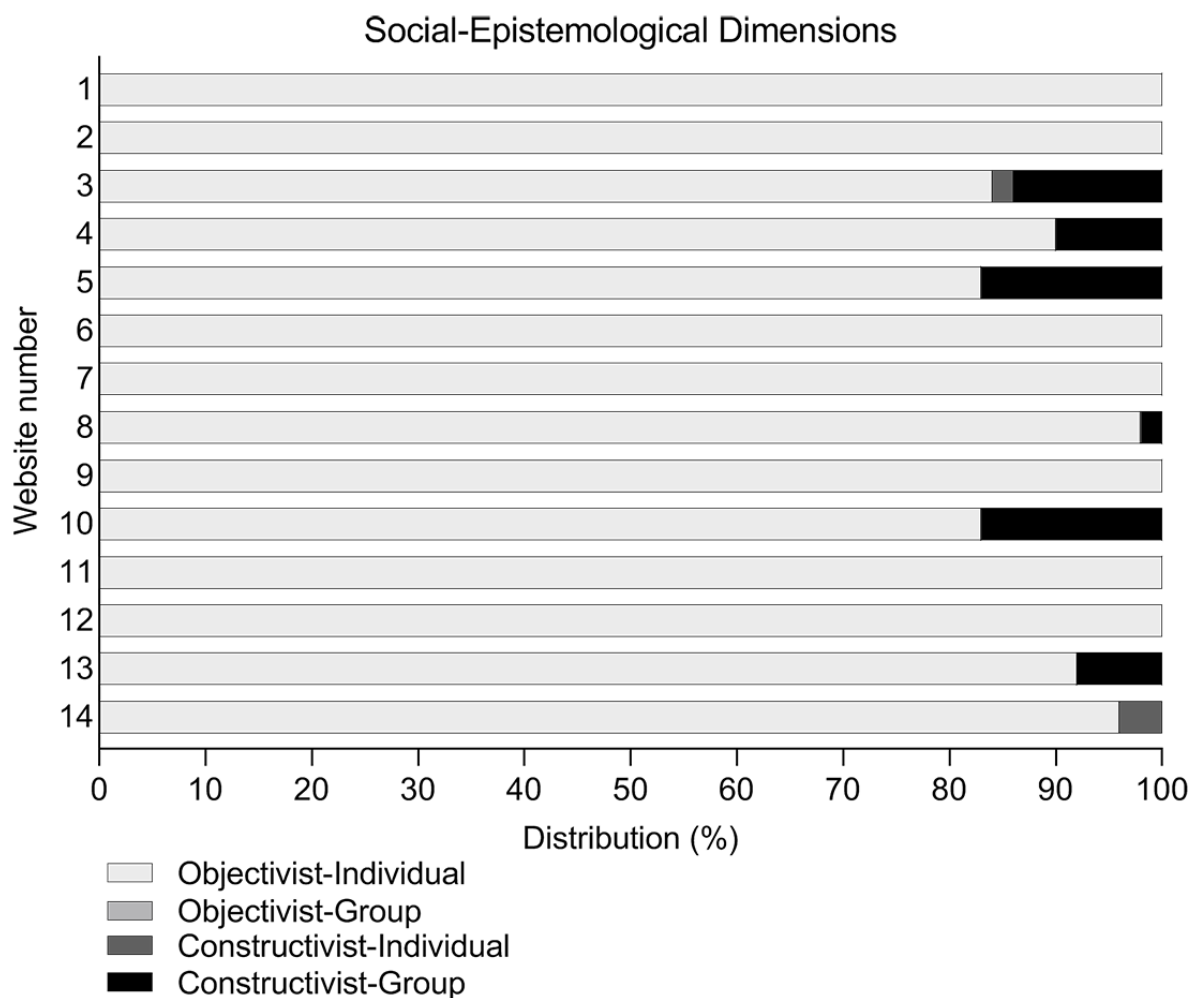
In addition to the previously categorized delivery modes, the 7 nonpredetermined modes were classified into social-epistemological dimensions (Table 3) [21,33]. Text-to-speech functions, instruction videos, and documentaries were classified as objectivist-individual, whereas question submission forms and surveys were categorized as constructivist-individual. Webinars and one-on-one chats were categorized as constructivist-group. Of the 13 different delivery modes available, 8 (62%) were objectivist-individual, 2 (15%)

were constructivist-individual, and 3 (23%) were constructivist-group. None of the offered delivery modes were within the objectivist-group dimension. The websites did not vary in the most commonly observed social-epistemological dimension (Figure 4). All of the 14 examined websites included objectivist-individual delivery modes, whereas constructivist-individual and constructivist-group modes were only offered by 2 (14%) and 6 (43%) websites, respectively. Individual-oriented delivery modes were observed the most frequently, with a minimum of 83% and a maximum of 100% per website.



**Table 3.** Social-epistemological dimensions of the delivery modes based on analysis of the social-epistemological dimensions according to the Teaching Approach Framework of Arbaugh and Benbunan-Fich [27] (N=270), n (%).

Delivery mode	Objectivist-individual	Objectivist-group	Constructivist-individual	Constructivist-group
Text	30 (11.1)	— <sup>a</sup>	—	—
Link to external web-based resource	210 (77.8)	—	—	—
Video of instructor talking to camera	1 (0.4)	—	—	—
Illustration or simulation	9 (1.9)	—	—	—
Digital textbook	2 (0.7)	—	—	—
Documentary	1 (0.4)	—	—	—
Instruction video	2 (0.7)	—	—	—
Text-to-speech	1 (0.4)	—	—	—
Discussion board for dialogue	—	—	—	5 (1.9)
One-on-one chat	—	—	—	2 (0.7)
Question submission form	—	—	1 (0.4)	—
Survey	—	—	1 (0.4)	—
Webinar	—	—	—	5 (1.9)
Total	256 (94.8)	0 (0.0)	2 (0.7)	12 (4.4)

<sup>a</sup>—: not applicable.**Figure 4.** Distribution of social-epistemological dimensions per website.

## Discussion

Since the COVID-19 outbreak, patients undergoing renal transplant and living donors have relied more than ever on telecommunication and web-based information because of fear and uncertainty about COVID-19 and limited in-person visits between the patient and health care provider. The aim of this study was to draw lessons from the topics conveyed and delivery modes used in web-based sources on COVID-19 and renal transplantation. The results show that the variety of content topics, delivery modes, and social-epistemological dimensions was limited. Additionally, the majority of the identified delivery modes focused on objectivistic and individual learning.

In our study, only 3 of the 7 main themes were discussed on all the websites. Additionally, 18 of the 30 webpages (60%) referred users to external sources, and 20 (67%) referred them to health care providers for the latest information and general advice. A logical explanation is that information on COVID-19 rapidly becomes outdated because of the dynamics of the pandemic, and time and financial investments are required to provide the latest information. Regarding the content, the most frequently discussed topics were related to posttransplant care, such as risks of contracting COVID-19 after transplant and whether recipients should continue taking immunosuppressive therapy. The variety of content topics and number of webpages for living donors were limited. Each of the content topics discussed for living donors was also found for transplant recipients. General information about infection prevention, symptoms of COVID-19, and suggestions of what to do if experiencing said symptoms were observed frequently. In addition, only a few webpages discussed issues relating to mental health, employment, insurance, and support. This is in concordance with a previous study that demonstrated that after transplant, the main focus of health care providers is often on dealing with the disease and treatment, whereas patients would also prefer information on managing life after transplantation, including social and emotional support [35].

Previous studies have already focused on the medical quality of web-based information about renal transplants [36,37]. They found that this information is often unvalidated, inaccurate, and unreliable. Here, we focused on information delivery modes because these may influence patients' health literacy (eg, understanding medical information and indirectly promoting behavioral changes and coping strategies) [8,9,27,28]. We found a limited variety of delivery modes; the majority of these modes were instructional, mainly consisting of text and links to external resources. Additionally, we did not find any assessment modes (eg, quizzes) on the websites of interest. Literature shows that there is a need for more interactive patient education. However, the desired education modalities of patients with renal disease are currently unknown [38]. We would suggest that web-based sources should offer more assessment modes because these are crucial to evaluating knowledge and can assist patients by providing insight into their personal goals [21,39]. Moreover, assessing patients' understanding is an important element of promoting prevention behavior [40].

The limited variety in delivery modes is in contrast to another web-based education platform, MOOCs, of which a greater variety can be found in the medical field. This is probably because there are fewer time constraints with MOOCs and a wide team of people are involved in their development, including education professionals [33,41]. MOOCs are not fully comparable to websites, as a MOOC is a course with a beginning and end and contains learning objectives to be achieved. However, taking the educational design of MOOCs into account when developing website content could help improve the knowledge or change the behavior of learners who access the websites. It is interesting to note that the London School of Hygiene & Tropical Medicine recently developed a MOOC for the general public, which focused on understanding and responding to COVID-19 by providing multiple modes: articles, videos, peer reviews, and quizzes [42].

Previous studies demonstrated that actively constructing information instead of passively transferring it results in better and deeper understanding and embedding of knowledge and behavioral changes [24-26]. The vast majority of the delivery modes in our study contained an objectivist-individual dimension and almost no applied constructivist learning. As renal patients must cope with new lifestyle regimens and in-person contact between patients and health care providers is often replaced by telemedicine, patients are expected to take a more active role. This includes monitoring their blood pressure and weight at home. To maintain these behavioral changes, a shift to more constructivist modes of information delivery is recommended.

Moreover, the integration of group learning, such as interaction with peers and participation in group activities, is favorable to learning [27]. Previous studies showed that group-based education for patients with type 2 diabetes promotes disease-specific knowledge, self-empowerment, and drug adherence, and it even improves clinical outcomes compared to individual education [43]. In addition to its effectiveness, patients favor group learning because it enables them to immediately receive answers to questions, discuss experiences and questions with peers, and experience a feeling of community [44]. To incorporate group learning during times of social distancing and limited in-person visits, a webinar may be a good option because this delivery mode offers the possibility of synchronous web-based interactive conversation between patients, living donors, and health care providers [45]. Additionally, contact with peers can increase patients' self-management, and group education settings can help patients to overcome feelings of isolation [46]. Moreover, implementing webinars in transplant care could help health care providers tailor information to patients' information needs because patients and living donors can submit questions beforehand.

A limitation of our study is that we mapped the available delivery modes on the included webpages at a single time point in April 2020. However, the variety of delivery modes may change over time. Therefore, future studies should analyze available delivery modes at multiple time points to investigate the dynamics and compare differences.

In conclusion, the variety of topics and delivery modes of web-based information on COVID-19 for patients who undergo renal transplants and living donors is limited. We therefore recommend providing information on COVID-19 in more diverse and interactive ways. Additionally, web-based sources

should focus more on knowledge construction than on passive information transfer, and they should take interactivity into account. This is particularly important in times of crisis, when rapid knowledge transfer is needed.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Keyword combinations used to search for websites on renal transplantation and COVID-19 using the search engines Google.com and Google.nl.

[PNG File , 26 KB - [jmir\\_v22i10e22068\\_app1.png](#) ]

## References

1. de Vries APJ, Alwayn IPJ, Hoek RAS, van den Berg AP, Ultee FCW, Vogelaar SM, et al. Immediate impact of COVID-19 on transplant activity in the Netherlands. *Transpl Immunol* 2020 Aug;61:101304 [FREE Full text] [doi: [10.1016/j.trim.2020.101304](#)] [Medline: [32371150](#)]
2. Kumar D, Manuel O, Natori Y, Egawa H, Grossi P, Han S, et al. COVID-19: A global transplant perspective on successfully navigating a pandemic. *Am J Transplant* 2020 Jul;20(7):1773-1779 [FREE Full text] [doi: [10.1111/ajt.15876](#)] [Medline: [32202064](#)]
3. Loupy A, Aubert O, Reese PP, Bastien O, Bayer F, Jacquelinet C. Organ procurement and transplantation during the COVID-19 pandemic. *Lancet* 2020 May 23;395(10237):e95-e96 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)31040-0](#)] [Medline: [32407668](#)]
4. Correia MITD, Ramos RF, Bahten LCV. The surgeons and the COVID-19 pandemic. *Rev Col Bras Cir* 2020;47:e20202536. [doi: [10.1590/0100-6991e-20202536](#)] [Medline: [32236295](#)]
5. Vidal-Alaball J, Acosta-Roja R, Pastor Hernández N, Sanchez Luque U, Morrison D, Narejos Pérez S, et al. Telemedicine in the face of the COVID-19 pandemic. *Aten Primaria* 2020;52(6):418-422 [FREE Full text] [doi: [10.1016/j.aprim.2020.04.003](#)] [Medline: [32402477](#)]
6. Fisk M, Livingstone A, Pit SW. Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States. *J Med Internet Res* 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: [10.2196/19264](#)] [Medline: [32463377](#)]
7. Capurro D, Ganzinger M, Perez-Lu J, Knaup P. Effectiveness of eHealth interventions and information needs in palliative care: a systematic literature review. *J Med Internet Res* 2014 Mar 07;16(3):e72 [FREE Full text] [doi: [10.2196/jmir.2812](#)] [Medline: [24610324](#)]
8. Hunt HF, Rodrigue JR, Dew MA, Schaffer RL, Henderson ML, Bloom R, et al. Strategies for Increasing Knowledge, Communication, and Access to Living Donor Transplantation: an Evidence Review to Inform Patient Education. *Curr Transplant Rep* 2018 Mar;5(1):27-44 [FREE Full text] [doi: [10.1007/s40472-018-0181-1](#)] [Medline: [30873335](#)]
9. Hanif F, Read JC, Clancy MJ. A snapshot of renal transplant patients using medical web browsing. *Exp Clin Transplant* 2012 Jun;10(3):219-223. [doi: [10.6002/ect.2011.0154](#)] [Medline: [22631056](#)]
10. Jiang S, Beaudoin CE. Health literacy and the internet: An exploratory study on the 2013 HINTS survey. *Computers in Human Behavior* 2016 May 27;58(1):240-248 [FREE Full text] [doi: [10.1016/j.chb.2016.01.007](#)]
11. Institute of Medicine, Board on Neuroscience and Behavioral Health. 1. Introduction. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. *Health Literacy: A Prescription to End Confusion*. Washington, DC: National Academies Press; 2004.
12. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011 Jul 19;155(2):97-107. [doi: [10.7326/0003-4819-155-2-201107190-00005](#)] [Medline: [21768583](#)]
13. Taylor DM, Fraser S, Dudley C, Oniscu GC, Tomson C, Ramanan R, ATTOM investigators. Health literacy and patient outcomes in chronic kidney disease: a systematic review. *Nephrol Dial Transplant* 2018 Sep 01;33(9):1545-1558. [doi: [10.1093/ndt/gfx293](#)] [Medline: [29165627](#)]
14. Rutherford EJ, Kelly J, Lehane EA, Livingstone V, Cotter B, Butt A, et al. Health literacy and the perception of risk in a breast cancer family history clinic. *Surgeon* 2018 Apr;16(2):82-88. [doi: [10.1016/j.surge.2016.06.003](#)] [Medline: [27908542](#)]
15. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. *Am J Health Behav* 2007;31 Suppl 1:S19-S26. [doi: [10.5555/ajhb.2007.31.supp.S19](#)] [Medline: [17931132](#)]
16. Chisholm-Burns MA, Spivey CA, Pickett LR. Health literacy in solid-organ transplantation: a model to improve understanding. *Patient Prefer Adherence* 2018;12:2325-2338 [FREE Full text] [doi: [10.2147/PPA.S183092](#)] [Medline: [30464420](#)]

17. Taylor DM, Fraser SDS, Bradley JA, Bradley C, Draper H, Metcalfe W, ATTOM investigators. A Systematic Review of the Prevalence and Associations of Limited Health Literacy in CKD. *Clin J Am Soc Nephrol* 2017 Jul 07;12(7):1070-1084. [doi: [10.2215/CJN.12921216](https://doi.org/10.2215/CJN.12921216)] [Medline: [28487346](https://pubmed.ncbi.nlm.nih.gov/28487346/)]
18. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. *Ann Behav Med* 2009 Aug;38(1):18-27 [FREE Full text] [doi: [10.1007/s12160-009-9133-4](https://doi.org/10.1007/s12160-009-9133-4)] [Medline: [19802647](https://pubmed.ncbi.nlm.nih.gov/19802647/)]
19. Bellamy R. An introduction to patient education: theory and practice. *Med Teach* 2004 Jun;26(4):359-365. [doi: [10.1080/01421590410001679398](https://doi.org/10.1080/01421590410001679398)] [Medline: [15203851](https://pubmed.ncbi.nlm.nih.gov/15203851/)]
20. Pappot N, Taarnhøj GA, Pappot H. Telemedicine and e-Health Solutions for COVID-19: Patients' Perspective. *Telemed J E Health* 2020 Apr 24;26(7):online. [doi: [10.1089/tmj.2020.0099](https://doi.org/10.1089/tmj.2020.0099)] [Medline: [32329654](https://pubmed.ncbi.nlm.nih.gov/32329654/)]
21. Toven-Lindsey B, Rhoads R, Lozano J. Virtually unlimited classrooms: Pedagogical practices in massive open online courses. *The Internet and Higher Education* 2015 Jan;24:1-12. [doi: [10.1016/j.iheduc.2014.07.001](https://doi.org/10.1016/j.iheduc.2014.07.001)]
22. Lustria MLA. Can interactivity make a difference? Effects of interactivity on the comprehension of and attitudes toward online health content. *J Am Soc Inf Sci* 2007 Apr;58(6):766-776. [doi: [10.1002/asi.20557](https://doi.org/10.1002/asi.20557)]
23. Gance S. Are Constructivism and Computer-Based Learning Environments Incompatible? *JAHN* 2002 May;5(1):online.
24. Piaget J. The psychogenesis of knowledge and its epistemological significance. In: Piattelli-Palmarini M, editor. *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press; Dec 1981.
25. Phillips D. The Good, the Bad, and the Ugly: The Many Faces of Constructivism. *Educational Researcher* 1995 Oct;24(7):5-12. [doi: [10.2307/1177059](https://doi.org/10.2307/1177059)]
26. Bada SO. Constructivism Learning Theory: A Paradigm for Teaching and Learning. *IOSR-JRME* 2015;5(6):66-70 [FREE Full text]
27. Arbaugh JB, Benbunan-Finch R. An Investigation of Epistemological and Social Dimensions of Teaching in Online Learning Environments. *AMLE* 2006 Dec;5(4):435-447. [doi: [10.5465/amle.2006.23473204](https://doi.org/10.5465/amle.2006.23473204)]
28. Ubbes V, Black J, Ausherman J. Constructing knowledge for understanding in health education: The role of critical and creative thinking skills within constructivism theory. *J Health Educ* 1999;30(2):67-72.
29. Search Engine Host Market Share Worldwide. Statcounter. URL: <https://gs.statcounter.com/search-engine-host-market-share> [accessed 2020-03-27]
30. Google (United States). URL: <https://www.google.com/> [accessed 2020-10-02]
31. Google (Netherlands). URL: <https://www.google.nl/> [accessed 2020-10-02]
32. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. *Cyberpsychol Behav* 2004 Oct;7(5):497-510. [doi: [10.1089/cpb.2004.7.497](https://doi.org/10.1089/cpb.2004.7.497)] [Medline: [15667044](https://pubmed.ncbi.nlm.nih.gov/15667044/)]
33. Hendriks RA, de Jong PGM, Admiraal WF, Reinders MEJ. Teaching modes and social-epistemological dimensions in medical Massive Open Online Courses: Lessons for integration in campus education. *Med Teach* 2019 Aug;41(8):917-926. [doi: [10.1080/0142159X.2019.1592140](https://doi.org/10.1080/0142159X.2019.1592140)] [Medline: [31008667](https://pubmed.ncbi.nlm.nih.gov/31008667/)]
34. WikiKids. Webpage in Dutch. URL: <https://wikikids.nl> [accessed 2020-10-02]
35. Schmid-Mohler G, Schäfer-Keller P, Frei A, Fehr T, Spirig R. A mixed-method study to explore patients' perspective of self-management tasks in the early phase after kidney transplant. *Prog Transplant* 2014 Mar;24(1):8-18. [doi: [10.7182/pit2014728](https://doi.org/10.7182/pit2014728)] [Medline: [24598560](https://pubmed.ncbi.nlm.nih.gov/24598560/)]
36. Hanif F, Abayasekara K, Willcocks L, Jolly EC, Jamieson NV, Praseedom RK, et al. The quality of information about kidney transplantation on the World Wide Web. *Clin Transplant* 2007;21(3):371-376. [doi: [10.1111/j.1399-0012.2006.00652.x](https://doi.org/10.1111/j.1399-0012.2006.00652.x)] [Medline: [17488387](https://pubmed.ncbi.nlm.nih.gov/17488387/)]
37. Valizadeh-Haghi S, Rahmatizadeh S. Evaluation of the Quality and Accessibility of Available Websites on Kidney Transplantation. *Urol J* 2018 Sep 26;15(5):261-265 [FREE Full text] [doi: [10.22037/uj.v0i0.4252](https://doi.org/10.22037/uj.v0i0.4252)] [Medline: [30058062](https://pubmed.ncbi.nlm.nih.gov/30058062/)]
38. Cassidy BP, Getchell LE, Harwood L, Hemmett J, Moist LM. Barriers to Education and Shared Decision Making in the Chronic Kidney Disease Population: A Narrative Review. *Can J Kidney Health Dis* 2018;5:2054358118803322. [doi: [10.1177/2054358118803322](https://doi.org/10.1177/2054358118803322)] [Medline: [30542621](https://pubmed.ncbi.nlm.nih.gov/30542621/)]
39. Rodrigue JR, Hanto DW, Curry MP. Patients' expectations and success criteria for liver transplantation. *Liver Transpl* 2011 Nov;17(11):1309-1317 [FREE Full text] [doi: [10.1002/lt.22355](https://doi.org/10.1002/lt.22355)] [Medline: [21656656](https://pubmed.ncbi.nlm.nih.gov/21656656/)]
40. Krist AH, Tong ST, Aycock RA, Longo DR. Engaging Patients in Decision-Making and Behavior Change to Promote Prevention. *Stud Health Technol Inform* 2017;240:284-302 [FREE Full text] [Medline: [28972524](https://pubmed.ncbi.nlm.nih.gov/28972524/)]
41. Reinders MEJ, de Jong PGM. Innovating clinical kidney transplant education by a massive open online course. *Transpl Immunol* 2016 Sep;38:1-2. [doi: [10.1016/j.trim.2016.06.003](https://doi.org/10.1016/j.trim.2016.06.003)] [Medline: [27373761](https://pubmed.ncbi.nlm.nih.gov/27373761/)]
42. COVID-19: Tackling the Novel Coronavirus. FutureLearn. URL: <https://www.futurelearn.com/courses/covid19-novel-coronavirus> [accessed 2020-10-02]
43. Deakin T, McShane CE, Cade JE, Williams RDRR. Group based training for self-management strategies in people with type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2005 Apr 18(2):CD003417. [doi: [10.1002/14651858.CD003417.pub2](https://doi.org/10.1002/14651858.CD003417.pub2)] [Medline: [15846663](https://pubmed.ncbi.nlm.nih.gov/15846663/)]

44. Salmasi S, Kwan L, MacGillivray J, Bansback N, De Vera MA, Barry AR, et al. Assessment of atrial fibrillation patients' education needs from patient and clinician perspectives: A qualitative descriptive study. *Thromb Res* 2019 Jan;173:109-116. [doi: [10.1016/j.thromres.2018.11.015](https://doi.org/10.1016/j.thromres.2018.11.015)] [Medline: [30502679](https://pubmed.ncbi.nlm.nih.gov/30502679/)]
45. Chiswell M, Smissen A, Ugalde A, Lawson D, Whiffen R, Brockington S, et al. Using Webinars for the Education of Health Professionals and People Affected by Cancer: Processes and Evaluation. *J Cancer Educ* 2018 Jun;33(3):583-591. [doi: [10.1007/s13187-016-1138-7](https://doi.org/10.1007/s13187-016-1138-7)] [Medline: [27981435](https://pubmed.ncbi.nlm.nih.gov/27981435/)]
46. Klang B, Björvell H, Clyne N. Predialysis education helps patients choose dialysis modality and increases disease-specific knowledge. *J Adv Nurs* 1999 Apr;29(4):869-876. [doi: [10.1046/j.1365-2648.1999.00957.x](https://doi.org/10.1046/j.1365-2648.1999.00957.x)] [Medline: [10215978](https://pubmed.ncbi.nlm.nih.gov/10215978/)]

## Abbreviations

**MOOC:** massive open online course

*Edited by G Eysenbach; submitted 02.07.20; peer-reviewed by A Banbury; comments to author 20.07.20; revised version received 28.07.20; accepted 10.09.20; published 08.10.20.*

### *Please cite as:*

van Klaveren CW, de Jong PGM, Hendriks RA, Luk F, de Vries APJ, van der Boog PJM, Reinders MEJ

*Topics, Delivery Modes, and Social-Epistemological Dimensions of Web-Based Information for Patients Undergoing Renal Transplant and Living Donors During the COVID-19 Pandemic: Content Analysis*

*J Med Internet Res* 2020;22(10):e22068

URL: <http://www.jmir.org/2020/10/e22068/>

doi: [10.2196/22068](https://doi.org/10.2196/22068)

PMID: [32946412](https://pubmed.ncbi.nlm.nih.gov/32946412/)

©Charlotte W van Klaveren, Peter G M de Jong, Renée A Hendriks, Franka Luk, Aiko P J de Vries, Paul J M van der Boog, Marlies E J Reinders. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Associations of Mental Health and Personal Preventive Measure Compliance With Exposure to COVID-19 Information During Work Resumption Following the COVID-19 Outbreak in China: Cross-Sectional Survey Study

Yihang Pan<sup>1,2\*</sup>, MD; Meiqi Xin<sup>3\*</sup>, PhD; Changhua Zhang<sup>4\*</sup>, PhD; Willa Dong<sup>5</sup>, MSPH; Yuan Fang<sup>6</sup>, PhD; Wenhui Wu<sup>4</sup>, PhD; Mingzhe Li<sup>4</sup>, MD; Jun Pang<sup>7</sup>, MD; Zilong Zheng<sup>2</sup>, MD; Zixin Wang<sup>3\*</sup>, PhD; Jinqiu Yuan<sup>1,2,8\*</sup>, PhD; Yulong He<sup>4\*</sup>, MD

<sup>1</sup>Precision Medicine Center, Scientific Research Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

<sup>2</sup>Big Data Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

<sup>3</sup>JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong

<sup>4</sup>Center for Digestive Disease, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

<sup>5</sup>Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

<sup>6</sup>Department of Early Childhood Education, The Education University of Hong Kong, Hong Kong

<sup>7</sup>Center for Urology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

<sup>8</sup>Clinical Research Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

\*these authors contributed equally

**Corresponding Author:**

Zixin Wang, PhD

JC School of Public Health and Primary Care

The Chinese University of Hong Kong

Room 508, School of Public Health, Prince of Wales Hospital

Shatin, NT

666888

Hong Kong

Phone: 852 22528740

Email: [wangzx@cuhk.edu.hk](mailto:wangzx@cuhk.edu.hk)

## Abstract

**Background:** Risk and crisis communication plays an essential role in public health emergency responses. The COVID-19 pandemic has triggered spontaneous and intensive media attention, which has affected people's adoption of personal preventive measures and their mental health.

**Objective:** The aim of this study was to investigate the associations between exposure to COVID-19-specific information and mental health (depression and sleep quality) and self-reported compliance with personal preventive measures (face mask wearing and hand sanitizing). We also tested whether these associations were moderated by thoughtful consideration of the veracity of the information to which people were exposed.

**Methods:** A cross-sectional, closed web-based survey was conducted among a sample of 3035 factory workers at the beginning of work resumption following the COVID-19 outbreak in Shenzhen, China. A stratified two-stage cluster sampling design was used for recruitment. Multivariate linear and logistic regression models were used for the analyses.

**Results:** The prevalence of probable moderate-to-severe depression was 170/3035 (5.6%), while that of good or excellent sleep quality was 2110/3035 (69.5%). The prevalence of self-reported consistent face mask wearing in public places was 2903/3035 (95.7%), while that of sanitizing hands every time after returning from public spaces or touching public installations was 2151/3035 (70.9%). Of the 3035 respondents, 1013 to 1638 (33.3% to 54.0%) reported >1 hour of daily exposure to COVID-19-specific information through web-based media and television. After controlling for significant background variables, higher information exposure via television and via newspapers and magazines was associated with better sleep quality and higher compliance with hand sanitizing. Higher exposure via unofficial web-based media was associated with higher compliance with hand sanitizing

but was also associated with higher depressive symptoms. In contrast, higher exposure through face-to-face communication was associated with higher depressive symptoms, worse sleep quality, and lower compliance with hand sanitizing. Exposure to information about positive outcomes for patients with COVID-19, development of vaccines and effective treatments, and heroic stories about frontline health care workers were associated with both better mental health and higher compliance with preventive measures. Higher overall information exposure was associated with higher depressive symptoms among participants who were less likely to carefully consider the veracity of the information to which they were exposed; it was also associated with better sleep quality among people who reported more thoughtful consideration of information veracity.

**Conclusions:** This study provides empirical evidence of how the amount, sources, and contents of information to which people were exposed influenced their mental health and compliance with personal preventive measures at the initial phase of work resumption in China. Thoughtful consideration of information quality was found to play an important moderating role. Our findings may inform strategic risk communication by government and public health authorities during the COVID-19 pandemic.

(*J Med Internet Res* 2020;22(10):e22596) doi:[10.2196/22596](https://doi.org/10.2196/22596)

## KEYWORDS

COVID-19; information exposure; risk; communication; mental health; personal preventive measures; China; cross-sectional; public health; prevention

## Introduction

Risk and crisis communication plays an essential role in public health emergency responses [1]. During public health emergencies, the media is an essential tool used by government and public health authorities to manage crises [2], and the public relies on media to understand the situation [3]. Infectious disease pandemics, such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) trigger spontaneous and intensive media attention, which could have effects on various public responses [4,5]. One of the most desirable public responses is the adoption of personal preventive measures. Universal face mask wearing [6] and hand hygiene [7] are strongly advocated by the World Health Organization (WHO) and have been widely implemented during the COVID-19 pandemic [8,9]. Media exposure has been demonstrated to increase health knowledge among the public, which in turn encourages desirable preventive behaviors during infectious disease pandemics [4,10].

In addition, people facing a public health emergency often experience a range of negative emotions. During the COVID-19 outbreak, the unknown cause of the disease, fatal outcomes, and interruption of daily routines likely increased levels of negative emotions [11]. The COVID-19 pandemic also triggered mental health problems among the general public, such as stress, panic, depression, and anxiety [12-14]. The media tend to overemphasize risk and sensationalize crises, and repeated information exposure through the media may result in public panic and fear. According to Social Amplification Theory, this panic and fear can be further amplified through the exchange of information on diverse media platforms or through face-to-face communication [15]. Mental distress provoked by information exposure can fuel the proliferation of misinformation and amplify adverse health outcomes and maladaptive responses [16-18].

The importance of web-based media during public health emergencies has been increasing over time. Individuals have turned to web-based media during pandemics to find information related to safety precautions, news updates, and protective equipment. For example, during the SARS and MERS outbreaks,

Chinese and South Korean people turned to the internet to find information that was unavailable on traditional media channels [19,20]. During the early phase of the COVID-19 outbreak in 2020, 94% of Hong Kong residents considered social media to be their most important information source [21]. Compared to traditional media, web-based media platforms (such as social media) provide not only information but also emotional expression. It has been observed that negative emotions are more likely to be conveyed on social media than positive emotions during an infectious disease pandemic [22]. Another study suggested that social media has had a significant impact on the spread of fear and panic during the COVID-19 outbreak [23]. Studies conducted during the MERS outbreak of 2015 indicated that increased exposure to MERS-specific information through social media was simultaneously associated with higher adoption of personal preventive measures and higher levels of negative emotions [4,24]. However, traditional media consumption (eg, television and newspapers) did not influence personal preventive measures or mental health [4,24].

Previous studies have shown that people are likely to react emotionally rather than rationally during a disease outbreak [25]. For example, in one study [26], participants were less able to pay attention to external information and did not trust authorities. Moreover, they were more likely to absorb negative rather than positive information. Although web-based media is a powerful tool for disseminating information, there are concerns related to inaccurate data, unverified rumors, and even malicious misinformation on these platforms [4]. With the COVID-19 outbreak, a global epidemic of misinformation has also been spreading rapidly through social media, which poses a serious additional challenge for public health efforts [27,28]. In response, the WHO and many other public health authorities are verifying rumors and providing evidence-based information [29,30]. In this information environment, it is also important to understand how audiences process the obtained information (eg, the extent of thoughtful consideration of the veracity of health information) and whether these processes will affect their behaviors and mental health during the pandemic.

This study targeted workers who resumed work in China at the beginning of work resumption. Although the implementation

of strict nationwide control measures (eg, closure of all unessential businesses) effectively controlled the pandemic, it severely damaged China's economy [31]. Therefore, the Chinese government scaled up work resumption starting on February 20, 2020 [32]. However, there were concerns that the increase in public contact after work resumption would result in a second wave of COVID-19 in China [33]. Implementing effective risk communication can help workers comply with personal preventive measures without amplifying panic or negative emotions, which is crucial to achieve a balance of disease control with revitalization of the economy.

To the best of our knowledge, there have been no studies investigating behavioral or mental health outcomes of exposure to COVID-19-specific information at the initial phase of work resumption following the COVID-19 outbreak. The objective of this study was to investigate the associations between COVID-19-specific information exposure and four outcome variables, including depression, sleep quality, self-reported consistent face mask wearing, and hand hygiene. The study also tested whether the associations between information exposure and outcomes varied as a function of thoughtful consideration

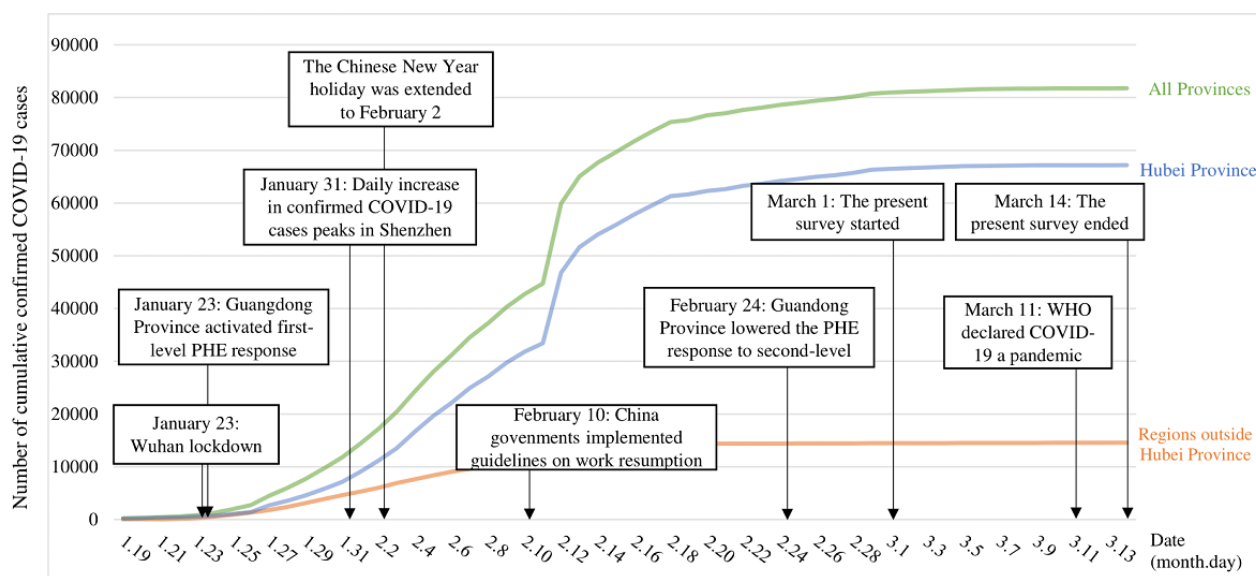
of the veracity of information to which the participants were exposed.

## Methods

### Settings and Participants

A cross-sectional, closed web-based survey was conducted from March 1 to 14, 2020, in Shenzhen, China. Approximately 65.1% of the residents of this city are internal migrants, and 34.3% are factory workers [34]. During the COVID-19 pandemic, Shenzhen has been one of the most severely affected regions in China outside of Hubei Province, which is the epicenter of the pandemic. The daily increase in confirmed COVID-19 cases peaked at 60 on January 31. By March 1, 2020, 100 factories in Shenzhen had resumed work. A stratified two-stage cluster sampling approach was used for recruitment. First, 14 factories were randomly selected by the research team. Of these 14 factories, 10 (71%) manufactured electronic devices, 2 (14%) manufactured watches, 1 (7%) manufactured beverages, and 1 (7%) manufactured biotechnology products. All full-time employees aged  $\geq 18$  years who had resumed work in these factories were then invited to complete a web-based survey. Details of the study background are shown in Figure 1.

**Figure 1.** Background of the present survey, including the trend of cumulative confirmed COVID-19 cases in mainland China and critical responses to COVID-19 in Shenzhen, a city in Guangdong Province. PHE: public health emergency; WHO: World Health Organization.



### Procedure

We developed a web-based questionnaire using Questionnaire Star, a commonly used web-based survey platform in China. The link to access the web-based questionnaire was shared in WeChat, a popular messaging app. In addition to national guidelines, the Shenzhen government requested that each factory establish WeChat groups including all employees as part of the preparation for work resumption. A designated coordinator responsible for COVID-19 control in each factory facilitated the data collection. This coordinator posted the study information and the link to access the web-based questionnaire in the WeChat group and invited all eligible workers who had resumed work to participate. The coordinator also sent out

reminders in the WeChat groups twice per week during the study period. These designated coordinators did not participate in the survey. The coordinators and participants were asked not to disseminate the survey access link to people outside their factories. Before starting the web-based survey, the participants read a statement indicating that participation was voluntary, refusal to participate would have no effect on them, the survey would not collect personal contacts or identifying information, and the data would be kept strictly confidential and would only be used for research purposes. Web-based informed consent was sought. Each individual WeChat account was allowed to access the web-based survey once. The web-based survey platform performed completeness checks before each questionnaire was submitted. Participants were able to review

and change their responses using a Back button. Upon completion of the survey, an electronic coupon of ¥10 (US \$1.30) was sent to each participant. All data were stored in the server of the survey platform and were protected by a password. Only the corresponding authors had access to the database. Of 4158 workers who had resumed work in these factories on March 1, 2020, 3035 completed the web-based survey. The overall response rate was 73.0%. Ethics approval was obtained from the Seventh Affiliated Hospital, Sun Yat-sen University (reference: KY-2020-005-001).

## Measures

### Background Variables

Sociodemographic data such as age, gender, highest education level, marital status, monthly personal income, status as frontline workers or management staff, and type of factories the respondents were working in were collected.

### COVID-19-Specific Information Exposure

Five items assessed daily average time (hours) of COVID-19-specific information exposure through different sources from January 24 (the first day of the Chinese New Year holiday) through the survey date. Sources of health information included official web-based media (media accounts of government institutions), unofficial web-based media (media accounts of private organizations or self-published accounts), television, newspapers and magazines, and face-to-face communication. The Overall Information Exposure Scale was constructed by summing the individual item scores (1=almost none, 2=<1 hour, 3=1-2 hours, 4=3-4 hours, and 5=>4 hours). A higher score on the scale indicated exposure to a higher amount of COVID-19-specific information.

Participants were also asked to report their frequencies of exposure to information on the eight most popular topics related to COVID-19 during the outbreak (0=almost never to 3=always). These topics included statistics about the COVID-19 pandemic (number of confirmed or suspected cases, deaths, and recoveries), positive information about governmental responses to COVID-19 (eg, building mobile clinics, mobilizing national resources to epidemic areas), negative information about governmental responses to COVID-19 (eg, insufficient supply of protective equipment by some local governmental organizations), information about the development of vaccines and effective treatments, heroic stories about frontline health care workers, positive information about patients with COVID-19 (eg, recovery, discharge from hospital), negative information about patients with COVID-19 (eg, severe symptoms and deaths, other negative consequences for patients and their families), and information about negative impacts of the COVID-19 pandemic on the economy (eg, bankruptcy of businesses, layoffs, pay cuts).

A single item assessed the frequency of thoughtful consideration of the veracity of health information: “When you viewed information about COVID-19, how often did you carefully think about its veracity?” (0=almost never to 3=always). This variable was treated as a moderator.

### Mental Health

Depressive symptoms were measured using the validated Chinese version of the 9-item Patient Health Questionnaire (PHQ-9) [35]. Participants were asked to rate the frequency at which they experienced each of nine depressive symptoms over the past two weeks (0=never to 3=nearly every day). A mean score was calculated and used for analysis (alpha reliability=.91). A summative score  $\geq 10$  indicated the presence of probable moderate to severe depression [36].

Global sleep quality over a 7-day recall period was measured using a single-item Sleep Quality Scale with a rating range from 0 to 10. A higher score indicated better sleep quality. The original score was then recoded into four categories: 1=poor or terrible (0-3), 2=fair (4-6), 3=good (7-9), and 4=excellent (10) [37]. This instrument has been demonstrated to be a reliable and valid measure without significantly increasing respondents' survey burden.

### Self-Reported Compliance With Personal Preventive Measures

Participants were asked to report the frequency at which they wore face masks in the workplace and in other public settings (public places and transportation) in the past month (response categories: every time, often, sometimes, never). A composite variable of “consistent face mask wearing” was created to represent respondents who reported wearing a face mask every time both in the workplace and in other public settings (0=no, 1=yes). Participants also reported the frequency at which they sanitized their hands using soaps, liquid soaps, or alcohol-based hand sanitizers after returning from public spaces or touching public installations and equipment (eg, handrails, escalator control panels, or door knobs) (response categories: every time, often, sometimes, never). This item was also recoded into a binary variable of “sanitizing hands every time” (0=no, 1=yes).

### Statistical Analysis

The characteristics of all studied variables were described first. The associations between independent variables (including overall amount, sources, and content of COVID-19-specific information exposure) and mental health outcomes (depressive symptoms and sleep quality) were tested using multivariable linear regression analysis; the associations between the independent variables and behavioral outcomes (self-reported consistent face mask wearing and sanitizing hands every time) were tested using multivariate logistic regression analysis. Background characteristics that were significantly associated with the dependent variables were adjusted for in the regression models. Adjusted unstandardized coefficients (B) or adjusted odds ratios (AORs) and their 95% CIs were reported. An interaction term was further created by multiplying overall information exposure and thoughtful consideration of information veracity, and this term was included in the multivariable linear and logistic regression models to test its significance. SPSS version 24 (IBM Corporation) was used to conduct all analyses, with a two-sided *P* value <.05 indicating statistical significance.



## Results

### Background Characteristics

Over half of the 3035 participants were <30 years of age (1552, 51.1%), male (1612, 53.1%), married (1812, 59.7%), had attained an education level lower than college or university (2004, 66%), had a monthly income level lower than ¥5000 (US \$714) (1542, 50.8%), were frontline workers (1847, 60.9%), and were working in an electronic device manufacturing factory (2353, 77.5%) ([Table 1](#)).

### COVID-19–Specific Information Exposure

Of the 3035 participants, 1638 (54.0%), 1057 (34.8%), and 1013 (33.3%) had been exposed to COVID-19–specific information

via official web-based media, unofficial web-based media, and television for >1 hour per day, respectively. Smaller proportions reported a daily exposure of >1 hour via newspaper (499/3035, 16.5%) and face-to-face communication (506/3035, 16.7%). Regarding the contents of the information, over half of the 3035 participants were always exposed to statistics about the COVID-19 pandemic (1720, 56.7%), positive information about governmental responses to COVID-19 (1617, 53.3%), and heroic stories about frontline health care workers (1542, 50.8%). Only 1065 of the 3035 participants (35.1%) always thought carefully about the veracity of the COVID-19–specific information ([Table 2](#)).



**Table 1.** Background characteristics of the study sample (N=3035), n (%).

Characteristic	Value
<b>Age (years)</b>	
18-25	653 (21.5)
26-30	899 (29.6)
31-40	1195 (39.4)
>40	288 (9.5)
<b>Gender</b>	
Male	1612 (53.1)
Female	1423 (46.9)
<b>Marital status</b>	
Unmarried	1223 (40.3)
Married	1812 (59.7)
<b>Highest education level attained</b>	
Junior high or below	1163 (38.3)
Senior high or equivalent	841 (27.7)
College or university	895 (29.5)
Postgraduate	136 (4.5)
<b>Monthly personal income (¥)<sup>a</sup></b>	
<3000	179 (5.9)
3000-4999	1363 (44.9)
5000-6999	763 (25.1)
7000-9999	327 (10.8)
≥10,000	403 (13.3)
<b>Type of work</b>	
Frontline worker	1847 (60.9)
Manager	1188 (39.1)
<b>Factory type</b>	
Electronic device manufacturing	2353 (77.5)
Watchmaking	307 (10.1)
Beverage manufacturing	191 (6.3)
Biotechnology product manufacturing	184 (6.1)

<sup>a</sup>1 ¥=US \$0.14 on March 1, 2020.

**Table 2.** Exposure to COVID-19–specific information in the study sample (N=3035).

Characteristic	Value
<b>Daily average hours of exposure to COVID-19–specific information through different channels, n (%)</b>	
<b>Official web-based media</b>	
Almost never	134 (4.4)
Less than 1 hour	1263 (41.6)
1-2 hours	911 (30.0)
3-4 hours	258 (8.5)
More than 4 hours	469 (15.5)
<b>Unofficial web-based media</b>	
Almost never	543 (17.9)
Less than 1 hour	1435 (47.3)
1-2 hours	572 (18.8)
3-4 hours	186 (6.1)
>4 hours	299 (9.9)
<b>Television</b>	
Almost never	614 (20.2)
Less than 1 hour	1408 (46.4)
1-2 hours	608 (20.0)
3-4 hours	147 (4.8)
>4 hours	258 (8.5)
<b>Newspapers and magazines</b>	
Almost never	1628 (53.6)
Less than 1 hour	908 (29.9)
1-2 hours	294 (9.7)
3-4 hours	78 (2.6)
>4 hours	127 (4.2)
<b>Face-to-face communication</b>	
Almost never	1269 (41.9)
Less than 1 hour	1260 (41.5)
1-2 hours	309 (10.2)
3-4 hours	76 (2.5)
>4 hours	121 (4.0)
Overall Information Exposure Scale (5 items; sum score), mean (SD)	6.26 (3.92)
<b>Frequency of exposure to different topics of COVID-19–specific information, n (%)</b>	
<b>Statistics about the COVID-19 pandemic</b>	
Almost never	248 (8.2)
Seldom	318 (10.5)
Sometimes	749 (24.7)
Always	1720 (56.7)
<b>Positive information about governmental responses to COVID-19</b>	
Almost never	155 (5.1)
Seldom	260 (8.6)
Sometimes	1003 (33.0)

Characteristic	Value
Always	1617 (53.3)
<b>Negative information about governmental responses to COVID-19</b>	
Almost never	290 (9.6)
Seldom	445 (14.7)
Sometimes	1138 (37.5)
Always	1162 (38.3)
<b>Information about development of vaccines and effective treatment</b>	
Almost never	196 (6.5)
Seldom	404 (13.3)
Sometimes	1156 (38.1)
Always	1279 (42.1)
<b>Heroic stories about frontline health care workers</b>	
Almost never	145 (4.8)
Seldom	227 (7.5)
Sometimes	1121 (36.9)
Always	1542 (50.8)
<b>Positive information about patients with COVID-19</b>	
Almost never	155 (5.1)
Seldom	297 (9.8)
Sometimes	1174 (38.7)
Always	1409 (46.4)
<b>Negative information about patients with COVID-19</b>	
Almost never	357 (11.8)
Seldom	594 (19.6)
Sometimes	1190 (39.2)
Always	894 (29.5)
<b>Information about negative impacts of the COVID-19 pandemic on the economy</b>	
Almost never	515 (17.0)
Seldom	651 (21.4)
Sometimes	1097 (36.1)
Always	772 (25.4)
<b>Thoughtful consideration of the veracity of COVID-19-specific information, n (%)</b>	
Almost never	207 (6.8)
Seldom	346 (11.4)
Sometimes	1417 (46.7)
Always	1065 (35.1)

### Mental Health and Self-Reported Compliance With Personal Preventive Measures

The prevalence of probable moderate to severe depression was 170/3035 (5.6%), and that of good or excellent sleep quality

was 2110/3035 (69.5%) (Table 3). Regarding behavioral responses, 2903/3035 (95.7%) and 2151/3035 (70.9%) participants reported consistent face mask wearing in any public places and sanitizing hands every time after returning from public spaces or touching installations, respectively.

**Table 3.** Mental health and behavioral outcomes in the study sample (N=3035).

Characteristic	Value
<b>Mental health outcomes</b>	
Depressive symptoms (mean score of the PHQ-9 <sup>a</sup> ), mean (SD)	2.12 (4.02)
<b>Sleep quality</b>	
Poor or terrible, n (%)	112 (3.7)
Fair, n (%)	813 (26.8)
Good, n (%)	1384 (45.6)
Excellent, n (%)	726 (23.9)
Mean score (SD)	2.90 (0.80)
<b>Self-reported compliance with personal preventive measures, n (%)</b>	
<b>Self-reported consistent face mask wearing (wearing a face mask every time in the workplace and in other public spaces or transportation)</b>	
No	132 (4.3)
Yes	2903 (95.7)
<b>Self-reported hand sanitizing every time after returning from public spaces or touching installations</b>	
No	884 (29.1)
Yes	2151 (70.9)

<sup>a</sup>PHQ-9: 9-item Patient Health Questionnaire.

### Associations Between Information Exposure and Mental Health Outcomes

In the univariate analysis, age, marital status, education level, monthly personal income, status as frontline worker or manager, and factory type were significantly associated with depressive symptoms and sleep quality ([Multimedia Appendix 1](#)).

After adjusting for these significant background characteristics, a higher overall amount of COVID-19-specific information exposure was associated with higher depressive symptoms (adjusted B=0.05,  $P=.006$ ). Specifically, higher exposure via unofficial web-based media (adjusted B=0.20,  $P=.001$ ) and face-to-face communication (adjusted B=0.45,  $P<.001$ ) was associated with higher depressive symptoms. Exposure to negative information about patients with COVID-19 (adjusted B=0.15,  $P=.049$ ) and about effects on the economy (adjusted B=0.31,  $P<.001$ ) was associated with higher depressive

symptoms, while exposure to information about development of vaccines and effective treatments was associated with lower depressive symptoms (adjusted B =-0.16,  $P=.045$ ) ([Table 4](#)).

A higher overall amount of COVID-19-specific information exposure was associated with better sleep quality (adjusted B=0.01,  $P=.01$ ). Specifically, exposure via face-to-face communication was associated with poorer sleep quality (adjusted B=-0.04,  $P=.01$ ), while exposure via television (adjusted B 0.05,  $P<.001$ ) and newspapers and magazines (adjusted B=0.07,  $P<.001$ ) was associated with better sleep quality. Exposure to negative information about patients with COVID-19 (adjusted B=-0.03,  $P=.04$ ) and negative impacts on the economy (adjusted B=-0.04,  $P=.02$ ) was associated with poorer sleep quality. In contrast, exposure to heroic stories about frontline health care workers (adjusted B=0.06,  $P=.002$ ) and positive information about patients with COVID-19 (adjusted B=0.04,  $P=.02$ ) was associated with better sleep quality.

**Table 4.** Linear regression on exposure to COVID-19–specific information and mental health outcomes (N=3035).

Exposure	Depressive symptoms		Sleep quality	
	Adjusted B (95% CI)	P value	Adjusted B (95% CI)	P value
<b>Frequency of exposure to COVID-19–specific information through different channels</b>				
Official web-based media	0.10 (–0.02 to 0.23)	.11	0.01 (–0.01 to 0.04)	.30
Unofficial web-based media	0.20 (0.08 to 0.33)	.001	0.02 (–0.01 to 0.04)	.20
Television	–0.02 (–0.15 to 0.11)	.76	0.05 (0.03 to 0.08)	<.001
Newspapers and magazines	–0.01 (–0.15 to 0.13)	.87	0.07 (0.04 to 0.10)	<.001
Face-to-face communication	0.45 (0.30 to 0.60)	<.001	–0.04 (–0.07 to –0.01)	.01
Overall information exposure	0.05 (0.02 to 0.09)	.006	0.01 (0.00 to 0.02)	.01
<b>Frequency of exposure to different contents of COVID-19–specific information</b>				
Statistics about the COVID-19 pandemic	0.02 (–0.13 to 0.18)	.76	–0.02 (–0.05 to 0.01)	.15
Positive information about governmental responses to COVID-19	–0.13 (–0.30 to 0.04)	.12	0.02 (–0.01 to 0.06)	.15
Negative information about governmental responses to COVID-19	0.01 (–0.14 to 0.16)	.86	0.00 (–0.03 to 0.03)	.89
Information about development of vaccines and effective treatment	–0.16 (–0.32 to –0.00)	.045	0.03 (–0.00 to 0.06)	.08
Heroic stories about frontline health care workers	–0.13 (–0.30 to 0.05)	.16	0.06 (0.02 to 0.09)	.002
Positive information about patients with COVID-19	–0.11 (–0.28 to 0.06)	.21	0.04 (0.01 to 0.08)	.02
Negative information about patients with COVID-19	0.15 (0.00 to 0.29)	0.049	–0.03 (–0.06 to –0.00)	.04
Information about negative impacts of the COVID-19 pandemic on the economy	0.32 (0.18 to 0.46)	<.001	–0.04 (–0.06 to –0.01)	.02

### Associations Between Information Exposure and Behavioral Outcomes

In the univariate analysis, age, gender, marital status, education level, and factory type were associated with face mask wearing and sanitizing hands every time ([Multimedia Appendix 1](#)).

After adjusting for these significant background characteristics, the overall or source-specific amount of COVID-19–specific information exposure was not significantly associated with consistent face mask wearing. Exposure to statistics about the COVID-19 pandemic (AOR 1.23, 95% CI 1.04-1.45;  $P=.02$ ), negative information about governmental responses (AOR 1.32, 95% CI 1.11-1.56;  $P=.001$ ), heroic stories about frontline health care workers (AOR 1.30, 95% CI 1.08-1.56;  $P=.007$ ), and

positive information about patients with COVID-19 (AOR 1.33, 95% CI 1.11-1.60;  $P=.002$ ) were positively associated with this outcome.

Overall amount of COVID-19–specific information exposure (AOR=1.03, 95% CI 1.01-1.05;  $P=.003$ ) and specific exposure through official web-based media, unofficial web-based media, television, and newspapers and magazines, with AORs ranging from 1.07 (95% CI 1.00-1.15) to 1.21 (95% CI 1.11-1.32), were positively associated with sanitizing hands every time, while a negative association was found for exposure through face-to-face communication (AOR=0.91, 95% CI 0.84-0.98;  $P=.02$ ). Exposure to all eight information topics was positively associated with this outcome, with AORs ranging from 1.09 (95% CI 1.01-1.18) to 1.36 (95% CI 1.24-1.49) ([Table 5](#)).



**Table 5.** Logistic regression of media exposure to COVID-19–specific information and behavioral outcomes (N=3035).

Media exposure	Consistent face mask wearing		Sanitizing hands every time	
	Adjusted odds ratio (95% CI)	P value	Adjusted odds ratio (95% CI)	P value
<b>Frequency of exposure to COVID-19–specific information through different channels</b>				
Official web-based media	1.01 (0.87-1.18)	.90	1.07 (1.00-1.15)	.05
Unofficial web-based media	1.03 (0.88-1.20)	.74	1.08 (1.00-1.16)	.04
Television	1.11 (0.94-1.30)	.22	1.18 (1.09-1.27)	<.001
Newspapers and magazines	1.03 (0.87-1.23)	.70	1.21 (1.11-1.32)	<.001
Face-to-face communication	1.13 (0.92-1.37)	.24	0.91 (0.84-0.98)	.02
Overall information exposure	1.02 (0.97-1.07)	.39	1.03 (1.01-1.05)	.003
<b>Frequency of exposure to different COVID-19–specific information</b>				
Statistics about the COVID-19 pandemic	1.23 (1.04-1.45)	.02	1.23 (1.13-1.33)	<.001
Positive information about governmental responses to COVID-19	1.16 (0.96-1.40)	.12	1.34 (1.22-1.47)	<.001
Negative information about governmental responses to COVID-19	1.32 (1.11-1.56)	.001	1.20 (1.10-1.30)	<.001
Information about development of vaccines and effective treatment	1.16 (0.97-1.39)	.11	1.36 (1.25-1.49)	<.001
Heroic stories about frontline health care workers	1.30 (1.08-1.56)	.007	1.32 (1.20-1.46)	<.001
Positive information about patients with COVID-19	1.33 (1.11-1.60)	.002	1.36 (1.24-1.49)	<.001
Negative information about patients with COVID-19	1.17 (0.98-1.38)	.08	1.14 (1.05-1.23)	.002
Information about negative impacts of the COVID-19 pandemic on the economy	1.13 (0.95-1.35)	.16	1.09 (1.01-1.18)	.03

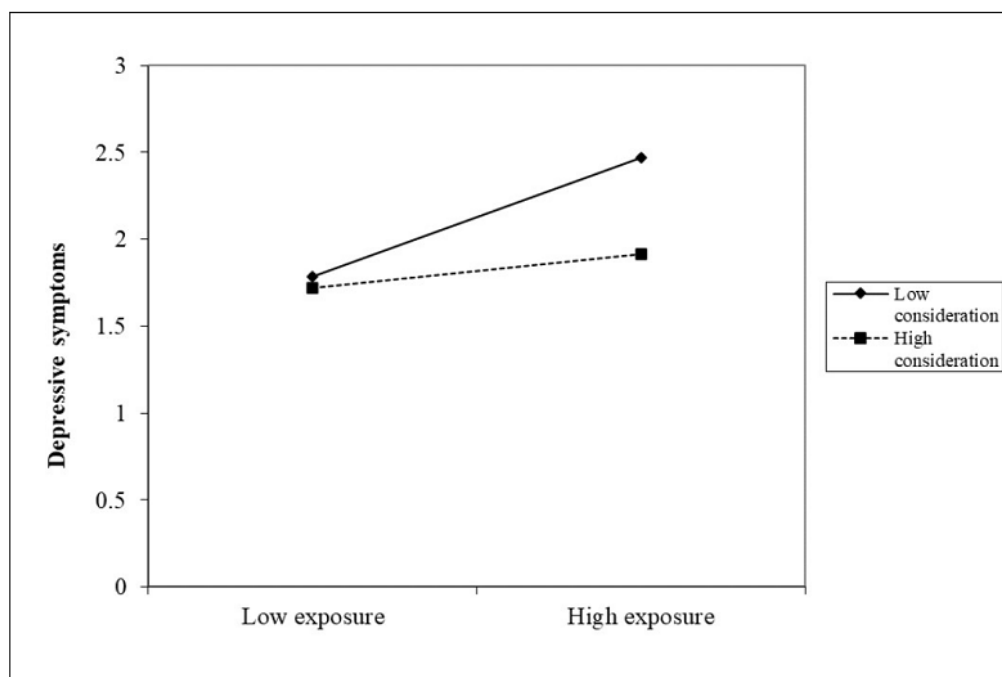
### Moderation Effects of Thoughtful Consideration of Information Veracity

The interaction term was significantly associated with depressive symptoms (adjusted  $B=-0.04$ ,  $P=.047$ ) and sleep quality (adjusted  $B=0.01$ ,  $P<.001$ ). Overall amount of information exposure was associated with higher depressive symptoms among participants who were less likely to consider the veracity of the information to which they were exposed; this association was not significant among those who were more likely to

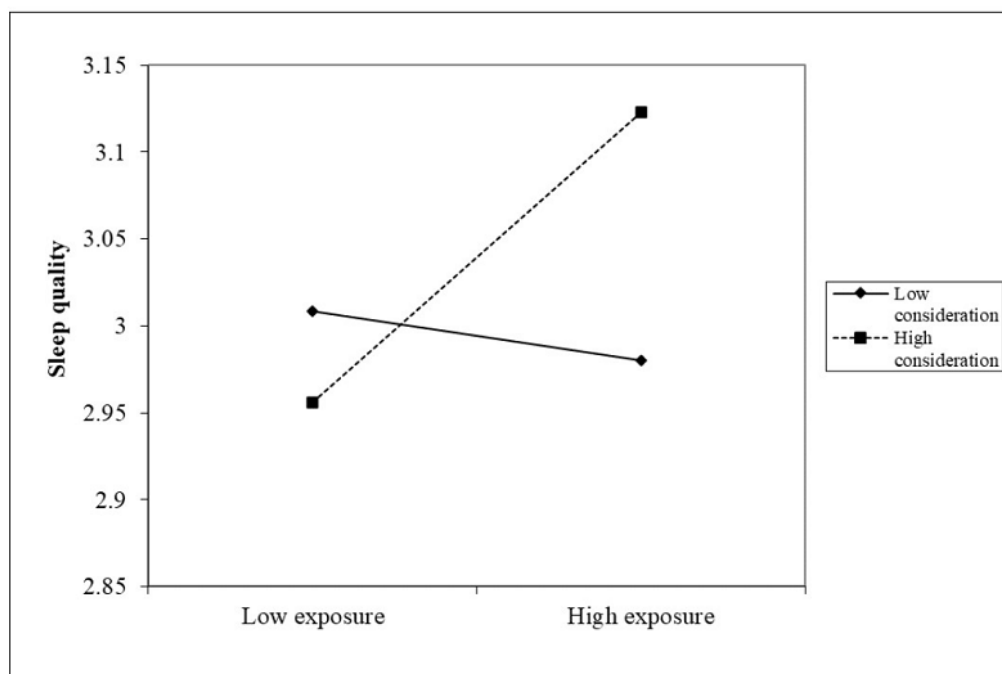
consider the veracity of this information. In contrast, overall amount of information exposure was associated with better sleep quality among participants who were more likely to consider the veracity of the information to which they were exposed; this association was not significant among those who were less likely to consider the veracity of this information (Figure 2 and Multimedia Appendix 2). Thoughtful consideration of information veracity did not moderate the association between overall amount of information exposure and behavioral responses.

**Figure 2.** The moderation effects of thoughtful consideration about the veracity of COVID-19–specific information on associations between overall information exposure and the mental health outcomes of (1) depressive symptoms and (2) sleep quality with adjustment for significant background factors.

(1) Outcome: depressive symptoms



(2) Outcome: sleep quality



## Discussion

### Principal Findings

The COVID-19 pandemic may have increased mental health problems among Chinese factory workers after work resumption, as their prevalence of moderate to severe depression was higher than that observed in the general population before the pandemic

(5.6% versus 2.4%) [38]. The prevalence of terrible or poor sleep quality was lower than that of Chinese workers who returned to work in Jiangsu Province (14.9%) [39]. Regarding self-reported compliance with personal preventive measures, the prevalence of consistent face mask wearing in all public spaces was very high, which is crucial for COVID-19 prevention in factories where physical distancing cannot be guaranteed.

Despite the WHO recommendation on hand hygiene [7], few participants always sanitized their hands. The importance of hand hygiene must be emphasized during work resumption.

Our findings suggest that the COVID-19 pandemic triggered intensive media responses, as our participants had been exposed to large amounts of COVID-19–specific information every day following the outbreak in China. Web-based media and television were the most common information sources, while newspapers and magazines and face-to-face communication were less common. These findings are consistent with previous reports, which showed a substantial increase in the use of web-based media and television following the COVID-19 outbreak [40].

Consistent with previous studies [4,24], we found that different sources of information exposure had differing effects on mental health and behavioral outcomes. First, television, newspapers, and magazines are traditional media outlets that are mainly operated by governmental organizations in China. Exposure to COVID-19–specific information through these channels not only increased compliance with personal preventive measures but also improved mental health and well-being. These media channels mainly report information verified by expert sources [4,24]. Previous studies suggested that the more people read newspapers and watched television reports about MERS, the more knowledge they acquired about the disease and its prevention strategies [4,24]. Being knowledgeable about COVID-19 was associated with better compliance with personal preventive measures [21] and lower levels of worry or panic [41]. In addition to knowledge dissemination, these traditional media outlets in China often disseminate information about the government's preparedness plans, actions, and progress in controlling COVID-19. This information is helpful to build trust in the government's capacity to counter the pandemic. Trust in government was positively associated with adoption of personal preventive measures and mental health during an infectious disease pandemic [42,43].

It was interesting to find that information exposure through official web-based media was associated with better hand hygiene but did not affect mental health status. In China, these media outlets are operated by governmental organizations and disseminate similar information to traditional media. However, instead of one-way communication, official web-based media platforms are interactive and allow readers to leave comments. We speculate that people viewed both positive and negative feedback for specific topics, which may have offset the positive influence on mental health. Information exposure through unofficial web-based media sources was likely to have both positive and negative aspects, as it motivated personal preventive measures but triggered mental health problems at the same time. These findings are similar to studies conducted during the 2015 MERS outbreak in South Korea [4,24]. Compared to official media sources (official web-based media and traditional media in China), unofficial web-based media sources contained not only factual information but also emotional content [22,44]. Negative emotions and unnecessary sensationalism are more likely to be present and amplified in these media outlets during an infectious disease outbreak than in official media channels.

Information exposure through face-to-face communication was limited, probably due to strict control measures implemented by the Chinese government to curb the COVID-19 pandemic (eg, intracity travel restrictions, community lockdown). However, the influence of face-to-face communication cannot be neglected. Higher levels of information exposure through this source were associated with decreased levels of hand hygiene and poorer mental health. It is likely that false and unverified information was disseminated by face-to-face communication during the pandemic. Studies have shown that “bottom-up” misinformation accounted for a significant proportion of information sharing between laypersons during the COVID-19 pandemic [45]. The consequences of misinformation can be long-lasting and should not be underestimated in health crisis management [29].

Our study covered eight types of COVID-19–related content that were active topics attracting public attention from the Chinese population during the pandemic [44,46]. Exposure to all these types of content was associated with increased compliance with face mask wearing and/or hand hygiene. A sufficient amount of information exposure about COVID-19, regardless of topic and valence (positive, negative, or neutral), could potentially cultivate a global sense of emergency and enhance compliance with these preventive measures. Exposure to content about positive outcomes for patients with COVID-19, development of effective treatments and vaccines, and heroic stories about frontline health care workers were associated with better mental health. This is understandable, as these types of information can increase people's confidence in the control of the pandemic and hence reduce their concern. Moreover, heroic stories may increase altruism, which has a protective effect on mental health during a pandemic. In contrast, negative information about patients with COVID-19 and about impacts on the economy were associated with decreased mental health. Observing the detrimental consequences of COVID-19 may induce fear, according to fear appeal theory [47]. Pay cuts or layoffs resulting from COVID-19 are potential stressors that may lead to worse mental health [48].

The measure of thoughtful consideration inherently assessed the participants' motivation to scrutinize information quality but also the ability to acquire, discern, and understand accurate health information [49–51]. During the global epidemic of misinformation that has accompanied the COVID-19 pandemic, some individuals were able to actively protect themselves against the crisis. Our results highlighted that thoughtful consideration of the veracity of information to which the participants were exposed significantly moderated the association between overall amount of information exposure and mental health. Participants who thought more about information veracity suffered less from the adverse effect of high information exposure on depression and even had better sleep quality compared to people with low information exposure. People with inadequate skills and resources to analyze and appraise the information to which they are exposed may be particularly vulnerable to mental distress due to high information anxiety, confusion about information quality, and unsatisfied information needs [12,52,53]. There is a significant need for

intervention, as only 35% of the participants always thought carefully about the veracity of COVID-19-specific information.

### Implications

Our findings may inform strategic risk communication by government and public health authorities during the COVID-19 outbreak and future public health emergencies to build public trust and facilitate prevention without amplifying panic. Future risk communication related to COVID-19 in China should make use of official web-based media and traditional media platforms; increase dissemination of information about patient rehabilitation, development of effective treatments and vaccines, and heroic deeds of frontline health care workers; and inform the public about effective strategies to reactivate the economy. Pilot testing of communication campaign messages among the target audience is also recommended. Furthermore, it is imperative to invest in building communication capacity to combat the crisis of misinformation. Government and public health authorities should proactively identify misinformation, provide clarifications, and empower the public with adequate skills to critically evaluate the veracity of information.

### Limitations

The present study has some limitations. First, causal relationships between information exposure and the studied outcomes cannot be determined due to the cross-sectional design. Reverse relationships are possible; for example, individuals with mental distress may spend more time searching for COVID-19-related information [16]. Second, the survey did not capture the temporal pattern of information exposure and its correlations with health outcomes [46]; instead, global

indicators were used to measure the average exposure levels over a critical period from the initial outbreak to the containment phase. Third, the measures did not differentiate active and passive information gathering; however, the majority of the Chinese population proactively sought related information during the COVID-19 pandemic [21,40]. Fourth, generalization of the evidence from a workforce sample should be made cautiously. The mental health impact of information exposure may be different for people who are unemployed or retired. This study was conducted in one city; however, the participants had resided in 29 of the 34 provincial regions in China prior to work resumption, and the observed findings are thus not only limited to the context of Shenzhen. Moreover, the data were self-reported and verification was not feasible. Participants may also overreport their compliance with personal preventive measures due to social desirability. The web-based survey was anonymous and did not collect participants' personal details. These measures may have reduced the degree of social desirability bias.

### Conclusion

This study provides empirical evidence of how different dimensions of COVID-19-specific information exposure influenced mental health status and compliance with personal preventive measures at the initial phase of work resumption in China. Thoughtful consideration of the veracity of information played an important role in moderating the associations between information exposure and mental health. These findings can help improve crisis communication during the response to the ongoing pandemic.

### Acknowledgments

This study was funded by the Startup Fund for the 100 Top Talents Program, Sun Yat-sen University (grant number 392012), and the National Key Research and Development Program (grant number 2018YFA0902801).

### Authors' Contributions

YP, MX, and CZ contributed equally as first authors. ZW, JY, and YH contributed equally as corresponding authors.

### Conflicts of Interest

None declared.

#### Multimedia Appendix 1

Associations between background factors and behavioral/mental health outcomes (N=3035).

[DOCX File, 25 KB - [jmir\\_v22i10e22596\\_app1.docx](#)]

#### Multimedia Appendix 2

Regression coefficients for moderation analyses.

[DOCX File, 13 KB - [jmir\\_v22i10e22596\\_app2.docx](#)]

### References

1. Risk communication. World Health Organization. 2017. URL: <https://www.who.int/emergencies/risk-communications> [accessed 2020-09-28]
2. Reynolds B, W Seeger M. Crisis and emergency risk communication as an integrative model. J Health Commun 2005;10(1):43-55. [doi: [10.1080/10810730590904571](https://doi.org/10.1080/10810730590904571)] [Medline: [15764443](#)]

3. Zhang L, Kong Y, Chang H. Media Use and Health Behavior in H1N1 Flu Crisis: The Mediating Role of Perceived Knowledge and Fear. *Atl J Commun* 2015 Apr 30;23(2):67-80. [doi: [10.1080/15456870.2015.1013101](https://doi.org/10.1080/15456870.2015.1013101)]
4. Seo M. Amplifying Panic and Facilitating Prevention: Multifaceted Effects of Traditional and Social Media Use During the 2015 MERS Crisis in South Korea. *Journal Mass Commun Q* 2019 Jul 26;107769901985769. [doi: [10.1177/1077699019857693](https://doi.org/10.1177/1077699019857693)]
5. Posid JM, Bruce SM, Guarnizo JT, Taylor ML, Garza BW. SARS: mobilizing and maintaining a public health emergency response. *J Public Health Manag Pract* 2005;11(3):208-215. [doi: [10.1097/00124784-200505000-00005](https://doi.org/10.1097/00124784-200505000-00005)] [Medline: [15829833](https://pubmed.ncbi.nlm.nih.gov/15829833/)]
6. Esposito S, Principi N, Leung CC, Migliori GB. Universal use of face masks for success against COVID-19: evidence and implications for prevention policies. *Eur Respir J* 2020 Jun;55(6):2001260. [doi: [10.1183/13993003.01260-2020](https://doi.org/10.1183/13993003.01260-2020)] [Medline: [32350103](https://pubmed.ncbi.nlm.nih.gov/32350103/)]
7. Recommendations to Member States to improve hand hygiene practices to help prevent the transmission of the COVID-19 virus. World Health Organization. 2020 Apr 01. URL: <https://www.who.int/publications-detail/recommendations-to-member-states-to-improve-hand-hygiene-practices-to-help-prevent-the-transmission-of-the-covid-19-virus> [accessed 2020-09-28]
8. Wipatayotin A. 70% Thais stick to social distancing. *Bangkok Post*. 2020 Apr 02. URL: <https://www.bangkokpost.com/thailand/general/1891290/70-thais-stick-to-social-distancing> [accessed 2020-09-28]
9. Briscese G, Lacetera N, Macis M, Tonin M. Compliance with COVID-19 social-distancing measures in Italy: the role of expectation and duration. *NBER Working Papers*. 2020 Mar. URL: <https://www.nber.org/papers/w26916> [accessed 2020-09-28]
10. Gammage KL, Klentrou P. Predicting osteoporosis prevention behaviors: health beliefs and knowledge. *Am J Health Behav* 2011 May 01;35(3):371-382. [doi: [10.5993/ajhb.35.3.10](https://doi.org/10.5993/ajhb.35.3.10)] [Medline: [21683025](https://pubmed.ncbi.nlm.nih.gov/21683025/)]
11. Lee D, Kim J, Kang H. The Emotional Distress and Fear of Contagion Related to Middle East Respiratory Syndrome (MERS) on General Public in Korea. Article in Korean. *Kor J Psychol: Gen* 2016 Jun 30;35(2):355. [doi: [10.22257/kjp.2016.06.35.2.355](https://doi.org/10.22257/kjp.2016.06.35.2.355)]
12. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. *Int J Environ Res Public Health* 2020 Mar 06;17(5):1729 [FREE Full text] [doi: [10.3390/ijerph17051729](https://doi.org/10.3390/ijerph17051729)] [Medline: [32155789](https://pubmed.ncbi.nlm.nih.gov/32155789/)]
13. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. *Gen Psych* 2020 Mar 06;33(2):e100213. [doi: [10.1136/gpsych-2020-100213](https://doi.org/10.1136/gpsych-2020-100213)]
14. Qian M, Wu Q, Wu P, Hou Z, Liang Y, Cowling B. Psychological responses, behavioral changes and public perceptions during the early phase of the COVID-19 outbreak in China: a population based cross-sectional survey. *medRxiv*. Preprint posted online on February 20, 2020. [doi: [10.1101/2020.02.18.20024448](https://doi.org/10.1101/2020.02.18.20024448)]
15. Renn O, Burns W, Kasperson R, Kasperson J, Slovic P. The social amplification of risk: Theoretical foundations and empirical application. *J Soc Issues* 1992;48(4):137-160. [doi: [10.1111/j.1540-4560.1992.tb01949.x](https://doi.org/10.1111/j.1540-4560.1992.tb01949.x)]
16. Garfin D, Silver R, Holman E. The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. *Health Psychol* 2020 May;39(5):355-357. [doi: [10.1037/hea0000875](https://doi.org/10.1037/hea0000875)] [Medline: [32202824](https://pubmed.ncbi.nlm.nih.gov/32202824/)]
17. Ho CSH, Chee CYI, Ho RCM. Mental Health Strategies to Combat the Psychological Impact of COVID-19 Beyond Paranoia and Panic. *Ann Acad Med Singap* 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: [32200399](https://pubmed.ncbi.nlm.nih.gov/32200399/)]
18. Yuen KF, Wang X, Ma F, Li KX. The Psychological Causes of Panic Buying Following a Health Crisis. *Int J Environ Res Public Health* 2020 May 18;17(10):3513 [FREE Full text] [doi: [10.3390/ijerph17103513](https://doi.org/10.3390/ijerph17103513)] [Medline: [32443427](https://pubmed.ncbi.nlm.nih.gov/32443427/)]
19. Tai Z, Sun T. Media dependencies in a changing media environment: the case of the 2003 SARS epidemic in China. *New Media Soc* 2016 Jun 30;9(6):987-1009. [doi: [10.1177/1461444807082691](https://doi.org/10.1177/1461444807082691)]
20. Kim Y. An Essay on Korean Media's Coverage of Middle East Respiratory Syndrome Coronavirus. Article in Korean. *Korean J Health Commun* 2016 Jun;11(1):39-50. [doi: [10.15715/kjhcom.2016.11.1.39](https://doi.org/10.15715/kjhcom.2016.11.1.39)]
21. Kwok KO, Li KK, Chan HHH, Yi YY, Tang A, Wei WI, et al. Community Responses during Early Phase of COVID-19 Epidemic, Hong Kong. *Emerg Infect Dis* 2020 Jul;26(7):1575-1579 [FREE Full text] [doi: [10.3201/eid2607.200500](https://doi.org/10.3201/eid2607.200500)] [Medline: [32298227](https://pubmed.ncbi.nlm.nih.gov/32298227/)]
22. Choi D, Yoo W, Noh G, Park K. The impact of social media on risk perceptions during the MERS outbreak in South Korea. *Comput Human Behav* 2017 Jul;72:422-431 [FREE Full text] [doi: [10.1016/j.chb.2017.03.004](https://doi.org/10.1016/j.chb.2017.03.004)] [Medline: [32288176](https://pubmed.ncbi.nlm.nih.gov/32288176/)]
23. Ahmad AR, Murad HR. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. *J Med Internet Res* 2020 May 19;22(5):e19556 [FREE Full text] [doi: [10.2196/19556](https://doi.org/10.2196/19556)] [Medline: [32369026](https://pubmed.ncbi.nlm.nih.gov/32369026/)]
24. Ludolph R, Schulz P, Chen L. Investigating the Effects of Mass Media Exposure on the Uptake of Preventive Measures by Hong Kong Residents during the 2015 MERS Outbreak: The Mediating Role of Interpersonal Communication and the Perception of Concern. *J Health Commun* 2018;23(1):1-8. [doi: [10.1080/10810730.2017.1388455](https://doi.org/10.1080/10810730.2017.1388455)] [Medline: [29261430](https://pubmed.ncbi.nlm.nih.gov/29261430/)]
25. Kreps GL, Alibek K, Neuhauser LE, Rowan K, Sparks L. Emergency/risk communication to promote public health respond to biological threats. In: Harden M, editor. *Global Public Health Communication*. Sudbury, MA: Jones and Bulwer; 2005:349-362.



26. Covello V. Best practices in public health risk and crisis communication. *J Health Commun* 2003;8 Suppl 1:5-8; discussion 148. [doi: [10.1080/713851971](https://doi.org/10.1080/713851971)] [Medline: [14692565](https://pubmed.ncbi.nlm.nih.gov/14692565/)]
27. Zarocostas J. How to fight an infodemic. *Lancet* 2020 Feb;395(10225):676. [doi: [10.1016/s0140-6736\(20\)30461-x](https://doi.org/10.1016/s0140-6736(20)30461-x)]
28. Bao H, Cao B, Xiong Y, Tang W. Digital Media's Role in the COVID-19 Pandemic. *JMIR Mhealth Uhealth* 2020 Sep 18;8(9):e20156 [FREE Full text] [doi: [10.2196/20156](https://doi.org/10.2196/20156)] [Medline: [32530817](https://pubmed.ncbi.nlm.nih.gov/32530817/)]
29. Countering misinformation about COVID-19. World Health Organization. 2020 May 11. URL: <https://www.who.int/news-room/feature-stories/detail/countering-misinformation-about-covid-19> [accessed 2020-09-28]
30. Clarifications. The Government of the Hong Kong Special Administration Region. URL: <https://www.coronavirus.gov.hk/eng/clarifications.html> [accessed 2020-09-28]
31. 2020 first quarter GDP. Webpage in Chinese. National Bureau of Statistics. 2020. URL: [http://www.stats.gov.cn/tjsj/zxfb/202004/t20200417\\_1739602.html](http://www.stats.gov.cn/tjsj/zxfb/202004/t20200417_1739602.html) [accessed 2020-09-28]
32. A Letter to Enterprises in Shenzhen. Office of Shenzhen Headquarters for Prevention and Control of Pneumonia Epidemic Caused by Novel Coronavirus. 2020 Feb 02. URL: [http://fao.sz.gov.cn/ydmh/xxgk/zyxw/202002/t20200205\\_18998951.htm](http://fao.sz.gov.cn/ydmh/xxgk/zyxw/202002/t20200205_18998951.htm) [accessed 2020-09-28]
33. Zhang L, Shen M, Ma X, Su S, Gong W, Wang J. What is required to prevent a second major outbreak of the novel coronavirus SARS-CoV-2 upon lifting the metropolitan-wide quarantine of Wuhan city, China. medRxiv. Preprint posted online on March 30, 2020. [doi: [10.1101/2020.03.24.20042374](https://doi.org/10.1101/2020.03.24.20042374)]
34. Shenzhen Statistics Yearbook (2019). Statistics Bureau of Shenzhen Municipality. 2019. URL: <http://tjj.sz.gov.cn/attachment/0/695/695422/7971762.pdf> [accessed 2020-09-28]
35. Liao W, Lau JTF, Tsui HY, Gu J, Wang Z. Relationship between sexual compulsivity and sexual risk behaviors among Chinese sexually active males. *Arch Sex Behav* 2015 Apr 17;44(3):791-798. [doi: [10.1007/s10508-014-0317-z](https://doi.org/10.1007/s10508-014-0317-z)] [Medline: [25030121](https://pubmed.ncbi.nlm.nih.gov/25030121/)]
36. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Gen Hosp Psychiatry* 2015;37(6):567-576. [doi: [10.1016/j.genhosppsych.2015.06.012](https://doi.org/10.1016/j.genhosppsych.2015.06.012)] [Medline: [26195347](https://pubmed.ncbi.nlm.nih.gov/26195347/)]
37. Snyder E, Cai B, DeMuro C, Morrison MF, Ball W. A New Single-Item Sleep Quality Scale: Results of Psychometric Evaluation in Patients With Chronic Primary Insomnia and Depression. *J Clin Sleep Med* 2018 Nov 15;14(11):1849-1857 [FREE Full text] [doi: [10.5664/jcsm.7478](https://doi.org/10.5664/jcsm.7478)] [Medline: [30373688](https://pubmed.ncbi.nlm.nih.gov/30373688/)]
38. Chen Y, Bennett D, Clarke R, Guo Y, Yu C, Bian Z, et al. Patterns and correlates of major depression in Chinese adults: a cross-sectional study of 0.5 million men and women. *Psychol Med* 2016 Dec 06;47(5):958-970. [doi: [10.1017/s0033291716002889](https://doi.org/10.1017/s0033291716002889)]
39. Yang Y, Zhu J, Yang S, Lin H, Chen Y, Zhao Q, et al. Prevalence and associated factors of poor sleep quality among Chinese returning workers during the COVID-19 pandemic. *Sleep Med* 2020 Sep;73:47-52 [FREE Full text] [doi: [10.1016/j.sleep.2020.06.034](https://doi.org/10.1016/j.sleep.2020.06.034)] [Medline: [32771927](https://pubmed.ncbi.nlm.nih.gov/32771927/)]
40. Hua J, Shaw R. Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China. *Int J Environ Res Public Health* 2020 Mar 30;17(7):2309 [FREE Full text] [doi: [10.3390/ijerph17072309](https://doi.org/10.3390/ijerph17072309)] [Medline: [32235433](https://pubmed.ncbi.nlm.nih.gov/32235433/)]
41. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. *Int J Biol Sci* 2020;16(10):1745-1752 [FREE Full text] [doi: [10.7150/ijbs.45221](https://doi.org/10.7150/ijbs.45221)] [Medline: [32226294](https://pubmed.ncbi.nlm.nih.gov/32226294/)]
42. Sibley CG, Greaves LM, Satherley N, Wilson MS, Overall NC, Lee CHJ, et al. Effects of the COVID-19 pandemic and nationwide lockdown on trust, attitudes toward government, and well-being. *Am Psychol* 2020;75(5):618-630. [doi: [10.1037/amp0000662](https://doi.org/10.1037/amp0000662)] [Medline: [32496074](https://pubmed.ncbi.nlm.nih.gov/32496074/)]
43. Chuang Y, Huang Y, Tseng K, Yen C, Yang L. Social capital and health-protective behavior intentions in an influenza pandemic. *PLoS One* 2015;10(4):e0122970 [FREE Full text] [doi: [10.1371/journal.pone.0122970](https://doi.org/10.1371/journal.pone.0122970)] [Medline: [25874625](https://pubmed.ncbi.nlm.nih.gov/25874625/)]
44. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public Engagement and Government Responsiveness in the Communications About COVID-19 During the Early Epidemic Stage in China: Infodemiology Study on Social Media Data. *J Med Internet Res* 2020 May 26;22(5):e18796 [FREE Full text] [doi: [10.2196/18796](https://doi.org/10.2196/18796)] [Medline: [32412414](https://pubmed.ncbi.nlm.nih.gov/32412414/)]
45. Brennan J, Simon F, Howard P, Nielsen RK. Types, sources, and claims of COVID-19 misinformation. Reuters Institute. 2020 Apr 07. URL: <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation> [accessed 2020-09-28]
46. Zhao Y, Cheng S, Yu X, Xu H. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. *J Med Internet Res* 2020 May 04;22(5):e18825 [FREE Full text] [doi: [10.2196/18825](https://doi.org/10.2196/18825)] [Medline: [32314976](https://pubmed.ncbi.nlm.nih.gov/32314976/)]
47. Witte K. Putting the fear back into fear appeals: The extended parallel process model. *Commun Monogr* 1992 Dec;59(4):329-349. [doi: [10.1080/03637759209376276](https://doi.org/10.1080/03637759209376276)]
48. Sanchez TH, Zlotorzynska M, Rai M, Baral SD. Characterizing the Impact of COVID-19 on Men Who Have Sex with Men Across the United States in April, 2020. *AIDS Behav* 2020 Jul;24(7):2024-2032 [FREE Full text] [doi: [10.1007/s10461-020-02894-2](https://doi.org/10.1007/s10461-020-02894-2)] [Medline: [32350773](https://pubmed.ncbi.nlm.nih.gov/32350773/)]

49. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. *Lancet Public Health* 2020 May;5(5):e249-e250. [doi: [10.1016/s2468-2667\(20\)30086-4](https://doi.org/10.1016/s2468-2667(20)30086-4)]
50. Seng J, Yeam C, Huang W, Tan N, Low L. Pandemic related Health literacy - A Systematic Review of literature in COVID-19, SARS and MERS pandemics. *medRxiv*. Preprint posted online on May 11, 2020 . [doi: [10.1101/2020.05.07.20094227](https://doi.org/10.1101/2020.05.07.20094227)]
51. Sentell T, Vamos S, Okan O. Interdisciplinary Perspectives on Health Literacy Research Around the World: More Important Than Ever in a Time of COVID-19. *Int J Environ Res Public Health* 2020 Apr 26;17(9):3010 [FREE Full text] [doi: [10.3390/ijerph17093010](https://doi.org/10.3390/ijerph17093010)] [Medline: [32357457](https://pubmed.ncbi.nlm.nih.gov/32357457/)]
52. Qian M, Wu Q, Wu P, Hou Z, Liang Y, Cowling B. Psychological responses, behavioral changes and public perceptions during the early phase of the COVID-19 outbreak in China: a population based cross-sectional survey. *medRxiv*. Preprint posted online on February 20, 2020 . [doi: [10.1101/2020.02.18.20024448](https://doi.org/10.1101/2020.02.18.20024448)]
53. Chong YY, Cheng HY, Chan HYL, Chien WT, Wong SYS. COVID-19 pandemic, infodemic and the role of eHealth literacy. *Int J Nurs Stud* 2020 Aug;108:103644 [FREE Full text] [doi: [10.1016/j.ijnurstu.2020.103644](https://doi.org/10.1016/j.ijnurstu.2020.103644)] [Medline: [32447127](https://pubmed.ncbi.nlm.nih.gov/32447127/)]

## Abbreviations

**AOR:** adjusted odds ratio

**MERS:** Middle East respiratory syndrome

**SARS:** severe acute respiratory syndrome

**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 21.07.20; peer-reviewed by J Xu, P Banik, A Ramazan; comments to author 31.08.20; revised version received 10.09.20; accepted 13.09.20; published 08.10.20.*

### *Please cite as:*

Pan Y, Xin M, Zhang C, Dong W, Fang Y, Wu W, Li M, Pang J, Zheng Z, Wang Z, Yuan J, He Y

Associations of Mental Health and Personal Preventive Measure Compliance With Exposure to COVID-19 Information During Work Resumption Following the COVID-19 Outbreak in China: Cross-Sectional Survey Study

*J Med Internet Res* 2020;22(10):e22596

URL: <http://www.jmir.org/2020/10/e22596/>

doi: [10.2196/22596](https://doi.org/10.2196/22596)

PMID: [32936776](https://pubmed.ncbi.nlm.nih.gov/32936776/)

©Yihang Pan, Meiqi Xin, Changhua Zhang, Willa Dong, Yuan Fang, Wenhui Wu, Mingzhe Li, Jun Pang, Zilong Zheng, Zixin Wang, Jinqiu Yuan, Yulong He. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Social Media Use, eHealth Literacy, Disease Knowledge, and Preventive Behaviors in the COVID-19 Pandemic: Cross-Sectional Study on Chinese Netizens

Xiaojing Li<sup>1\*</sup>, Prof Dr; Qinliang Liu<sup>1\*</sup>, DPhil

Center for Health and Medical Communication, School of Media & Communication, Shanghai Jiao Tong University, Shanghai, China

\* all authors contributed equally

**Corresponding Author:**

Xiaojing Li, Prof Dr

Center for Health and Medical Communication

School of Media & Communication

Shanghai Jiao Tong University

800 Dongchuan Rd

Shanghai, 200240

China

Phone: 86 13918611103

Email: [lixiaojing@sjtu.edu.cn](mailto:lixiaojing@sjtu.edu.cn)

## Abstract

**Background:** Since its outbreak in January 2020, COVID-19 has quickly spread worldwide and has become a global pandemic. Social media platforms have been recognized as important tools for health-promoting practices in public health, and the use of social media is widespread among the public. However, little is known about the effects of social media use on health promotion during a pandemic such as COVID-19.

**Objective:** In this study, we aimed to explore the predictive role of social media use on public preventive behaviors in China during the COVID-19 pandemic and how disease knowledge and eHealth literacy moderated the relationship between social media use and preventive behaviors.

**Methods:** A national web-based cross-sectional survey was conducted by a proportionate probability sampling among 802 Chinese internet users ("netizens") in February 2020. Descriptive statistics, Pearson correlations, and hierarchical multiple regressions were employed to examine and explore the relationships among all the variables.

**Results:** Almost half the 802 study participants were male (416, 51.9%), and the average age of the participants was 32.65 years. Most of the 802 participants had high education levels (624, 77.7%), had high income >¥5000 (US \$736.29) (525, 65.3%), were married (496, 61.8%), and were in good health (486, 60.6%). The average time of social media use was approximately 2 to 3 hours per day (mean 2.34 hours, SD 1.11), and the most frequently used media types were public social media (mean score 4.49/5, SD 0.78) and aggregated social media (mean score 4.07/5, SD 1.07). Social media use frequency ( $\beta=.20$ ,  $P<.001$ ) rather than time significantly predicted preventive behaviors for COVID-19. Respondents were also equipped with high levels of disease knowledge (mean score 8.15/10, SD 1.43) and eHealth literacy (mean score 3.79/5, SD 0.59). Disease knowledge ( $\beta=.11$ ,  $P=.001$ ) and eHealth literacy ( $\beta=.27$ ,  $P<.001$ ) were also significant predictors of preventive behaviors. Furthermore, eHealth literacy ( $P=.038$ ) and disease knowledge ( $P=.03$ ) positively moderated the relationship between social media use frequency and preventive behaviors, while eHealth literacy ( $\beta=.07$ ) affected this relationship positively and disease knowledge ( $\beta=-.07$ ) affected it negatively. Different social media types differed in predicting an individual's preventive behaviors for COVID-19. Aggregated social media ( $\beta=.22$ ,  $P<.001$ ) was the best predictor, followed by public social media ( $\beta=.14$ ,  $P<.001$ ) and professional social media ( $\beta=.11$ ,  $P=.002$ ). However, official social media ( $\beta=.02$ ,  $P=.597$ ) was an insignificant predictor.

**Conclusions:** Social media is an effective tool to promote behaviors to prevent COVID-19 among the public. Health literacy is essential for promotion of individual health and influences the extent to which the public engages in preventive behaviors during a pandemic. Our results not only enrich the theoretical paradigm of public health management and health communication but also have practical implications in pandemic control for China and other countries.

(*J Med Internet Res* 2020;22(10):e19684) doi:[10.2196/19684](https://doi.org/10.2196/19684)

**KEYWORDS**

social media; media use; COVID-19; pandemic; disease knowledge; eHealth literacy; public health; preventive behaviors

## Introduction

### Background

COVID-19, an acute infectious disease, quickly spread worldwide after it emerged in December 2019 and has evolved from an epidemic to a pandemic. As of the end of May 2020, over 200 countries and territories had reported laboratory-confirmed cases of COVID-19, and the global number of confirmed cases of COVID-19 had exceeded 6,000,000 [1]. As a global pandemic, SARS-CoV-2, the novel coronavirus that causes COVID-19, has infected more people than either of its two predecessors, severe acute respiratory syndrome coronavirus (SARS-CoV) in 2003 and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 [2]; thus, COVID-19 poses a serious threat to global development. There has been an obvious rise in the number of emerging and reemerging infectious diseases over the past two decades, such as severe acute respiratory syndrome (SARS, 2003), H1N1 (2009), Middle East respiratory syndrome (MERS, 2012), Ebola virus (2014), and Zika virus (2016). All these infections were difficult to control due to a lack of effective vaccines and medicines, which led to great concern and anxiety among the public and to challenges for public health systems [3,4].

Preventive behaviors are essential to control infectious diseases from both public and individual perspectives. Authorities and public health agencies should implement a variety of pharmaceutical and nonpharmaceutical interventions to prevent pandemic expansion, including vaccination and medical prophylaxis, hygienic precautions, patient isolation, and other social distancing measures [5]. Individuals should also take preventive measures to protect themselves, such as washing hands frequently with soap or hand sanitizer, avoiding crowded gatherings, and wearing face masks when going outside [6]. Because many infectious diseases erupt in a short time and have high morbidity and mortality rates, it is difficult for executive agencies to impose sufficient interventions to control these diseases in a timely fashion. Thus, effective disease-management activities benefit greatly from preventive measures by individuals [7]. Therefore, educating the public to enhance health awareness and increase disease knowledge is crucial in a pandemic.

Information communication and media use are well suited to achieve this goal by providing the public with professional information, decreasing public panic, disseminating health knowledge, and expressing appreciation to the public for their cooperation [8]. Regarding the COVID-19 pandemic, information communication is still crucial for disease prevention. China has potential advantages in the area of social media. Since the rapid development of the internet and emerging mobile media technologies, China has made remarkable achievements in mobile digital communication. Chinese internet users are also called “netizens,” defined as Chinese citizens who use the internet for at least 1 hour per week by the China Internet Network Information Center (CNNIC); these netizens have been

marked by the rise of a highly connected and digitally empowered general public [9]. As of June 2019, the number of Chinese netizens had reached 847 million according to the CNNIC [10]. Social media applications are becoming increasingly diversified; WeChat, Weibo, QQ, and TikTok are the most frequently used platforms by Chinese netizens.

Also, social media is widely used by Chinese authorities to inform the public about the latest news, disseminate public health knowledge, refute rumors, and facilitate effective coordination of medical, public, and pharmaceutical resources. Although social media has been broadly used in China, the effects of social media on disease prevention have still not been greatly investigated. In this study, we hope to explore the predictive role of social media use in public preventive behaviors and how health literacy moderates the causality between individuals' social media use and preventive behaviors during the COVID-19 pandemic in Chinese contexts.

### Literature Review and Hypotheses

The mechanisms underlying the effects of social media use on health behavioral changes is that coverage of a pandemic on social media can magnify the public's fear and urge the public to take preventive actions [11]. Prior studies indicated that mass media use can produce positive changes or prevent negative changes in health-related behaviors across large populations [12]; for example, frequency of listening to the radio and reading the newspaper were associated with increased odds of being vaccinated [13], while time spent watching television was positively correlated with water, sanitation, and hygiene behaviors [14]. Comparatively, social media (eg, Facebook, Twitter, WeChat, Weibo) has provided the public and health institutes with new avenues for disease prevention during an epidemic or pandemic, as it allows two-way communication between health authorities and the public. Social media has also been found to be useful in terms of health-promotion interventions, such as preventing increases in risky sexual behavior [15], contributing to improved knowledge and attitudes toward skin cancer [16], positively influencing maternal influenza vaccine uptake [17], and targeting lifestyle changes among users with chronic diseases [18]. Additionally, studies on the effects of social media have shed light on its utility in public health domains. For example, Facebook was used for strategic crisis communication by health authorities in Singapore during the Zika virus pandemic [19]; moreover, WeChat and Weibo use were found to significantly increase preventive behaviors for haze health [20]. Scholars are paying increasing attention to the role of social media during pandemics; however, the question of whether social media use can affect the public's affective responses or preventive behaviors still deserves exploration. Thus, we propose the first research question:

*RQ1: Does social media use predict preventive behaviors among Chinese netizens during the COVID-19 pandemic?*

Social cognitive theory is used to explain how people learn behaviors by observing others. It emphasizes the reciprocal



causation of individual behaviors between personal factors (eg, values, self-efficacy, outcome expectations), behavioral factors (eg, prior behavior) and social environmental factors (eg, others' behaviors, feedback). This theory provides a conceptual framework of how media use influences human beings' thoughts, affect, and actions. Media use leads to behavioral changes by communicating information through two pathways. On one hand, media use promotes changes by informing, enabling, motivating, and guiding users to take direct action to effect change [21]. On the other hand, people adopt, support, spread, and share innovative ideas or behaviors in the socially mediated pathways of social media [22]. As a socially mediated factor, social media frames and reinforces social norms and enriches the ability of the public to receive health information, such as news, knowledge, and health behavior patterns. This knowledge can be rapidly and widely diffused by exerting social influences on people's health behaviors through observational learning [23]. Therefore, the degree to which people's use of social media to access health information for disease management may influence an individual's health behavioral outcomes.

As media use is a composite concept that comprises a cluster of measurements, research questions about media use and health behaviors are usually presented as "how many hours did you spend on [social media platform, such as Facebook, Twitter, or YouTube] per day?" [13] or "how many times did you use a particular social media platform?" [24], which can be respectively summarized as "time of media use" (ie, how long) and "frequency of media use" (ie, how often). Time and frequency are also known to be the key variables of social media use. Thus, we proposed two hypotheses:

*H1: Social media use time is positively associated with preventive behaviors during the COVID-19 pandemic.*

*H2: Social media use frequency is positively associated with preventive behaviors during the COVID-19 pandemic.*

In addition to time and frequency, type is a crucial dimension of social media use. As the media landscape has changed dramatically, media types have rapidly become diversified in the new media environment [25]. In China, users usually obtain news or information via mobile news channels. The number of web-based news users has been reported to be 686 million, which accounts for 80.3% of Chinese netizens [10]. Web-based mobile news channels mostly consist of various applications that are characterized by social interactive functions such as reading, commenting, retweeting, and timely interaction. These platforms can be divided into different types by their functions. Official social media outlets, such as China Central Television (CCTV) and People's Daily, often serve as the voice of government or administrative institutions. Professional social media is an emerging form of social media that focuses on news in the professional domain. For example, Caixin News focuses on finance. Aggregated social media is a new type of media that collects and distributes news or information from different agencies. The scope of news on aggregated social media is widespread, including politics, the economy, culture, sports, and entertainment. Public social media (eg, WeChat, Weibo,

TikTok), also called interpersonal social media, is produced and disseminated by individuals. Netizens can use public social media to share news with their friends or strangers. All the above types of social media include almost all the social media platforms in China, and each media type is aimed at particular users. For instance, traditional official media represents the official voice of the government, while public or aggregated social media provides voices to grassroots organizations or individuals [26].

At the same time, various types of social media appear to have different effects. Web-based content has been reported to facilitate safer sex literacy and information-sharing intentions on social networking sites [27]. Traditional media (eg, television and radio) can be a more effective tool for managing crises than social media and websites; meanwhile, social media should also be considered to be effective during public health interventions, as younger people heavily rely on social media to seek information [28]. Additionally, when messages are transmitted through reliable web-based personal broadcasting channels, they can induce new attitudes or intentions to change in users [29]. In particular, previous studies have examined the associations of particular types of media access with information-seeking behaviors. For example, Alhuwail and Abdulsalam [30] indicated that people searched YouTube most for health information, but they did not place a high value on other social media platforms such as Twitter, Snapchat, and Facebook. Stawarz et al [31] found in their investigation that people used mobile technologies to support their mental health for specific purposes. Hence, inspired by previous results, it is essential to examine the relationship between different social media types and the public's preventive behaviors for COVID-19. Here, we propose another research question:

*RQ2: Do social media types (official social media, professional social media, public social media, aggregated social media) differ in terms of predicting users' preventive behaviors during the COVID-19 pandemic?*

## Health Literacy and Preventive Behaviors

### eHealth Literacy

The predictors of preventive measures are not merely based on the external impact of social media but also involve internal "assets," including the set of health knowledge, skills, and capabilities that is called *health literacy*. As a discrete form of literacy, health literacy is becoming increasingly important in predicting health promotion and prevention [32]. In 2004, the US Institute of Medicine [33] defined health literacy as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions." This concept is also interpreted and has evolved as a wide range of skills that people develop to seek out, comprehend, evaluate, and use health information.

The internet is now widely used and has drastically changed how health information is disseminated [34]. eHealth literacy combines information and media literacies and applies them to eHealth promotion. It has been defined as "the ability to seek,



find, understand, and appraise health information from electronic sources and to apply the knowledge gained to addressing or solving health problems [35].” eHealth literacy is becoming increasingly important as individuals continue to seek medical advice from various web-based sources, especially social media. Empirical studies have also found that eHealth literacy positively influences health outcomes, such as health-promoting behaviors among people with diabetes [36] and people’s health-related quality of life [37]. College students with higher eHealth literacy were found to be less likely to consume unhealthy food [38].

### Disease Knowledge

In addition to eHealth literacy, disease knowledge is a vital component of health literacy; it enables people to recognize the symptoms, understand the causes, and perceive the risks of chronic diseases or infectious diseases [39]. Disease knowledge is also effective in improving health management, and it even acts as a predictor of change in an individual’s health behaviors. Authorities are generally implementing additional measures to improve the level of disease knowledge among the public, with the aim of changing the attitudes of citizens toward public health prevention [40]. For example, disease knowledge can change attitudes and practices toward rabies prevention [41], levels of oncological knowledge had an impact on individuals’ decisions to consent to particular medical procedures [42], and higher

public health knowledge was positively associated with more frequent handwashing [14].

Additionally, disease knowledge and eHealth literacy can combine as intermediate factors linking to health status [43]. eHealth literacy has been independently related to disease knowledge; it also further influences disease knowledge by an indirect pathway [44]. For example, diabetes knowledge was the most important factor associated with glycemic control, and health literacy through diabetes knowledge exerted an indirect influence on self-care and medication adherence [45].

Therefore, we propose four hypotheses here:

*H3: eHealth literacy is positively associated with preventive behaviors during the COVID-19 pandemic.*

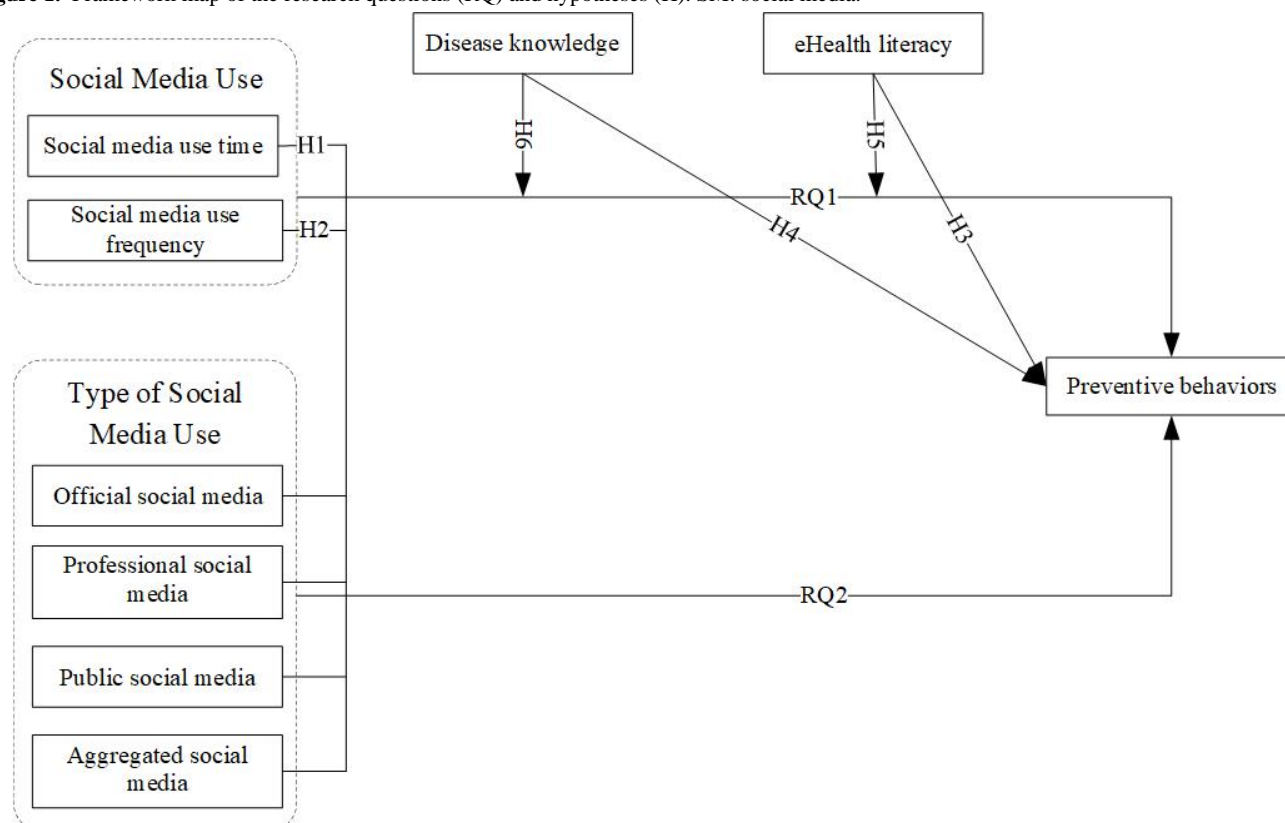
*H4: Disease knowledge is positively associated with preventive behaviors during the COVID-19 pandemic.*

*H5: eHealth literacy moderates the relationship between social media use and preventive behaviors during the COVID-19.*

*H6: Disease knowledge moderates the relationship between social media use and preventive behaviors during the COVID-19 pandemic.*

Figure 1 presents all the core variables and research hypotheses examined in this study.

**Figure 1.** Framework map of the research questions (RQ) and hypotheses (H). SM: social media.



## Methods

### Design and Recruitment

A national web-based cross-sectional survey was executed by proportionate probability sampling in this study to examine

whether social media use predicted Chinese netizens’ preventive behaviors during the COVID-19 pandemic and to explore the moderators of disease knowledge and eHealth literacy. The proportionate probability sampling method was employed according to the gender and age distributions of Chinese netizens reported in the 44th Statistical Report on Internet Development

of China (SRIDC) [10]. The SRIDC is an authoritative report that is released annually by the CNNIC and is based on a representative national survey with a sample size of 60,000. As the report showed, people 20 to 60 years of age were the main body of Chinese netizens; they represented 72.3% of the entire sample. In our survey, the web-based sample pool had an age limitation in that participants >60 years of age were rare. Thus, we selected 20 to 60 years of age as the target sample age range. We set the age intervals and proportions as 20 to 29 years of age (34.02%), 30 to 39 years of age (32.78%), 40 to 49 years of age (23.93%) and 50 to 59 years of age (9.27%); also, the proportions of men and women for each age range were 52.4% and 47.6%, respectively, according to the population distribution of Chinese netizens; these proportions were also in line with the SRIDC.

Participants were recruited using a web-based platform from the Questionnaire Star survey company [46], which contains over 2.6 million registered panelists in its sample pool. A structured questionnaire was developed and pretested for this study (Multimedia Appendix 1). Then, the web-based survey was partially adjusted and formally executed. The survey was conducted from February 13 to 21, 2020. After excluding ineligible samples (eg, incomplete or completed in a very short time), we finally collected 802 valid questionnaires from 952 respondents. The valid response rate was 84.24%.

## Ethics Statement

Authorization to conduct the research and recruit participants was obtained from the Institutional Review Board of the authors' university (ID: 20200203). In addition, the purpose of this study was elucidated by the "Notification of Sample Service" (Survey ID: 57071374). Consent was obtained from all the participants before the web-based survey was conducted by the survey agency [46]. Participation was completely voluntary, and the participants could choose to quit at any time for any reason during the process of answering the web-based questionnaire.

## Instruments

### Demographic Information

The six most frequently used sociodemographic variables were collected, including gender (0=female and 1=male), age (the respondents reported their birth year and we computed their actual age, eg, if the respondent entered "1980," we computed 2020 - 1980 to obtain an age of 40 years), education (from 1=middle school or less to 5=master's degree and above), monthly income (1, <¥1500; 2, ¥1500 to 3000; 3, ¥3001 to 5000; 4, ¥5001 to 8000; 5, ¥8001 to 12,000; 6, ¥12,001 to 20,000; 7, >¥20,000; 1 ¥=US \$0.14), marital status (1, single; 2, divorced or widowed; 3, separated; 4, cohabiting; 5, married), and health status (from 1=severe disease to 5=good).

### Social Media Use

Media use was measured by the following questions: social media use time ("In the past week, how much time did you spend using social media every day to learn about news of the COVID-19 pandemic?") with answers ranging from "less than one hour" to "5 hours and more"); type of social media use ("Which channel do you use often to obtain COVID-19

information every day?") with four types of social media channels, including "Official social media, such as People's Daily," "Professional social media, such as Ding Xiang Doctor," "Public social media, such as WeChat," and "Aggregated social media, such as Tencent News," with possible answers for each social media channel of 1, never used; 2, 1 to 2 times per week; 3, 3 to 4 times per week; 4, 5 to 6 times per week, and 5, one or more times per day). Additionally, the variable of social media use frequency was measured by the sum score of the frequencies of all four types of social media channels (maximum score: 20), and a higher score indicates more frequent use of social media.

### Preventive Behaviors

Preventive behaviors were measured by 10 items consisting of basic protective recommendations during the COVID-19 pandemic (eg, "Washing your hands after going home" and "Covering your mouth and nose with a tissue or sleeves when you cough or sneeze"). The 10 items were assessed by a self-reported measurement scale. Firstly, the measures of disease knowledge were drawn from the COVID-19 Protection Manual (Hong Kong version, February 2020) [47] and COVID-19 Protection Manual (China Mainland version, January 2020) [48]. 20 items were generated as alternative metrics. Second, we consulted with medical experts on all the metrics. According to their suggestions, we selected 10 items as the final measurement metrics. Before the formal survey was conducted, we invited 10 adults to conduct a pilot study and modified the survey correspondingly until the validity and reliability were acceptable. Finally, we adopted the adapted measures. Respondents were asked to indicate the extent to which they agreed with the statements on a 5-point Likert scale ranging from 1=never executed to 5=do it every time (Cronbach  $\alpha$ =.75).

### Disease Knowledge

Disease knowledge was assessed by a self-reported measurement scale consisting of 10 items (eg, "The incubation period of COVID-19 infections is generally 3-7 days, with a maximum of 14 days," "The coronavirus volume is about 3 microns"). Like the measurement process of preventive behaviors, the instrument of disease knowledge was drawn from the COVID-19 Protection Manual (Hong Kong Version, February 2020) [47] and COVID-19 Protection Manual (China Mainland version, January 2020) [48]. We generated 20 items, also in consultation with medical experts. Finally, 10 items were used as the final measurement metrics via a pilot study. The answer options were "yes" or "no" for each item. Participants were given 1 point for the correct answer and 0 points for an incorrect response for each item. The variable of disease knowledge had possible scores of 0 to 10 (Cronbach  $\alpha$ =.70).

### eHealth Literacy

eHealth Literacy was assessed by the 8-item eHealth Literacy Scale (eHEALS) [34]. The eHEALS is a reliable computer-based measure of patients' knowledge and self-efficacy for obtaining and evaluating web-based health resources. This brief scale assesses an individual's perceived ability to find, understand, and appraise health information from web-based sources and apply that knowledge to address health concerns (eg, "I know

what health resources are available on the internet” and “I know where to find helpful health resources on the internet”). The eHEALS was developed in English. It was translated into a Chinese version for our questionnaire, and we invited 5 adults to conduct a pilot study. The results indicated that the reliability of the Chinese version is high; therefore, we adopted it. Response options included a 5-point Likert scale ranging from 1=totally disagree to 5=totally agree (Cronbach  $\alpha=.82$ ).

### Statistical Analysis

Descriptive statistics were used to assess the sociodemographic characteristics of the respondents, including gender, age, education, monthly income, marital status, and health status. Category variables were described as *n* (%). Continuous variables were expressed as mean (SD). Category variables (including education, monthly income, marital status, and health status) were also dummy-coded, and one group was set as a reference group in each category. Pearson correlation analysis and hierarchical multiple regression were employed. Two-tailed Pearson correlations were used to examine the correlations between the control variables and the independent and dependent variables, respectively.

Two hierarchical regression models were used to test the research questions and hypotheses. The first hierarchical multiple regression was used to investigate RQ1, H1, H2, H3, H4, H5, and H6, in which the demographics were set as the control variables for Model 1. Then, the social media use time and social media use frequency were introduced in Model 2, and disease knowledge and eHealth literacy were introduced in Model 3. Finally, the two interaction items of social media use

frequency  $\times$  disease knowledge and social media use frequency  $\times$  eHealth literacy were entered in Model 4. Two additional interaction items, time  $\times$  eHealth literacy and time  $\times$  disease knowledge, were entered in Model 5. The second hierarchical regression was carried out to explore the predictors of the four social media types (RQ2). The demographics were set as the control variables for Model 1, and four types of social media channels (official social media, professional social media, public social media, aggregated social media) were introduced in Model 2. All statistical analyses were calculated with SPSS for Windows version 22.0 (IBM Corporation).

## Results

### Descriptive Statistics

#### *Sociodemographic Profiles*

Among the 802 participants, 416 (51.9%) were male and 386 (48.1%) were female. The ages of the respondents ranged from 20 to 60 years, which is representative of Chinese netizens according to 2019 CNNIC statistics [10]. The sample overrepresented high education (above bachelor's degree, 624/902, 77.7%) and high monthly income of  $>¥5000$  (US \$736.29, 525/802, 65.3%) compared with the respective values of 9.7% and 27.1% reported in the SRIDC. Most of the respondents had a bachelor's (undergraduate) degree or higher, and nearly half of the respondents' monthly income was  $>¥8000$  (US \$1178). Additionally, the majority of the respondents in our sample were married (496/802, 61.8%) and in good health (486/802, 60.6%). A detailed comparison of our sample profile and the CNNIC sample is presented in Table 1.

**Table 1.** Sociodemographic characteristics of our research sample and the CNNIC sample.

Characteristic	Research sample (N=802), n (%)	CNNIC <sup>a</sup> sample (N=60,000), %
<b>Gender</b>		
Female	386 (48.1)	47.6
Male	416 (51.9)	52.4
<b>Age (years)</b>		
<20	N/A <sup>b</sup>	20.9
20-29	318 (39.7)	24.6
30-39	288 (35.9)	23.7
40-59	196 (24.4)	24.0
>60	N/A	6.9
<b>Education</b>		
Primary school and below	N/A	18.0
Middle school	9 (1.1)	38.1
High school	54 (6.7)	23.8
Associate degree	115 (14.4)	10.5
Bachelor's degree	547 (68.1)	N/A
Bachelor's degree and above <sup>c</sup>	N/A	9.7
Master's degree and above	77 (9.6)	N/A
<b>Monthly income (¥)<sup>d</sup></b>		
<1500	50 (6.2)	31.7
1500-3000	68 (8.5)	20.3
3001-5000	159 (19.9)	20.8
5001-8000	242 (30.1)	14.1
8001-12,000	191 (23.8)	13.0
12,001-20,000	78 (9.7)	N/A
>20,000	14 (1.7)	N/A

<sup>a</sup>CNNIC: China Internet Network Information Center.<sup>b</sup>N/A: not applicable.<sup>c</sup>In the CNNIC survey, "Bachelor's degree and above" was a single category.<sup>d</sup>1 ¥=US \$0.14 on February 13, 2020.

### **Characteristics of Social Media Use, Health Literacy, and Preventive Behaviors**

**Table 2** presents the basic characteristics of social media users in terms of social media use, disease knowledge, eHealth literacy, and preventive behaviors. Respondents did not spend much more time on social media every day to learn about the COVID-19 pandemic, as the average social media use time was approximately 2 to 3 hours per day (mean 2.34, SD 1.12). By contrast, the respondents used social media more often (mean

score 13.59/20, SD 2.42) when compared with reference point 12. As the types of social media channels, respondents liked to use public social media and aggregated social media more than official social media and professional social media. Respondents had a high level of disease knowledge (mean score 8.15/10, SD 1.43) and eHealth literacy (mean score 3.79/5, SD 0.59). Moreover, respondents also took many preventive behaviors (mean score 4.30/5, SD 0.44) for health management during the COVID-19 pandemic.

**Table 2.** Characteristics of social media use, disease knowledge, eHealth literacy and preventive behaviors (N=802), mean (SD).

Characteristic	Value
Social media use time (hours)	2.34 (1.11)
Social media use frequency <sup>a</sup>	13.59 (2.42)
<b>Social media type<sup>b</sup></b>	
Official social media	2.54 (1.20)
Professional social media	2.48 (1.11)
Public social media	4.49 (0.78)
Aggregated social media	4.07 (1.07)
Disease knowledge <sup>c</sup>	8.15 (1.43)
eHealth literacy <sup>d</sup>	3.79 (0.59)
Preventive behaviors <sup>e</sup>	4.30 (0.44)

<sup>a</sup>Measured by the sum score of the frequencies of all four types of social media channels (maximum score: 20).

<sup>b</sup>Measured on a scale with scores of 1=never used to 5=one or more times per day.

<sup>c</sup>Measured by 10 yes/no questions with a possible score of 1 to 10 (Cronbach  $\alpha=.70$ ).

<sup>d</sup>Measured by the 8-item eHealth Literacy Scale with scores of 1=totally disagree to 5=totally agree (Cronbach  $\alpha=.82$ ).

<sup>e</sup>Measured by a 10-item scale with scores of 1=never executed to 5=do it every time (Cronbach  $\alpha=.75$ ).

## Predictors and Moderators of Preventive Behaviors

Before the two hierarchical multiple regressions were conducted, Pearson correlations were employed to assess the correlations between independent variables and dependent variables. As displayed in the correlation table in [Multimedia Appendix 2](#), significant correlations exist between demographics, social media use, disease knowledge, eHealth literacy, and preventive behaviors; however, social media use time ( $\beta=.07$ ,  $P>.05$ ) did not predict preventive behaviors. Thus, H1 was not supported.

To examine the predictors and moderators of the preventive behaviors, the first hierarchical multiple regression was carried out, and the full results are shown in [Table 3](#) (the change in  $R^2$  upon adding the interaction of the last step of Model 5 was insignificant; therefore, we selected Model 4 as our final model). Social media use frequency ( $\beta=.20$ ,  $P<.001$ ), disease knowledge ( $\beta=.11$ ,  $P=.001$ ), and eHealth literacy ( $\beta=.27$ ,  $P<.001$ ) significantly and positively predicted preventive behaviors, respectively, when controlling sociodemographic variables (gender, age, education, monthly income, marital status, and health status). eHealth literacy ( $\beta=.27$ ) also emerged as the main effect. These results supported H2, H3, and H4; they also partly answered RQ1, which states that social media use frequency rather than social media use time can predict preventive behaviors during the COVID-19 pandemic.

The results showed significant correlations of the social media use frequency  $\times$  disease knowledge and social media use frequency  $\times$  eHealth literacy interactions with preventive

behaviors ( $\beta=-.07$ ,  $P=.03$ , and  $\beta=.07$ ,  $P=.04$ , respectively). These results indicate that disease knowledge and eHealth literacy significantly moderate the relationship between social media use frequency and preventive behaviors. Moreover, eHealth literacy positively moderated the relationship between social media use frequency and preventive behaviors, while disease knowledge negatively moderated this relationship. We also checked the moderator effects of social media use time  $\times$  eHealth literacy ( $\beta=.02$ ,  $P=.51$ ) and social media use time  $\times$  disease knowledge ( $\beta=.05$ ,  $P=.15$ ); however, both these correlations were insignificant. Thus, H5 and H6 were partly supported.

The slope test is often applied to test the magnitude of a moderated effect on the conditional value of a moderator. Given that the interaction items were significant, we performed slope tests and plotted the predicted preventive behaviors separately for high and low eHealth literacy or disease knowledge (1 SD above the mean and 1 SD below the mean, respectively; see [Figure 2](#) and [Figure 3](#)). The simple slope analyses indicated that for social media users who had lower levels of eHealth literacy, a higher level of frequency of social media use (mean  $-1$ SD) was associated with higher levels of preventive behaviors ( $\beta$  simple $=.02$ ,  $P<.001$ ). For people with higher levels of eHealth literacy (mean  $+1$ SD), the positive association between the frequency of social media and preventive behaviors was also significant ( $\beta$  simple $=.044$ ,  $P<.001$ ), and the magnitude of this association was greater than that for lower levels of eHealth literacy.



**Table 3.** Hierarchical multiple regression examining the predictors and moderators of preventive behaviors during the COVID-19 pandemic.

Variable	Model 1 <sup>a</sup>		Model 2 <sup>b</sup>		Model 3 <sup>c</sup>		Model 4 <sup>d</sup>		Model 5 <sup>e</sup>	
	$\beta^f$	P value	$\beta$	P value	$\beta$	P value	$\beta$	P value	$\beta$	P value
<b>Demographic</b>										
Female gender	-.11	.001	-.12	<.001	-.12	<.001	-.12	<.001	-.13	<.001
Age	.20	<.001	.20	<.001	.24	<.001	.24	<.001	.23	<.001
<b>Education</b>										
Middle school	Reference	N/A <sup>g</sup>	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
High school	.13	.13	.12	.17	.08	.33	.08	.34	.09	.29
Associate degree	.13	.27	.12	.29	.07	.50	.07	.54	.08	.49
Bachelor's degree	.18	.25	.16	.29	.07	.63	.06	.67	.07	.62
Master's degree and above	.09	.38	.08	.42	.02	.83	.01	.95	.01	.91
<b>Income (¥)<sup>h</sup></b>										
<1500	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
1500-3000	.05	.34	.02	.71	.06	.24	.06	.21	.07	.17
3001-5000	.12	.08	.05	.44	.09	.17	.09	.16	.08	.18
5001-8000	.19	.01	.10	.18	.11	.11	.12	.09	.11	.10
8001-12,000	.23	.001	.12	.08	.13	.055	.13	.06	.13	.06
12,001-20,000	.17	.002	.11	.049	.10	.07	.10	.070	.10	.07
>20,000	.08	.06	.05	.26	.05	.23	.05	.18	.05	.17
<b>Marital status</b>										
Married	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
Single	.03	.54	.06	.21	.08	.08	.08	.07	.07	.09
Divorced	.01	.78	.02	.50	.03	.34	.03	.31	.03	.33
Separated	-.06	.09	-.05	.11	-.05	.16	-.04	.18	-.04	.18
Cohabiting	-.07	.05	-.05	.14	-.05	.16	-.05	.13	-.05	.12
<b>Health status</b>										
Good	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A	Reference	N/A
Severe disease	-.01	.88	.01	.80	.02	.61	.01	.72	.01	.71
Chronic disease	-.07	.06	-.07	.03	-.07	.03	-.07	.03	-.07	.03
Suboptimal health	-.12	.001	-.12	<.001	-.12	<.001	-.11	.001	-.11	.001
Fair	-.17	<.001	-.15	<.001	-.13	<.001	-.13	<.001	-.13	<.001
<b>Social media use</b>										
Time	— <sup>i</sup>	—	-.03	.46	-.02	.51	-.02	.46	-.03	.40
Frequency	—	—	.25	<.001	.20	<.001	.20	<.001	.20	<.001
<b>Health literacy</b>										
eHealth literacy	—	—	—	—	.26	<.001	.27	<.001	.27	<.001
Disease knowledge	—	—	—	—	.11	.001	.11	.001	.11	.001
<b>Interactions</b>										
1. Social media use frequency × eHealth literacy	—	—	—	—	—	—	.07	.04	.05	.11

Variable	Model 1 <sup>a</sup>		Model 2 <sup>b</sup>		Model 3 <sup>c</sup>		Model 4 <sup>d</sup>		Model 5 <sup>e</sup>	
	$\beta^f$	<i>P</i> value	$\beta$	<i>P</i> value	$\beta$	<i>P</i> value	$\beta$	<i>P</i> value	$\beta$	<i>P</i> value
2. Social media use frequency $\times$ disease knowledge	—	—	—	—	—	—	-.07	.03	-.07	.03
3. Social media use time $\times$ eHealth literacy	—	—	—	—	—	—	—	—	.02	.51
4. Social media use time $\times$ disease knowledge	—	—	—	—	—	—	—	—	.05	.15

<sup>a</sup>Adjusted  $R^2=0.11$ ,  $\Delta R^2=0.13$ ,  $P<.001$ .

<sup>b</sup>Adjusted  $R^2=0.16$ ,  $\Delta R^2=0.05$ ,  $P<.001$ .

<sup>c</sup>Adjusted  $R^2=0.23$ ,  $\Delta R^2=0.07$ ,  $P<.001$ .

<sup>d</sup>Adjusted  $R^2=0.24$ ,  $\Delta R^2=0.01$ ,  $P=.01$ .

<sup>e</sup>Adjusted  $R^2=0.24$ ,  $\Delta R^2=0.002$ ,  $P=.28$ .

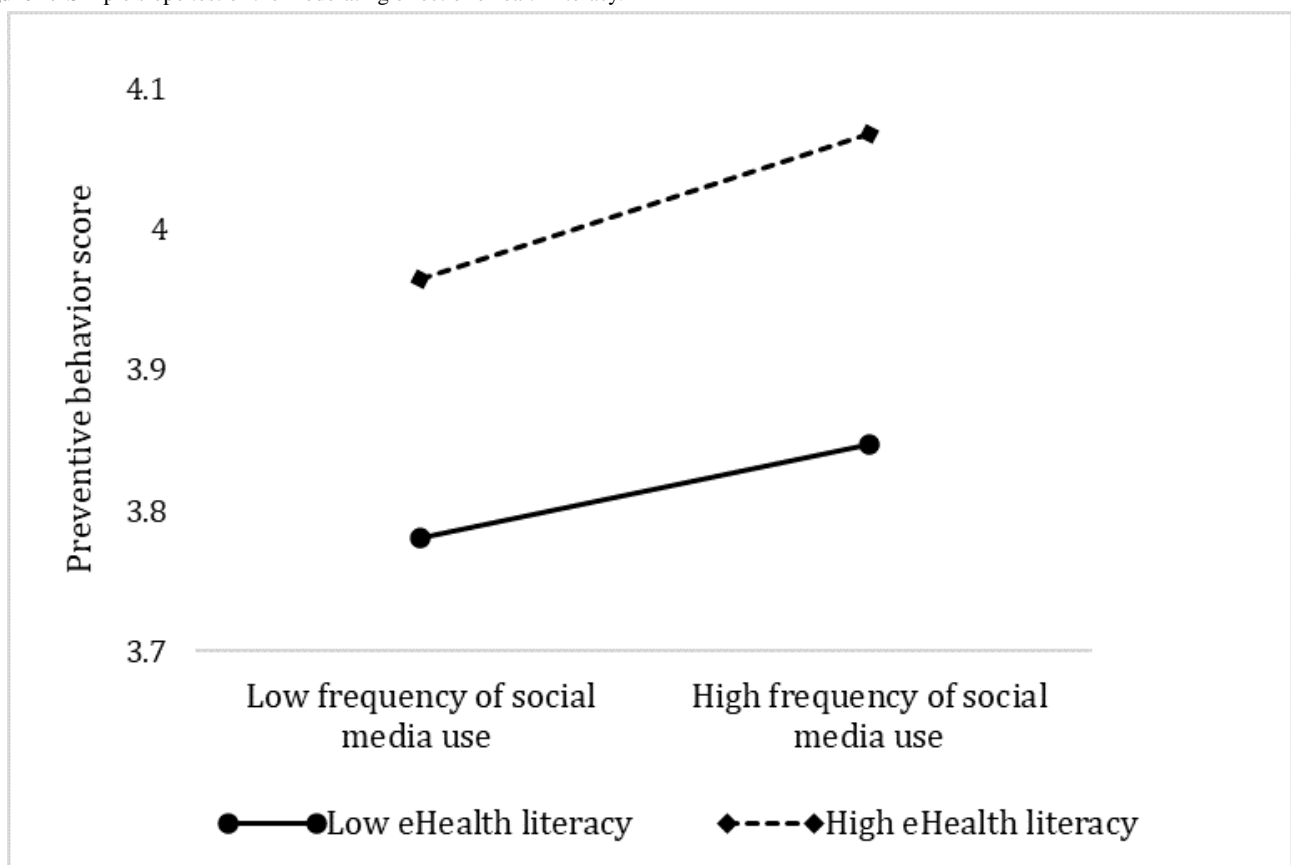
<sup>f</sup> $\beta$ : standardized regression coefficient.

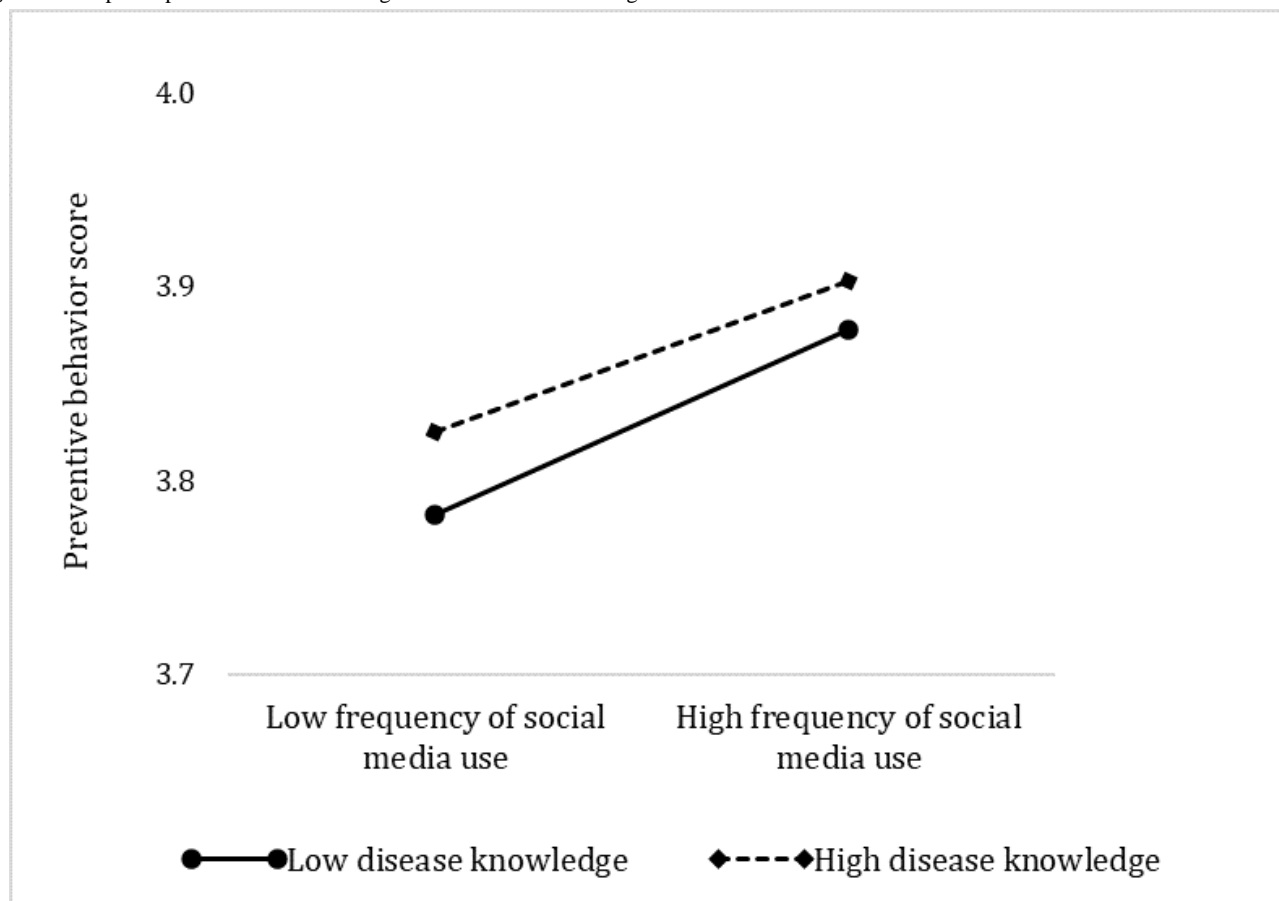
<sup>g</sup>N/A: not applicable.

<sup>h</sup>1 ¥=US \$0.14 on February 13, 2020.

<sup>i</sup>—: Not included in the model.

**Figure 2.** Simple slope test of the moderating effect of eHealth literacy.



**Figure 3.** Simple slope test of the moderating effect of disease knowledge.

For disease knowledge, simple slope analyses indicate that for social media users who had lower levels of disease knowledge (mean  $-1SD$ ), a higher level of social media use frequency was associated with higher levels of preventive behaviors ( $\beta$  simple=.060,  $P<.001$ ). For people with higher levels of disease knowledge (mean  $+1SD$ ), the positive association between frequency of social media and preventive behaviors was also significant; however, the magnitude of the association was smaller ( $\beta=.035$ ,  $P<.001$ ).

Concerning demographics, gender, age, monthly income, and health status were all found to significantly predict preventive behaviors. Age, monthly income, and health status positively predicted preventive behaviors. However, gender negatively predicted preventive behaviors. In detail, participants with a monthly income of more than ¥5000 engaged in more preventive behaviors than the reference group of people with a monthly income of less than ¥1500. Compared with participants who reported their health status as “good,” those who reported unhealthy status took fewer preventive measures. This suggests that social media users who were older, had higher monthly

income, and had better health status were more likely to take preventive measures during the COVID-19 pandemic. Women generally engaged in more preventive behaviors than men. However, marital status and education had no significant effects on preventive behaviors.

### Types of Social Media Use and Preventive Behaviors

RQ2 focused on comparisons among the four media genres, namely official social media, professional social media, public social media, and aggregated social media. As shown in Table 4, the multiple regression results indicated that professional social media ( $\beta=.11$ ,  $P=.002$ ), public social media ( $\beta=.14$ ,  $P<.001$ ), and aggregated social media ( $\beta=.22$ ,  $P<.001$ ) positively predicted preventive behaviors, while official social media ( $\beta=.02$ ,  $P=.60$ ) did not. Furthermore, aggregated social media was found to be the highest predictor of preventive behaviors, closely followed by public social media and professional social media. However, use of official social media in China did not predict netizens’ preventive behaviors. Additionally, health literacy positively moderated the relationship between social media use and preventive behaviors.

**Table 4.** Hierarchical multiple regression examining the predicting roles of different types of social media use on preventive behaviors.

Characteristic	Model 1		Model 2	
	$\beta^a$	<i>P</i> value	$\beta$	<i>P</i> value
<b>Demographic</b>				
Female gender	-.11	.001	-.11	.001
Age	.20	.000	.19	.000
<b>Education</b>				
Middle school	Reference	N/A <sup>b</sup>	Reference	N/A
High school	.13	.130	.08	.358
Associate degree	.13	.266	.06	.588
Bachelor's degree	.18	.248	.09	.560
Master's degree and above	.09	.384	.04	.683
<b>Income (¥)<sup>c</sup></b>				
<1500	Reference	N/A	Reference	N/A
1500-3000	.05	.348	.01	.909
3001-5000	.12	.076	.05	.473
5001-8000	.19	.011	.08	.284
8001-12,000	.23	.001	.10	.145
12,001-20,000	.17	.002	.09	.096
>20,000	.08	.056	.03	.471
<b>Marital status</b>				
Married	Reference	N/A	Reference	N/A
Single	.03	.542	.05	.260
Divorced	.01	.778	.02	.606
Separated	-.06	.086	-.04	.184
Cohabiting	-.07	.052	-.06	.067
<b>Health status</b>				
Good	Reference	N/A	Reference	N/A
Severe disease	-.01	.880	.01	.675
Chronic disease	-.07	.056	-.06	.072
Suboptimal	-.12	.001	-.12	.000
Fair	-.17	.000	-.15	.000
<b>Social media type</b>				
Official social media	N/A	N/A	.02	.597
Professional social media	N/A	N/A	.11	.002
Public social media	N/A	N/A	.14	.000
Aggregated social media	N/A	N/A	.22	.000

<sup>a</sup> $\beta$ : standardized regression coefficient.<sup>b</sup>N/A: not applicable.<sup>c</sup>1 ¥=US \$0.14 on February 13, 2020.

## Discussion

This study had three goals. The first goal was to explore the predictors of preventive behaviors during the COVID-19

pandemic, the second goal was to examine the roles of disease knowledge and eHealth literacy in moderating public preventive behaviors, and the third goal was to explain the relationship between demographics and people's preventive behaviors. The

findings revealed that social media use frequency, disease knowledge, and eHealth literacy all positively predicted an individual's preventive behaviors during the COVID-19 pandemic. Aggregated social media, public social media, and professional social media were the significant predictors of preventive behaviors within the four social media channels. Moreover, eHealth literacy positively moderated the relationship between social media use frequency and preventive behaviors, while disease knowledge negatively affected this relationship. Concerning demographics, female sex, older age, high monthly income, and good health status were likely to take more preventive measures during the COVID-19 pandemic in China.

### Social Media Use and Preventive Behaviors

For a long time, mass media (eg, television, radio, and newspapers) was recognized as an important strategy for health-promoting practice [49]. For example, a mass media campaign increased physical activity, produced positive changes, and prevented negative changes in health-related behaviors [50]. Government and executive agencies have generally used mass media and social media as convenient tools for supervising and preventing epidemics. According to the main results of this study, social media use (frequency) played a positive role in public preventive behaviors during the COVID-19 pandemic in China. This may be an important indicator of health promotion, which encourages the public to take more health measures during emergencies. Compared to mass media, social media provides the public with convenient channels to obtain news or disease knowledge and delivers information effectively. Thus, social media should be an effective strategy for public health promotion, especially during an epidemic or a pandemic.

In contrast with social media use time (which was nonsignificant), social media use frequency was a significant predictor of preventive behaviors. In other words, "how often" rather than "how long" social media was used was a good predictor of an individual's preventive behaviors; this was an unexpected but interesting finding in this work. Time and frequency are often used to measure the regularity of social media use [51]. We attempted to draw an explanation from previous studies that investigated the relationship between social media use frequency and behavioral outcomes; we found that "frequency" may be a direct indicator of the motivations of social media use, such as self-expression, social learning, social comparison, or filtering [52,53]. Therefore, we cautiously concluded that frequency of social media use indicates the degree of engagement or investment in social media. "Frequency" may thus be a more significant predictor of social media effects.

### Types of Social Media Use and Preventive Behaviors

The positive correlation of social media use and preventive behaviors extended the study of the relationships between different types of social media use and preventive behaviors. Aggregated social media use was found to be the most significant predictor of preventive behaviors among four types of social media channels, followed by public social media and professional social media use. In contrast, official social media use was not significant. These results indicate that new media access (aggregated social media, public social media, and

professional social media) deserved more attention in affecting public preventive measures than traditional media (official social media), particularly in Chinese contexts.

Aggregated social media, a novel type of news aggregator, has ensured that readers can read news stories of high quality from many outlets; this simplifies the search process of news stories and allows users to save time and effort in finding news [54]. News aggregators such as Tencent News, Sina News, and Toutiao have emerged as important components of digital content ecosystems in China, along with overseas Google News, Reddit, Bing News, etc. These aggregated social media sites have drastically changed the ways in which users access information and interact with each other. They can also generate a substitution effect when users switch from news outlets (official media) to news aggregators [55]. Consequently, aggregated social media is competing with official social media for more users' attention and has led to an intensified propaganda crisis of official social media. This may partially explain our finding that aggregated social media was the most significant predictor for preventive behaviors among the four social media types, while official social media was not significant. Furthermore, official media outlets, such as CCTV, People's Daily and Xinhua Net, are state-driven media platforms in Chinese contexts. The content of official social media platforms mainly focuses on party ideology or party image [56]; meanwhile, the content spectrum of more extensive social imperatives is limited [57]. Therefore, the readability and humanity of public health content on official social media are lower than on aggregated social media, which may be another reason for the insignificant effect of official social media on public preventive behaviors.

Additionally, we found that public social media (eg, WeChat, Weibo, and TikTok) played a vital role in affecting users' adoption of preventive behaviors. Because public social media is the most popular media type in China, it accelerates news diffusion among people or across regions and enables users to learn from each other [31]. On the other hand, public social media mostly disseminates information via interpersonal communication, which intensifies the perceived credibility of this type of social media [58,59]. Thus, public social media can act as a significant predictor for preventive behaviors. Finally, as an emerging web-based platform, professional medical social media sites such as Ding Xiang Doctor provide professional health knowledge with enormous medical resources and are a promising information channel for future public health emergencies.

All these results suggest that information communication during a pandemic should be built on perceived credibility or trust. Aggregated social media usually provides various sources. Users can compare different sources for a news theme and select the most trustworthy news. In contrast, media with a single source delivers only one voice and has lower perceived credibility. This media will be abandoned in a competitive context. Additionally, public social media platforms are the most popular channels of interpersonal communication in China. These platforms are usually used among acquaintances with higher levels of trust. This shows that the credibility of the information source is important for news dissemination during a pandemic.



Governments should deliver more credible news and dispel rumors, which may be helpful in controlling the pandemic.

### **eHealth Literacy and Disease Knowledge as Predictors and Moderators of Preventive Behaviors**

Health literacy is being increasingly emphasized in public health-related studies. The relationship between health literacy and health behaviors or health status has also been highly recognized and understood based on empirical evidence. For example, it was found that poor health literacy created barriers to fully understanding individual health, illness, and treatment for people with HIV/AIDS [60]. Unimproved public mental health literacy predicted denial of self-help [61], and limited health literacy was correlated with worse health outcomes in terms of a patient's motivation, problem-solving ability, self-efficacy, and disease knowledge, among other factors [62].

However, prior studies mainly focused on chronic disease or unhealthy lifestyles. Less attention has been paid to public health emergencies such as pandemics. In this study, we investigated if and how health literacy influenced public preventive behaviors during the COVID-19 pandemic in China. Disease knowledge and eHealth literacy were selected as the core indicators of health literacy, as concluded from previous studies [63–66]. In line with most previous findings, we verified that both disease knowledge and eHealth literacy significantly predicted Chinese respondents' preventive behaviors during the COVID-19 pandemic. Additionally, eHealth literacy had more weight in predicting preventive behaviors than disease knowledge. Moreover, eHealth literacy positively moderated the relationship between social media use and preventive behaviors, while disease knowledge had a significant but negative effect. These findings highlight the importance of health literacy for pandemic prevention. Improving the public's level of health literacy is essential for health promotion, not only during a pandemic but in all contexts of public health in the future.

However, it should be mentioned that health literacy is not always positively correlated with preventive behaviors. Health literacy has shown inverse effects on individuals' healthy behaviors; for example, misinformation toward vaccination may lead to denial of the influenza vaccine [67], and a higher level of health literacy is not always associated with health-promotion behaviors [45]. This evidence underscored a compelling need to increase public awareness of health literacy in different disease conditions.

### **Demographics and Preventive Behaviors**

Many studies have indicated that sociodemographic indicators are vital in predicting health promotion behaviors. Our study showed similar outcomes to previous findings. We found that women engaged in more preventive behaviors than men during the COVID-19 pandemic in China. This finding may be explained by a study indicating that women are more sensitive to and interested in health information on social media than men [68]. Moreover, women usually have higher levels of disease knowledge and health literacy than men [69], and they search more frequently for health information on the internet related to changes in diet [70].

Furthermore, age, monthly income, and health status were positive predictors of preventive behaviors. These results indicate that people who are older and have higher income or good health status are more likely to take measures to prevent COVID-19, which is consistent with previous findings [67]. Additionally, education and marital status were significant predictors in the existing literature; for example, in one study [71], the odds of having accurate knowledge of malaria increased as individuals' educational levels increased, and unmarried people were found to be more likely to have positive attitudes toward rabies prevention than married people [41]. However, these variables were not significant in this study, perhaps due to the different social contexts.

### **Limitations**

The results of our study should be considered in light of several limitations, and the following improvements can be implemented in future studies:

Firstly, the sample consisted of netizens between 20 and 60 years of age. Younger people (age <20 years) and older people (age >60 years) had very low response rates in the survey database. Thus, we selected 20 to 60 years of age as the target age range of our sample. People younger than 20 years or older than 60 years could be included in future studies. Furthermore, the sample consisted of much more high-income and educated netizens because our sampling was proportioned according to gender and age without consideration of income and education. Future studies are suggested to comprise netizens with lower income and less education to facilitate the generalizability of our findings.

Secondly, a single measurement of disease knowledge was used in this study, which may have led to a ceiling effect on the respondents and impaired the validity of our test. Thus, a more suitable, reasonable, and valid instrument of disease knowledge should be constructed in future studies.

Finally, this article mainly focused on the frequency and types of social media use, while other variables of media use, such as motivations and content, were not included in this study. With the rapid development of various social media platforms, such as WeChat, Weibo, Facebook, Twitter, and WhatsApp, they will continue to play a vital role in public health promotion, as we found in this study. Future research is necessary to explore how social media access affects health behaviors, including the information sources and information content accessed. Also, the experience, needs, and motivations of one's social media use are suggested to be explored in health behavior studies in the future.

### **Conclusions**

Using a national web-based cross-sectional survey of a representative sample of Chinese netizens, we fully investigated our hypotheses and answered the proposed questions. We present our conclusions as follows: social media use frequency and disease knowledge and eHealth literacy were significant predictive factors of preventive behaviors; eHealth literacy and disease knowledge moderated the relationship between social media use and preventive behaviors. Aggregated social media use and public social media use were significant predictors of

preventive behaviors, while official social media use was not. These results not only enrich the theoretical paradigm of public health management and health communication but also have practical implications in pandemic control both for China and for other countries.

On one hand, the confirmed predictive ability of social media use suggests that social media is helpful to disseminate pandemic news and disease knowledge, which can help the public to collectively adopt necessary preventive measures for disease

control. On the other hand, the predictive ability of disease knowledge and eHealth literacy provided an endorsement that improving one's level of health literacy is essential during a pandemic in the long term. Additionally, sociodemographic factors such as gender, age, monthly income, and health status should be taken into account in public health interventions. More attention should perhaps be paid to the people who are male, are younger, have lower income, and have poor health status during a pandemic.

## Acknowledgments

The authors would like to thank WJX Company for their help with the data collection. The authors wish to thank Yuxin Han for her methodological and statistical help with the formal analysis. Also, we are grateful to the journal editors and the anonymous reviewers for their helpful comments and suggestions.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

The questionnaire used in the survey.

[DOCX File, 38 KB - [jmir\\_v22i10e19684\\_app1.docx](#)]

### Multimedia Appendix 2

Pearson correlation coefficients for sociodemographic variables, social media use, disease knowledge, eHealth literacy, and preventive behaviors.

[DOCX File, 29 KB - [jmir\\_v22i10e19684\\_app2.docx](#)]

## References

1. Coronavirus disease 2019 (COVID-19) Situation Report 133. World Health Organization. 2020 Jun 01. URL: [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200601-covid-19-sitrep-133.pdf?sfvrsn=9a56f2ac\\_4](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200601-covid-19-sitrep-133.pdf?sfvrsn=9a56f2ac_4) [accessed 2019-06-01]
2. Guarner J. Three Emerging Coronaviruses in Two Decades. *Am J Clin Pathol* 2020 Mar 09;153(4):420-421 [FREE Full text] [doi: [10.1093/ajcp/aqaa029](#)] [Medline: [32053148](#)]
3. Al-Surimi K, Khalifa M, Bahkali S, El-Metwally A, Househ M. The Potential of Social Media and Internet-Based Data in Preventing and Fighting Infectious Diseases: From Internet to Twitter. *Adv Exp Med Biol* 2017;972:131-139 [FREE Full text] [doi: [10.1007/5584\\_2016\\_132](#)] [Medline: [28004307](#)]
4. Velavan TP, Meyer CG. The COVID-19 epidemic. *Trop Med Int Health* 2020 Mar;25(3):278-280 [FREE Full text] [doi: [10.1111/tmi.13383](#)] [Medline: [32052514](#)]
5. Herrera-Diestra JL, Meyers LA. Local risk perception enhances epidemic control. *PLoS One* 2019 Dec 3;14(12):e0225576 [FREE Full text] [doi: [10.1371/journal.pone.0225576](#)] [Medline: [31794551](#)]
6. Kan J, Zhang H. Effects of awareness diffusion and self-initiated awareness behavior on epidemic spreading - An approach based on multiplex networks. *Commun Nonlinear Sci Numer Simul* 2017 Mar;44:193-203 [FREE Full text] [doi: [10.1016/j.cnsns.2016.08.007](#)] [Medline: [32288421](#)]
7. Erraguntla M, Zapletal J, Lawley M. Framework for Infectious Disease Analysis: A comprehensive and integrative multi-modeling approach to disease prediction and management. *Health Informatics J* 2019 Dec 27;25(4):1170-1187 [FREE Full text] [doi: [10.1177/1460458217747112](#)] [Medline: [29278956](#)]
8. van Velsen L, Beaujean DJ, van Gemert-Pijnen JE, van Steenberghe JE, Timen A. Public knowledge and preventive behavior during a large-scale Salmonella outbreak: results from an online survey in the Netherlands. *BMC Public Health* 2014 Jan 31;14:100 [FREE Full text] [doi: [10.1186/1471-2458-14-100](#)] [Medline: [24479614](#)]
9. Roberts H, Seymour B, Fish SA, Robinson E, Zuckerman E. Digital Health Communication and Global Public Influence: A Study of the Ebola Epidemic. *J Health Commun* 2017 Aug 30;22(sup1):51-58. [doi: [10.1080/10810730.2016.1209598](#)] [Medline: [28854128](#)]
10. The 44th Statistical Report on Internet Development in China. Webpage in Chinese. China Internet Network Information Center. 2019 Aug 30. URL: [http://www.cnnic.net.cn/hlwfzyj/hlwxbzg/hlwtjbg/201908/t20190830\\_70800.htm](http://www.cnnic.net.cn/hlwfzyj/hlwxbzg/hlwtjbg/201908/t20190830_70800.htm) [accessed 2020-03-20]

11. Costantino C, Restivo V, Ventura G, D'Angelo C, Randazzo M, Casuccio N, et al. Increased Vaccination Coverage among Adolescents and Young Adults in the District of Palermo as a Result of a Public Health Strategy to Counteract an 'Epidemic Panic'. *Int J Environ Res Public Health* 2018 May 17;15(5):1014 [FREE Full text] [doi: [10.3390/ijerph15051014](https://doi.org/10.3390/ijerph15051014)] [Medline: [29772848](https://pubmed.ncbi.nlm.nih.gov/29772848/)]
12. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. *Lancet* 2010 Oct;376(9748):1261-1271. [doi: [10.1016/s0140-6736\(10\)60809-4](https://doi.org/10.1016/s0140-6736(10)60809-4)]
13. Kim J, Jung M. Associations between media use and health information-seeking behavior on vaccinations in South Korea. *BMC Public Health* 2017 Sep 11;17(1):700 [FREE Full text] [doi: [10.1186/s12889-017-4721-x](https://doi.org/10.1186/s12889-017-4721-x)] [Medline: [28893212](https://pubmed.ncbi.nlm.nih.gov/28893212/)]
14. Alexander CC, Shrestha S, Tounkara MD, Cooper S, Hunt L, Hoj TH, et al. Media Access is Associated with Knowledge of Optimal Water, Sanitation and Hygiene Practices in Tanzania. *Int J Environ Res Public Health* 2019 Jun 03;16(11):1963 [FREE Full text] [doi: [10.3390/ijerph16111963](https://doi.org/10.3390/ijerph16111963)] [Medline: [31163573](https://pubmed.ncbi.nlm.nih.gov/31163573/)]
15. Bull SS, Levine DK, Black SR, Schmiede SJ, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. *Am J Prev Med* 2012 Nov;43(5):467-474 [FREE Full text] [doi: [10.1016/j.amepre.2012.07.022](https://doi.org/10.1016/j.amepre.2012.07.022)] [Medline: [23079168](https://pubmed.ncbi.nlm.nih.gov/23079168/)]
16. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for Behavior Change: Using Social Media for the Dissemination of Public Health Messages. *JMIR Public Health Surveill* 2017 Mar 23;3(1):e14 [FREE Full text] [doi: [10.2196/publichealth.6313](https://doi.org/10.2196/publichealth.6313)] [Medline: [28336503](https://pubmed.ncbi.nlm.nih.gov/28336503/)]
17. O'Leary ST, Narwaney KJ, Wagner NM, Kraus CR, Omer SB, Glanz JM. Efficacy of a Web-Based Intervention to Increase Uptake of Maternal Vaccines: An RCT. *Am J Prev Med* 2019 Oct;57(4):e125-e133. [doi: [10.1016/j.amepre.2019.05.018](https://doi.org/10.1016/j.amepre.2019.05.018)] [Medline: [31471001](https://pubmed.ncbi.nlm.nih.gov/31471001/)]
18. Vasconcelos Silva C, Jayasinghe D, Janda M. What Can Twitter Tell Us about Skin Cancer Communication and Prevention on Social Media? *Dermatology* 2020 Feb 25;236(2):81-89 [FREE Full text] [doi: [10.1159/000506458](https://doi.org/10.1159/000506458)] [Medline: [32097934](https://pubmed.ncbi.nlm.nih.gov/32097934/)]
19. Lwin M, Lu J, Sheldenkar A, Schulz P. Strategic Uses of Facebook in Zika Outbreak Communication: Implications for the Crisis and Emergency Risk Communication Model. *Int J Environ Res Public Health* 2018 Sep 10;15(9):1974 [FREE Full text] [doi: [10.3390/ijerph15091974](https://doi.org/10.3390/ijerph15091974)] [Medline: [30201929](https://pubmed.ncbi.nlm.nih.gov/30201929/)]
20. Yang Q, Wu S. How Social Media Exposure to Health Information Influences Chinese People's Health Protective Behavior during Air Pollution: A Theory of Planned Behavior Perspective. *Health Commun* 2019 Nov 24:1-10. [doi: [10.1080/10410236.2019.1692486](https://doi.org/10.1080/10410236.2019.1692486)] [Medline: [31760816](https://pubmed.ncbi.nlm.nih.gov/31760816/)]
21. Bandura A. Social Cognitive Theory of Mass Communication. *Media Psychology* 2001 Aug;3(3):265-299. [doi: [10.1207/s1532785xmep0303\\_03](https://doi.org/10.1207/s1532785xmep0303_03)]
22. Lim JS, Choe M, Zhang J, Noh G. The role of wishful identification, emotional engagement, and parasocial relationships in repeated viewing of live-streaming games: A social cognitive theory perspective. *Computers in Human Behavior* 2020 Jul;108:106327. [doi: [10.1016/j.chb.2020.106327](https://doi.org/10.1016/j.chb.2020.106327)]
23. Sarkar U, Le GM, Lyles CR, Ramo D, Linos E, Bibbins-Domingo K. Using Social Media to Target Cancer Prevention in Young Adults: Viewpoint. *J Med Internet Res* 2018 Jun 05;20(6):e203 [FREE Full text] [doi: [10.2196/jmir.8882](https://doi.org/10.2196/jmir.8882)] [Medline: [29871850](https://pubmed.ncbi.nlm.nih.gov/29871850/)]
24. Hansen AH, Claudi T, Årsand E. Associations Between the Use of eHealth and Out-of-Hours Services in People With Type 1 Diabetes: Cross-Sectional Study. *J Med Internet Res* 2019 Mar 21;21(3):e13465 [FREE Full text] [doi: [10.2196/13465](https://doi.org/10.2196/13465)] [Medline: [30896437](https://pubmed.ncbi.nlm.nih.gov/30896437/)]
25. Moon M, Shim J. Social media effects? *J Commun Manag* 2019 Nov 04;23(4):281-297. [doi: [10.1108/jcom-01-2019-0002](https://doi.org/10.1108/jcom-01-2019-0002)]
26. Li X, Zhang G. Perceived Credibility of Chinese Social Media: Toward an Integrated Approach. *IJPOR* 2017 Feb 06;30(1):79-101. [doi: [10.1093/ijpor/edw035](https://doi.org/10.1093/ijpor/edw035)]
27. Lin W, Zhang X, Cao B. How Do New Media Influence Youths' Health Literacy? Exploring the Effects of Media Channel and Content on Safer Sex Literacy. *Int J Sex Health* 2018 Dec 11;30(4):354-365. [doi: [10.1080/19317611.2018.1509921](https://doi.org/10.1080/19317611.2018.1509921)]
28. Park S, Avery EJ. Effects of Media Channel, Crisis Type and Demographics on Audience Intent to Follow Instructing Information During Crisis. *J Contingencies Crisis Man* 2016 Nov 10;26(1):69-78. [doi: [10.1111/1468-5973.12137](https://doi.org/10.1111/1468-5973.12137)]
29. Kim H, An B. A Study on the Effects of the attractiveness and credibility of Online 1 Personal Media Broadcasting B.J. on the Viewing Engagement perceived on Media Channel, Interactivity, Perceived Enjoyment, and the User's Responses. *Article in Korean. Advert Res* 2018 Sep 30;118:78-126. [doi: [10.16914/ar.2018.118.78](https://doi.org/10.16914/ar.2018.118.78)]
30. Alhuwail D, Abdulsalam Y. Assessing Electronic Health Literacy in the State of Kuwait: Survey of Internet Users From an Arab State. *J Med Internet Res* 2019 May 24;21(5):e11174 [FREE Full text] [doi: [10.2196/11174](https://doi.org/10.2196/11174)] [Medline: [31127723](https://pubmed.ncbi.nlm.nih.gov/31127723/)]
31. Stawarz K, Preist C, Coyle D. Use of Smartphone Apps, Social Media, and Web-Based Resources to Support Mental Health and Well-Being: Online Survey. *JMIR Ment Health* 2019 Jul 12;6(7):e12546 [FREE Full text] [doi: [10.2196/12546](https://doi.org/10.2196/12546)] [Medline: [31301126](https://pubmed.ncbi.nlm.nih.gov/31301126/)]
32. Kickbusch IS. Health literacy: addressing the health and education divide. *Health Promot Int* 2001 Sep;16(3):289-297. [doi: [10.1093/heapro/16.3.289](https://doi.org/10.1093/heapro/16.3.289)] [Medline: [11509466](https://pubmed.ncbi.nlm.nih.gov/11509466/)]
33. Institute of Medicine, Board on Neuroscience and Behavioral Health, Committee on Health Literacy. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. *Health Literacy: A Prescription to End Confusion*. Washington, DC: The National Academies Press; 2004.

34. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. *J Med Internet Res* 2006 Nov 14;8(4):e27 [FREE Full text] [doi: [10.2196/jmir.8.4.e27](https://doi.org/10.2196/jmir.8.4.e27)] [Medline: [17213046](https://pubmed.ncbi.nlm.nih.gov/17213046/)]
35. Tennant B, Stelfox M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. *J Med Internet Res* 2015 Mar 17;17(3):e70 [FREE Full text] [doi: [10.2196/jmir.3992](https://doi.org/10.2196/jmir.3992)] [Medline: [25783036](https://pubmed.ncbi.nlm.nih.gov/25783036/)]
36. Kim KA, Kim YJ, Choi M. Association of Electronic Health Literacy With Health-Promoting Behaviors in Patients With Type 2 Diabetes. *CIN: Computers, Informatics, Nursing* 2018;36(9):438-447. [doi: [10.1097/cin.0000000000000438](https://doi.org/10.1097/cin.0000000000000438)]
37. Stelfox M, Paige SR, Alber JM, Chaney BH, Chaney D, Apperson A, et al. Association Between Health Literacy, Electronic Health Literacy, Disease-Specific Knowledge, and Health-Related Quality of Life Among Adults With Chronic Obstructive Pulmonary Disease: Cross-Sectional Study. *J Med Internet Res* 2019 Jun 06;21(6):e12165 [FREE Full text] [doi: [10.2196/12165](https://doi.org/10.2196/12165)] [Medline: [31172962](https://pubmed.ncbi.nlm.nih.gov/31172962/)]
38. Yang SC, Luo YF, Chiang C. Electronic Health Literacy and Dietary Behaviors in Taiwanese College Students: Cross-Sectional Study. *J Med Internet Res* 2019 Nov 26;21(11):e13140 [FREE Full text] [doi: [10.2196/13140](https://doi.org/10.2196/13140)] [Medline: [31769760](https://pubmed.ncbi.nlm.nih.gov/31769760/)]
39. Burns JR, Rapee RM. Adolescent mental health literacy: young people's knowledge of depression and help seeking. *J Adolesc* 2006 Apr;29(2):225-239. [doi: [10.1016/j.adolescence.2005.05.004](https://doi.org/10.1016/j.adolescence.2005.05.004)] [Medline: [15996727](https://pubmed.ncbi.nlm.nih.gov/15996727/)]
40. Kim S, Love F, Quistberg DA, Shea JA. Association of health literacy with self-management behavior in patients with diabetes. *Diabetes Care* 2004 Dec;27(12):2980-2982. [doi: [10.2337/diacare.27.12.2980](https://doi.org/10.2337/diacare.27.12.2980)] [Medline: [15562219](https://pubmed.ncbi.nlm.nih.gov/15562219/)]
41. Edukugho AA, Umoh JU, Diem M, Ajani G, Uba B, Okeke L, et al. Knowledge, attitudes and practices towards rabies prevention among residents of Abuja municipal area council, Federal Capital Territory, Nigeria. *Pan Afr Med J* 2018;31:21 [FREE Full text] [doi: [10.11604/pamj.2018.31.21.15120](https://doi.org/10.11604/pamj.2018.31.21.15120)] [Medline: [30918548](https://pubmed.ncbi.nlm.nih.gov/30918548/)]
42. Gajda M, Kowalska M. The Web-Based Randomized Controlled Intervention as the Enhancer of Cancer Prevention. *Medicina (Kaunas)* 2019 Aug 03;55(8):434 [FREE Full text] [doi: [10.3390/medicina55080434](https://doi.org/10.3390/medicina55080434)] [Medline: [31382569](https://pubmed.ncbi.nlm.nih.gov/31382569/)]
43. Cho YI, Lee SD, Arozullah AM, Crittenden KS. Effects of health literacy on health status and health service utilization amongst the elderly. *Soc Sci Med* 2008 Apr;66(8):1809-1816. [doi: [10.1016/j.socscimed.2008.01.003](https://doi.org/10.1016/j.socscimed.2008.01.003)] [Medline: [18295949](https://pubmed.ncbi.nlm.nih.gov/18295949/)]
44. Gazmararian JA, Williams MV, Peel J, Baker DW. Health literacy and knowledge of chronic disease. *Patient Education and Counseling* 2003 Nov;51(3):267-275. [doi: [10.1016/s0738-3991\(02\)00239-2](https://doi.org/10.1016/s0738-3991(02)00239-2)]
45. Bains SS, Egede LE. Associations between health literacy, diabetes knowledge, self-care behaviors, and glycemic control in a low income population with type 2 diabetes. *Diabetes Technol Ther* 2011 Mar;13(3):335-341 [FREE Full text] [doi: [10.1089/dia.2010.0160](https://doi.org/10.1089/dia.2010.0160)] [Medline: [21299402](https://pubmed.ncbi.nlm.nih.gov/21299402/)]
46. Questionnaire Star. Webpage in Chinese. URL: <https://www.wjx.cn/> [accessed 2020-10-06]
47. Prevention of Coronavirus Disease 2019. Government of the Hong Kong Special Administrative Region. URL: [https://www.chp.gov.hk/files/pdf/prevention\\_of\\_covid\\_19\\_en.pdf](https://www.chp.gov.hk/files/pdf/prevention_of_covid_19_en.pdf) [accessed 2020-01-07]
48. Protocol for Prevention and Control of COVID-19. Chinese Center for Disease Control and Prevention. URL: <http://weekly.chinacdc.cn/en/article/doi/10.46234/ccdcw2020.082> [accessed 2020-10-07]
49. Whitehead D. Using mass media within health-promoting practice: a nursing perspective. *J Adv Nurs* 2000 Oct;32(4):807-816. [doi: [10.1046/j.1365-2648.2000.01544.x](https://doi.org/10.1046/j.1365-2648.2000.01544.x)] [Medline: [11095218](https://pubmed.ncbi.nlm.nih.gov/11095218/)]
50. Huhman M, Potter LD, Wong FL, Banspach SW, Duke JC, Heitzler CD. Effects of a mass media campaign to increase physical activity among children: year-1 results of the VERB campaign. *Pediatrics* 2005 Aug;116(2):e277-e284. [doi: [10.1542/peds.2005-0043](https://doi.org/10.1542/peds.2005-0043)] [Medline: [16061581](https://pubmed.ncbi.nlm.nih.gov/16061581/)]
51. Hökby S, Hadlaczy G, Westerlund J, Wasserman D, Balazs J, Germanavicius A, et al. Are Mental Health Effects of Internet Use Attributable to the Web-Based Content or Perceived Consequences of Usage? A Longitudinal Study of European Adolescents. *JMIR Ment Health* 2016 Jul 13;3(3):e31 [FREE Full text] [doi: [10.2196/mental.5925](https://doi.org/10.2196/mental.5925)] [Medline: [27417665](https://pubmed.ncbi.nlm.nih.gov/27417665/)]
52. Cho H, Li W, Shen L, Cannon J. Mechanisms of Social Media Effects on Attitudes Toward E-Cigarette Use: Motivations, Mediators, and Moderators in a National Survey of Adolescents. *J Med Internet Res* 2019 Jun 27;21(6):e14303 [FREE Full text] [doi: [10.2196/14303](https://doi.org/10.2196/14303)] [Medline: [31250830](https://pubmed.ncbi.nlm.nih.gov/31250830/)]
53. Acun I. The Relationship among University Students' Trust, Self-Esteem, Satisfaction with Life and Social Media Use. *Int J Instruction* 2020 Jan 03;13(1):35-52. [doi: [10.29333/iji.2020.1313a](https://doi.org/10.29333/iji.2020.1313a)]
54. Dellarocas C, Sutanto J, Calin M, Palme E. Attention Allocation in Information-Rich Environments: The Case of News Aggregators. *Management Science* 2016 Sep;62(9):2543-2562. [doi: [10.1287/mnsc.2015.2237](https://doi.org/10.1287/mnsc.2015.2237)]
55. Calzada J, Gil R. What Do News Aggregators Do? Evidence from Google News in Spain and Germany. *Marketing Science* 2020 Jan;39(1):134-167. [doi: [10.1287/mksc.2019.1150](https://doi.org/10.1287/mksc.2019.1150)]
56. Repnikova M, Fang K. Digital Media Experiments in China: "Revolutionizing" Persuasion under Xi Jinping. *The China Quarterly* 2019 Apr 8;239:679-701. [doi: [10.1017/s0305741019000316](https://doi.org/10.1017/s0305741019000316)]
57. Deng J. Janus: How exceptionalism based on regaining influence and doing new media help a Chinese mobile news app negotiate censorship for better journalism. *Communication and the Public* 2018 Apr 09;3(2):113-133. [doi: [10.1177/2057047318770466](https://doi.org/10.1177/2057047318770466)]



58. Jang K, Baek YM. When Information from Public Health Officials is Untrustworthy: The Use of Online News, Interpersonal Networks, and Social Media during the MERS Outbreak in South Korea. *Health Commun* 2019 Aug;34(9):991-998. [doi: [10.1080/10410236.2018.1449552](https://doi.org/10.1080/10410236.2018.1449552)] [Medline: [29558170](https://pubmed.ncbi.nlm.nih.gov/29558170/)]
59. Calder K, D'Aeth L, Turner S, Begg A, Veer E, Scott J, et al. Evaluation of the All Right? Campaign's Facebook intervention post-disaster in Canterbury, New Zealand. *Health Promot Int* 2020 Feb 01;35(1):111-122. [doi: [10.1093/heapro/day106](https://doi.org/10.1093/heapro/day106)] [Medline: [30601952](https://pubmed.ncbi.nlm.nih.gov/30601952/)]
60. Kalichman SC, Benotsch E, Suarez T, Catz S, Miller J, Rompa D. Health literacy and health-related knowledge among persons living with HIV/AIDS. *American Journal of Preventive Medicine* 2000 May;18(4):325-331. [doi: [10.1016/s0749-3797\(00\)00121-5](https://doi.org/10.1016/s0749-3797(00)00121-5)]
61. Jorm AF. Mental health literacy. Public knowledge and beliefs about mental disorders. *Br J Psychiatry* 2000 Nov;177:396-401. [doi: [10.1192/bjp.177.5.396](https://doi.org/10.1192/bjp.177.5.396)] [Medline: [11059991](https://pubmed.ncbi.nlm.nih.gov/11059991/)]
62. Morris NS, Field TS, Wagner JL, Cutrona SL, Roblin DW, Gaglio B, et al. The association between health literacy and cancer-related attitudes, behaviors, and knowledge. *J Health Commun* 2013 Dec 04;18 Suppl 1(sup1):223-241 [FREE Full text] [doi: [10.1080/10810730.2013.825667](https://doi.org/10.1080/10810730.2013.825667)] [Medline: [24093358](https://pubmed.ncbi.nlm.nih.gov/24093358/)]
63. Mueller J, Davies A, Jay C, Harper S, Blackhall F, Summers Y, et al. Developing and testing a web-based intervention to encourage early help-seeking in people with symptoms associated with lung cancer. *Br J Health Psychol* 2019 Feb;24(1):31-65 [FREE Full text] [doi: [10.1111/bjhp.12325](https://doi.org/10.1111/bjhp.12325)] [Medline: [29999214](https://pubmed.ncbi.nlm.nih.gov/29999214/)]
64. Alsaifi A, Cheng A. Knowledge, Attitudes and Behaviours of Healthcare Workers in the Kingdom of Saudi Arabia to MERS Coronavirus and Other Emerging Infectious Diseases. *Int J Environ Res Public Health* 2016 Dec 06;13(12):1214 [FREE Full text] [doi: [10.3390/ijerph13121214](https://doi.org/10.3390/ijerph13121214)] [Medline: [27929452](https://pubmed.ncbi.nlm.nih.gov/27929452/)]
65. Ashida S, Goodman M, Pandya C, Koehly L, Lachance C, Stafford J, et al. Age differences in genetic knowledge, health literacy and causal beliefs for health conditions. *Public Health Genomics* 2011;14(4-5):307-316 [FREE Full text] [doi: [10.1159/000316234](https://doi.org/10.1159/000316234)] [Medline: [20829577](https://pubmed.ncbi.nlm.nih.gov/20829577/)]
66. Lindau ST, Tomori C, Lyons T, Langseth L, Bennett CL, Garcia P. The association of health literacy with cervical cancer prevention knowledge and health behaviors in a multiethnic cohort of women. *Am J Obstet Gynecol* 2002 May;186(5):938-943. [doi: [10.1067/mob.2002.122091](https://doi.org/10.1067/mob.2002.122091)] [Medline: [12015518](https://pubmed.ncbi.nlm.nih.gov/12015518/)]
67. El Khoury G, Salameh P. Influenza Vaccination: A Cross-Sectional Survey of Knowledge, Attitude and Practices among the Lebanese Adult Population. *Int J Environ Res Public Health* 2015 Dec 05;12(12):15486-15497 [FREE Full text] [doi: [10.3390/ijerph121215000](https://doi.org/10.3390/ijerph121215000)] [Medline: [26690189](https://pubmed.ncbi.nlm.nih.gov/26690189/)]
68. Huesch MD, Galstyan A, Ong MK, Doctor JN. Using Social Media, Online Social Networks, and Internet Search as Platforms for Public Health Interventions: A Pilot Study. *Health Serv Res* 2016 Jun 10;51 Suppl 2:1273-1290 [FREE Full text] [doi: [10.1111/1475-6773.12496](https://doi.org/10.1111/1475-6773.12496)] [Medline: [27161093](https://pubmed.ncbi.nlm.nih.gov/27161093/)]
69. Mouchtouri V, Papagiannis D, Katsioulis A, Rachiotis G, Dafopoulos K, Hadjichristodoulou C. Knowledge, Attitudes, and Practices about the Prevention of Mosquito Bites and Zika Virus Disease in Pregnant Women in Greece. *Int J Environ Res Public Health* 2017 Mar 31;14(4):367 [FREE Full text] [doi: [10.3390/ijerph14040367](https://doi.org/10.3390/ijerph14040367)] [Medline: [28362340](https://pubmed.ncbi.nlm.nih.gov/28362340/)]
70. Bujnowska-Fedak MM, Węgierek P. The Impact of Online Health Information on Patient Health Behaviours and Making Decisions Concerning Health. *Int J Environ Res Public Health* 2020 Jan 31;17(3):880 [FREE Full text] [doi: [10.3390/ijerph17030880](https://doi.org/10.3390/ijerph17030880)] [Medline: [32023828](https://pubmed.ncbi.nlm.nih.gov/32023828/)]
71. Yaya S, Bishwajit G, Ekholuenetale M, Shah V, Kadio B, Udenigwe O. Knowledge of prevention, cause, symptom and practices of malaria among women in Burkina Faso. *PLoS One* 2017;12(7):e0180508 [FREE Full text] [doi: [10.1371/journal.pone.0180508](https://doi.org/10.1371/journal.pone.0180508)] [Medline: [28671987](https://pubmed.ncbi.nlm.nih.gov/28671987/)]

## Abbreviations

**CCTV:** China Central Television  
**CNNIC:** China Internet Network Information Center  
**eHEALS:** eHealth Literacy Scale  
**MERS:** Middle East respiratory syndrome  
**MERS-CoV:** Middle East respiratory syndrome coronavirus  
**SARS:** severe acute respiratory syndrome  
**SARS-CoV:** severe acute respiratory syndrome coronavirus  
**SRIDC:** Statistical Report on Internet Development of China



*Edited by G Eysenbach; submitted 28.04.20; peer-reviewed by H Xu, C Mather, JR Bautista, Z Ma; comments to author 20.05.20; revised version received 10.06.20; accepted 23.09.20; published 09.10.20.*

*Please cite as:*

*Li X, Liu Q*

*Social Media Use, eHealth Literacy, Disease Knowledge, and Preventive Behaviors in the COVID-19 Pandemic: Cross-Sectional Study on Chinese Netizens*

*J Med Internet Res 2020;22(10):e19684*

URL: <http://www.jmir.org/2020/10/e19684/>

doi: [10.2196/19684](https://doi.org/10.2196/19684)

PMID: [33006940](https://pubmed.ncbi.nlm.nih.gov/33006940/)

©Xiaojing Li, Qinliang Liu. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 09.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

## Epidemiological Parameters of COVID-19: Case Series Study

Shujuan Ma<sup>1\*</sup>, PhD; Jiayue Zhang<sup>1\*</sup>, MPhil; Minyan Zeng<sup>2\*</sup>, MSc; Qingping Yun<sup>3\*</sup>, BSc; Wei Guo<sup>1\*</sup>, MPhil; Yixiang Zheng<sup>4\*</sup>, MD; Shi Zhao<sup>5</sup>, MPhil; Maggie H Wang<sup>5</sup>, PhD; Zuyao Yang<sup>5</sup>, PhD

<sup>1</sup>Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

<sup>2</sup>Peking University Shenzhen Hospital, Shenzhen, China

<sup>3</sup>Department of Social Medicine and Health Education, School of Public Health, Peking University, Beijing, China

<sup>4</sup>Xiangya Hospital, Central South University, Changsha, China

<sup>5</sup>JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong

\*these authors contributed equally

**Corresponding Author:**

Zuyao Yang, PhD

JC School of Public Health and Primary Care

The Chinese University of Hong Kong

Shatin, New Territories

Hong Kong

Phone: 852 2252 8739

Email: [zyang@cuhk.edu.hk](mailto:zyang@cuhk.edu.hk)

**Abstract**

**Background:** The estimates of several key epidemiological parameters of the COVID-19 pandemic are often based on small sample sizes or are inaccurate for various reasons.

**Objective:** The aim of this study is to obtain more robust estimates of the incubation period, serial interval, frequency of presymptomatic transmission, and basic reproduction number ( $R_0$ ) of COVID-19 based on a large case series.

**Methods:** We systematically retrieved and screened 20,658 reports of laboratory-confirmed COVID-19 cases released by the health authorities of China, Japan, and Singapore. In addition, 9942 publications were retrieved from PubMed and China National Knowledge Infrastructure (CNKI) through April 8, 2020. To be eligible, a report had to contain individual data that allowed for accurate estimation of at least one parameter. Widely used models such as gamma distributions were fitted to the data sets and the results with the best-fitting values were presented.

**Results:** In total, 1591 cases were included for the final analysis. The mean incubation period ( $n=687$ ) and mean serial interval ( $n=1015$  pairs) were estimated to be 7.04 (SD 4.27) days and 6.49 (SD 4.90) days, respectively. In 40 cases (5.82%), the incubation period was longer than 14 days. In 32 infector-infectee pairs (3.15%), infectees' symptom onsets occurred before those of infectors. Presymptomatic transmission occurred in 129 of 296 infector-infectee pairs (43.58%).  $R_0$  was estimated to be 1.85 (95% CI 1.37-2.60).

**Conclusions:** This study provides robust estimates of several epidemiological parameters of COVID-19. The findings support the current practice of 14-day quarantine of persons with potential exposure, but also suggest the need for additional measures. Presymptomatic transmission together with the asymptomatic transmission reported by previous studies highlight the importance of adequate testing, strict quarantine, and social distancing.

(*J Med Internet Res* 2020;22(10):e19994) doi:[10.2196/19994](https://doi.org/10.2196/19994)

**KEYWORDS**

coronavirus disease 2019; COVID-19; incubation period; serial interval; basic reproduction number; presymptomatic transmission

**Introduction**

In December 2019, a novel enveloped ribonucleic acid (RNA) beta-coronavirus, which was later named SARS-CoV-2, was first reported in Wuhan, the capital city of Hubei province,

China [1]. COVID-19, the disease caused by SARS-CoV-2, spread across and outside China rapidly, and was declared a pandemic by the World Health Organization on March 11, 2020.

Despite the explosive growth of the number of studies on COVID-19 [2-4], several key epidemiological parameters of

the disease remain to be clarified, among which are incubation period and serial interval. The mean or median incubation period and serial interval estimated by previous studies were mostly 4-5 days [5-16]. However, many of the studies included a limited number of cases (around or less than 100) [5,7,9,12-16]. For example, in a study by Zhang et al [16], which included 8579 cases, only 49 cases and 39 pairs could be used for estimating incubation period and serial interval, respectively. On the other hand, some studies might have been limited by the inaccuracy of the original data. For example, if the interval of exposure was long or unclear, determining the exact exposure date would be difficult [6,8,17], which could give rise to error.

Another important parameter of transmission dynamics is the basic reproduction number ( $R_0$ ), which is defined as the average number of secondary cases caused by a single infectious individual in a totally susceptible population [18]. As  $R_0$  is often estimated based on the serial interval [19], the abovementioned issues affecting previous estimates of serial interval might have affected the estimates of  $R_0$  as well. On the other hand, some studies estimated  $R_0$  based on the latent period and infectious period [19,20], with latent period approximated by incubation period. This may not be appropriate for COVID-19 as presymptomatic transmission might occur [21,22]. However, few published studies [23] estimated how often and approximately when the disease could be transmitted prior to symptom onset.

This study made use of the large amount of data reported by the health authorities of and outside China, as well as data from studies published as of April 8, 2020, to address the above issues. Specifically, we aimed to obtain more robust estimates of the following epidemiological parameters of COVID-19: (1) incubation period, defined as the time interval between exposure and onset of disease symptoms; (2) serial interval, defined as the duration between symptom onset of an infector (eg, a primary case) and that of an infectee (eg, a secondary case) in a transmission chain, with a negative value meaning that the infectee's symptoms occurred before the infector's symptoms; (3) frequency of presymptomatic transmission; and (4)  $R_0$ .

## Methods

### Data Sources

The details of data sources can be found in [Multimedia Appendix 1](#). For China, all provinces, autonomous regions, and municipalities (including mainland China, Hong Kong, Macau, and Taiwan) that had reported cases of COVID-19 were identified according to the daily updates by the National Health Commission of China [24]. Subsequently, the official websites and WeChat accounts (if any) of local governments and health authorities (eg, Municipal Health Commission, Center for Disease Control and Prevention, Department of Health) were checked manually through April 8, 2020, to identify and download the reports on laboratory-confirmed cases of COVID-19. Google and Baidu were searched to identify public media reports written based on official press releases. Chinese words for the following terms were used to perform the search:

("family" OR "household") AND ("cluster" OR "dinner" OR "party") AND "infection."

Other countries reporting COVID-19 cases were identified according to the COVID-19 situation reports published by the World Health Organization [25] and the data were searched through April 8, 2020. For Japan and Singapore, information regarding confirmed cases was retrieved from their respective Ministry of Health. We also searched for individual cases from relevant departments and public media of the United States, the United Kingdom, Canada, and Australia, but failed to find any with details allowing for parameters estimation for this study. Typically, the dates of exposure and symptom onset were lacking (eg, see reference [26]). Owing to the language barrier, we did not do a comprehensive search for other countries.

PubMed was searched to identify relevant publications by using the following terms: "coronavirus," "2019-nCov," "SARS-CoV-2," and "COVID-19." The China National Knowledge Infrastructure (CNKI) was searched to identify publications in Chinese journals using "novel coronavirus" ("Xin Xing Guan Zhuang Bing Du" in Chinese pinyin), which is the official Chinese name for SARS-CoV-2. Both databases were searched from December 2019 through April 8, 2020. The reference lists of eligible publications were also checked to see if there were other eligible studies not found by previous searches. We also kept an eye on the COVID-19 studies disseminated by public media and the official WeChat accounts of various academic entities in China and those recommended by experts in this field to the authors of this study.

### Definitions and Inclusion Criteria

To be eligible, a report had to contain individual data that allowed for estimation of at least one of the following parameters of laboratory-confirmed cases of COVID-19: incubation period, serial interval, and symptoms-to-transmission time, which was defined as the day of an infectee's contact with the infector relative to the latter's symptom onset date, with a negative value meaning that the transmission occurred before the infector developed symptoms (ie, presymptomatic transmission).

To obtain an accurate estimate of incubation period, only the cases with an exposure period spanning 3 days or less were included in the analysis. For those exposed for three continuous days and those exposed on two dates with one day apart (ie, exposed on the first and third days), the second day was uniformly used as the exposure date in estimations. For those exposed for two continuous days, the first day was uniformly used as the exposure date in estimations. This approach ensured the upper limit of error in the estimated incubation period be smaller than 1 day for the cases with a 2- or 3-day exposure, regardless of when exactly (ie, first, second, or third day) the transmission occurred. The actual overall error was bound to be much smaller than 1 day, as most included cases were exposed for only 1 day, which would dilute the error caused by 2- or 3-day exposures.

Serial interval was estimated based on the symptom onset dates of infector-infectee pairs, which were typically from cluster infections. For two cases to qualify as an infector-infectee pair

and be included in this study, the following two criteria must both be fulfilled. First, there must be evidence that the presumed infector had been exposed outside the cluster (eg, close contact with a confirmed case, travel history to Hubei, exposure to a person who returned from Hubei) before he/she attended the group activities (eg, family gathering, business conference) that led to cluster infections. Second, in the 14 days prior to symptom onset, the presumed infectee was exposed to the presumed infector only, without other exposure histories.

The estimation of symptoms-to-transmission time also involved determination of exposure date and judgement about transmission chain, hence the above principles applied in estimating incubation period and serial interval were followed as well.

### Screening, Data Extraction, and Quality Control

In total, 6 researchers were involved in data collection. The reports retrieved through the above searches were scrutinized one by one according to the inclusion criteria specified above. The following data were extracted from eligible reports by using a standard extraction form which was pilot tested with the reports from Liaoning province of China: the geographical location concerned, age, sex, type of exposure, first date and period (if applicable) of exposure, date of symptom onset, initial symptoms, and whether the case was from a cluster. For a clustering case, the generation he/she belonged to, the exposure date, and symptom onset date were also recorded. The retrieved reports were split into six parts, with each researcher responsible for one part. For each part, the eligibility of and data extracted from reports were first determined by one researcher and then cross-checked by another three researchers. All uncertainty and disagreements were discussed among the researchers. If no consensus could be reached after discussion, the concerned cases would be excluded to ensure the correctness and accuracy of data. For example, if it could not be determined who was infected through attending a group activity (ie, infectee) and who had been infected before he/she attended the activity (ie, infector), then all clustering cases related to the activity had to be excluded.

### Data Analysis

The basic characteristics of included cases were summarized descriptively. Lognormal, Weibull, and gamma distributions were fitted to the data sets of incubation period and serial interval and the one with the smallest Akaike information criterion (AIC) score was used for the final analysis. For serial interval, which had negative values, shifted distributions were fitted, with the best-fitting values determined by maximum

likelihood. The key parameters were estimated by using the maximum likelihood approach. The data on symptoms-to-transmission time was roughly symmetrical according to visual inspection and thus fitted by normal distribution. For each parameter, the range, median, selected percentiles, mean, and standard deviation were estimated. The 95% confidence intervals of mean and standard deviation were estimated by using the bootstrap technique.  $R_0$  was estimated by the well-studied Euler-Lotka equation [19]:  $R_0 = 1/M[-r/h(\cdot)]$ , where  $r$  is the exponential growth rate,  $h(\cdot)$  is the estimated distribution of the generation time, and the function  $M[\cdot]$  is the Laplace transform (ie, moment generating function known for statisticians) for  $h$ . As generation time is hard to observe directly, it was approximated by serial interval in this study. Compared with generation time, serial interval is expected to have the same mean but a larger variance due to possibly differing incubation periods of infectors and infectees. The exponential growth rate was obtained directly from a previous study reporting incidence data of the early stage of the epidemic [1], while the other parameters came from this study.

Sensitivity analyses were conducted to examine the robustness of the two parameters involving exposure date. Specifically, the third day for those whose exposure period spanned three days and the second day for those exposed for two continuous days were used as their exposure dates in sensitivity analyses. The difference between the estimates from sensitivity analysis and those from primary analysis represents the largest possible error in the latter. All statistical analyses were conducted with SAS software (Version 9.4, SAS Institute).

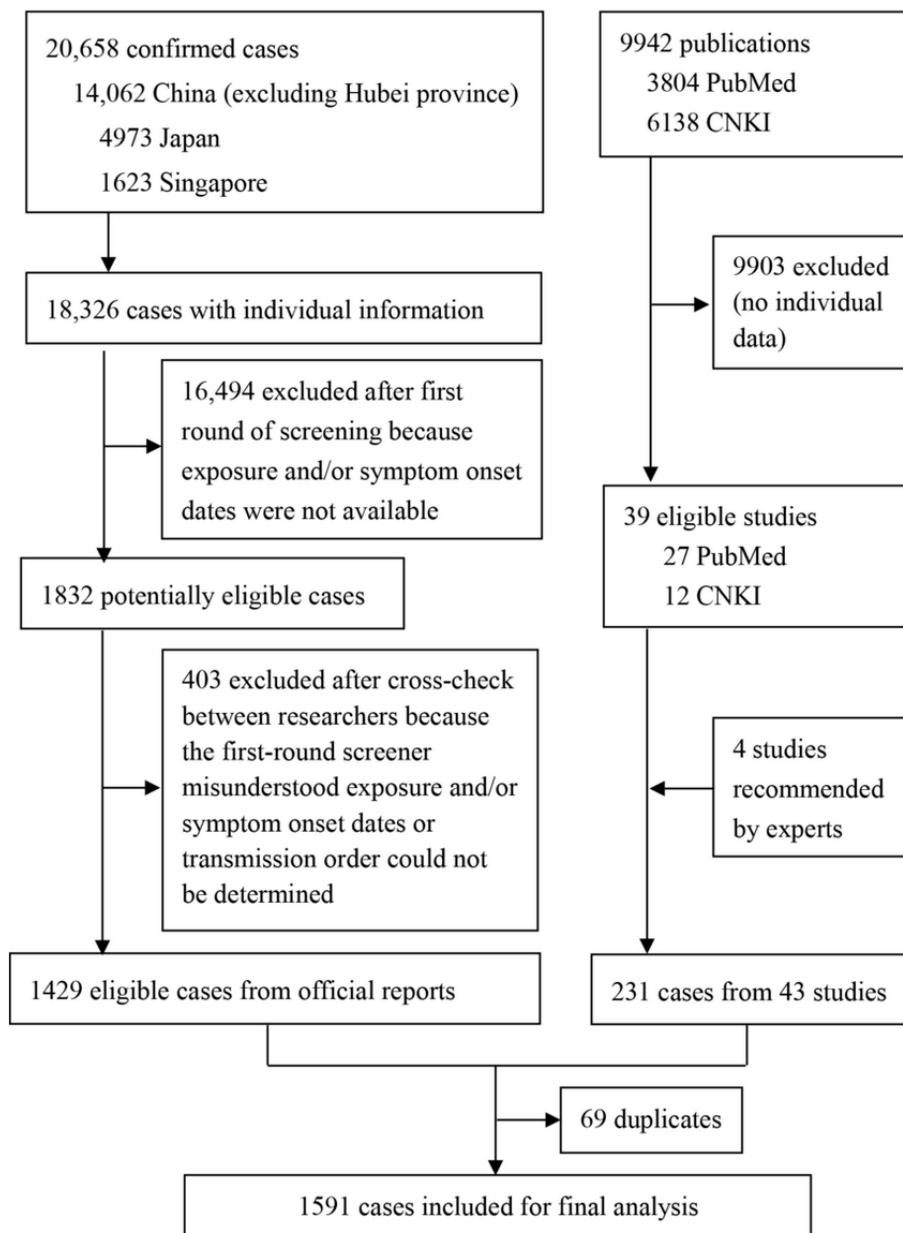
### Ethical Statement

No ethical approval was needed as the data used in this study was publicly available, either released by health authorities or retrieved from electronic databases.

## Results

### Overview

Figure 1 shows the process of data collection. We systematically retrieved and screened 20,658 case reports and 9942 publications. The 43 publications included in our analysis can be found in [Multimedia Appendix 2](#). In total, 1591 cases were included for the final analysis, including 1211 (76.12%) from China, 301 (18.92%) from Japan, 60 (3.77%) from Singapore, and 19 from five other countries. The cases from China covered 156 cities in 31 provinces of 34 in total.

**Figure 1.** The flowchart of cases selection from the official reports and published studies. CNKI: China National Knowledge Infrastructure.

### Characteristics of Included Cases

The age of included cases ranged from 5 days to 95 years (Table 1). There were 133 cases (10.74%) with a travel history to Hubei province, where Wuhan is the capital city. The source of exposure of 1302 cases (81.99%) was known confirmed cases. The exposure period was exactly 1 day in 508 cases (53.08%), spanned 2-3 days in 227 cases (23.72%), and spanned more

than 3 days in the other cases (which were not included in the estimation of parameters involving exposure date). The top 5 initial symptoms were fever (730/1009, 72.35%), cough (257/1009, 25.47%), fatigue (69/1009, 6.84%), sore throat (57/1009, 5.65%), and malaise (54/1009, 5.35%). Clustering cases accounted for 84.92% (1351/1591) of all, and most of them were second generation.



**Table 1.** Characteristics of included cases.

Characteristics	Value
Age range	5 days to 95 years
Mean age, years (SD)	45.90 (18.02)
<b>Age groups (n=1437), years, n (%)</b>	
0-18	89 (6.19)
19-64	1121 (78.01)
≥65	227 (15.80)
<b>Sex (n= 1559 ), n (%)</b>	
Male	768 (49.26)
Female	791 (50.74)
<b>Travel history to Hubei (n=1238, n (%))</b>	
Yes	133 (10.74)
No	1105 (89.26)
<b>Infectior was a known confirmed case (n= 1588 ), n (%)</b>	
Yes	1302 (81.99)
No	286 (18.01)
<b>Span of exposure period (n=957), n (%)</b>	
1 day	508 (53.08)
2-3 days	227 (23.72)
>3 days	222 (23.20)
<b>Top 5 initial symptoms (n=1009), n (%)</b>	
Fever	730 (72.35)
Cough	257 (25.47)
Fatigue	69 (6.84)
Sore throat	57 (5.65)
Malaise	54 (5.35)
<b>Clustering cases (n=1591), n (%)</b>	
Yes	1351 (84.92)
No	240 (15.08)
<b>Family cluster (n=1238), n (%)</b>	
Yes	514 (41.52)
No	724 (58.48)
<b>Generation of clustering cases (n=1350), n (%)</b>	
First	61 (4.52)
Second	1069 (79.19)
Third	188 (13.93)
Fourth or higher	32 (2.37)

## Epidemiological Parameters

Incubation period was estimated from 687 cases, of whom 482 (69.55%) were exposed for 1 day; the remainder were exposed for 2 or 3 days. The incubation periods of individual cases ranged from 0 to 23 days (Table 2), with a median of 6 days; in total, 5.82% (n=40) of incubation periods were longer than

14 days. The incubation period is best described by a gamma distribution (Multimedia Appendix 3, Figure 2A), with a mean of 7.04 days (95% CI 6.74-7.33) and standard deviation of 4.27 days (95% CI 3.92-4.43).

The serial interval was estimated for 1015 infector-infectee pairs and ranged from -5 to 29 days (Table 2), with a median of 6

days. For 32 pairs (3.15%), the serial interval was less than 0; for these 32 pairs, the mean was  $-2.50$  days, meaning that the infectee showed symptoms 2.50 days earlier than the infector. The serial interval is best described by a shifted gamma distribution (Multimedia Appendix 3, Figure 2B), with a mean of 6.49 days (95% CI 6.15-6.80) and standard deviation of 4.90 days (95% CI 4.68-5.25). Based on an exponential growth rate of 0.10 per day [1] and the mean and standard deviation of the serial interval,  $R_0$  was estimated at 1.85 (95% CI 1.37-2.60). The country-specific  $R_0$  for China, Japan, and Singapore was 1.83 (95% CI 1.37-2.54), 1.83 (95% CI 1.37-2.53), and 1.98 (95% CI 1.41-2.97), respectively.

The symptoms-to-transmission time was estimated from 296 infector-infectee pairs and ranged from  $-9$  to 14 days (Table 2), with a median of 0 day. Presymptomatic transmission

occurred in 129 infector-infectee pairs (43.58%, 95% CI 37.93%-49.23%); for these 129 pairs, the mean symptoms-to-transmission time was  $-2.88$  days, meaning that the transmission occurred 2.88 days before the infectors showed symptoms. Fitted by a normal distribution (Figure 2C), the mean of this parameter was  $-0.07$  days (95% CI  $-0.43$  to 0.31) and the standard deviation was 3.31 days (95% CI 2.96 to 3.66).

The symptoms-to-transmission time can also be inferred by mean serial interval minus mean incubation time ( $-0.55$  days), which is consistent with the direct estimate ( $-0.07$  days) as both suggest the mean symptoms-to-transmission time to be around the day before primary cases' symptom onset. Based on the above estimates, the timeline of infection for an "average" infector-infectee pair in a transmission chain is demonstrated in Multimedia Appendix 4.

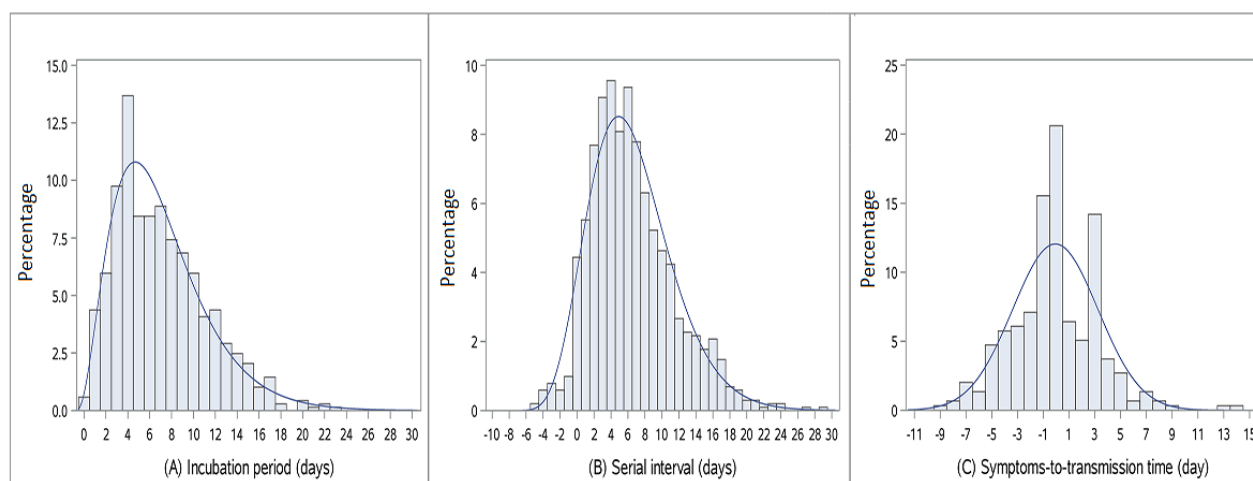
**Table 2.** The estimates of incubation period, serial interval, and symptoms-to-transmission time (unit: day).

Metrics	Incubation period	Serial interval	Symptoms-to-transmission time
Range	0 to 23.00	$-5$ to 29.00	$-9$ to 14.00
$P_{2.5}, P_{97.5}$	1.00, 17.00	$-1.00$ , 18.00	$-7.00$ , 7.00
Median ( $P_{25}, P_{75}$ )	6.00 (4.00, 10.00)	6.00 (3.00, 9.00)	0.00 ( $-2.00$ , 2.00)
Mean (95% CI)	7.04 (6.74 to 7.33)	6.49 (6.15 to 6.80)	$-0.07$ ( $-0.43$ to 0.31)
SD (95% CI)	4.27 (3.92 to 4.43)	4.90 (4.68 to 5.25)	3.31 (2.96 to 3.66)
Mean (95% CI) <sup>a</sup>	6.74 (6.43 to 7.03)	N/A <sup>b</sup>	0.17 ( $-0.19$ to 0.55)
SD (95% CI) <sup>a</sup>	4.34 (3.98 to 4.50)	N/A	3.32 (2.96 to 3.68)

<sup>a</sup>Results of sensitivity analyses. For incubation period, the third day for those whose exposure period spanned three days and the second day for those exposed for two continuous days were used as their exposure dates in sensitivity analyses. For symptoms-to-transmission time, exposure dates were taken in a different way in sensitivity analyses (like those of the incubation period).

<sup>b</sup>N/A: not applicable.

**Figure 2.** The distribution of (A) incubation period ( $n=687$ ), (B) serial interval ( $n=1015$  pairs), and (C) symptoms-to-transmission time ( $n=296$  pairs).



## Sensitivity and Stratified Analyses

In sensitivity analyses, the estimates of incubation period and symptoms-to-transmission time remained stable, with a difference of 0.24-0.3 days from those in primary analysis (Table 2), which represents the largest possible error caused by

inclusion of cases with an exposure period spanning 2 or 3 days. The results of the stratified analysis are summarized in Multimedia Appendix 5. The difference between strata was statistically significant in four of the analyses (see footnote of Multimedia Appendix 5). Briefly, the incubation period was longer for those exposed for only 1 day (as compared with those

exposed for 2-3 days), while serial interval and symptoms-to-transmission time were longer for those at lower generations of clustered infection. The serial interval was also longer for those infected through nonhousehold contact.

## Discussion

By pooling the individual data of 1591 cases, we estimated the mean incubation period, serial interval, and symptoms-to-transmission time of COVID-19 to be 7.04, 6.49, and -0.07 days, respectively. The incubation period was longer than 14 days in 5.82% of the cases. Infectees' symptom onsets occurred before those of infectors in 3.15% of the infector-infectee pairs. Presymptomatic transmission occurred in 43.58% of the infector-infectee pairs.  $R_0$  was estimated to be 1.85 (95% CI 1.37-2.60).

To our knowledge, this study represents the largest case series with accurate data on incubation period, serial interval, and frequency of presymptomatic transmission. Our estimates of incubation period and serial interval are longer than most of the previous estimates [5-10]. There are several possible reasons for the difference. First, the sample size is generally small in previous studies [5,7,9,12-15], but much larger in this one. Second, most cases in previous studies had a long or unclear interval of exposure, making it difficult to determine the exact exposure date [6,8,17] and giving rise to error. By contrast, this study applied strict inclusion criteria regarding the exposure period and a simple method to determine exposure date to ensure the potential error in the estimates was small (<0.3 days, according to the sensitivity analysis). Third, the order of transmission (ie, who is the infector and who is the infectee) in clustering cases, which is crucial to estimation of both parameters, is easy to be mistaken given the possibility of presymptomatic and asymptomatic transmission [21,22,27,28]. According to our experience in screening the publicly available data, for many clusters, there was no clear evidence (eg, who got infected outside and brought the virus to the cluster?) to establish the transmission order. To ensure the accuracy of data, we excluded such clusters in this study, while previous studies rarely described how this issue was handled [11].

Our finding that the incubation period was within 14 days for 94% of the cases lends support to the current practice of 14-day quarantine of persons with potential exposure to SARS-CoV-2. In line with other studies [17], we also found some cases who developed symptoms more than 14 days after exposure, indicating that longer quarantine periods might be justified for some people. However, as it is hard to know beforehand who will develop symptoms beyond 14 days after exposure, the cost of extending mandatory quarantine of many people and the potential consequence of failure to identify a few symptomatic cases must be weighed carefully. Testing at the end of quarantine and social distancing afterwards may help reduce the risk associated with this phenomenon.

The negative values of the serial interval and symptoms-to-transmission time provide evidence of presymptomatic transmission. Specifically, 43.58% of the

transmission events in this study occurred before infectors' symptom onsets. Asymptomatic transmission was also reported by the publications included in this study (Multimedia Appendix 2). These phenomena constitute a challenge in the control of the epidemic and highlight the importance of adequate testing, strict quarantine, and social distancing to reduce the transmission caused by "hidden" cases.

The  $R_0$  we estimated is smaller than those from previous estimates, which were mostly between 2 and 4 [20,29,30]. The difference may be due to either the methodological issues in obtaining parameters as discussed above or due to the estimating method itself. In estimating  $R_0$ , the serial interval was used to approximate generation time in the Euler-Lotka equation, which may lead to underestimation because of the possibly differing incubation periods of the infector and infectee. Assuming the incubation periods of the infector and infectee follow the same distribution, we found that the  $R_0$  was 1.90, almost the same as the original one, suggesting a small error in our estimation. In any case, a smaller  $R_0$  should not be interpreted as a low risk of transmission. Slow response of government, presymptomatic and asymptomatic transmission, and insufficient protection measures taken by the public together could lead to an out-of-control epidemic, as is the current situation in many countries.

This study has some limitations. First, it was based on publicly reported cases. Previous studies suggested that such cases may overrepresent the severe ones. However, as the publicly available information on severity was incomplete and inaccurate, we were unable to assess the magnitude and potential impact of this issue. Another problem with publicly reported cases is that their epidemiology history might be reported in more detail at the early stage of epidemic than when the number of cases grew larger. We found that this was especially true for Japan and Singapore. Thus, in general, the results of this study reflected more the situation during the early stage of the pandemic. Second, to ensure accuracy in estimating the incubation period, the cases exposed for a long (>3 days) or unclear period were excluded (which was inevitable and justifiable). This might have contributed partly to the longer incubation period in this study than others, as there was a trend toward a longer incubation period in the cases exposed for 1 day only (a relatively shorter period) as compared with those exposed for 2-3 days (Multimedia Appendix 5). Third,  $R_0$  might have been underestimated as the Euler-Lotka equation was used; however, as explained in the last paragraph, the magnitude of underestimation, if any, was very small.

In conclusion, this study obtained robust estimates of several key epidemiological parameters of COVID-19. It provides additional evidence on the mean incubation period of COVID-19, which supports the current practice of 14-day quarantine of persons with potential exposure but also suggests the need for additional measures. Presymptomatic transmission together with the asymptomatic transmission reported by previous studies highlight the importance of adequate testing, strict quarantine, and social distancing.

## Acknowledgments

We thank all the people who worked hard, risked their health, and even sacrificed their lives to fight against COVID-19. They are real heroes. We also thank the numerous individuals who reduced their social activities to a minimum to help contain the spread of SARS-CoV-2.

## Authors' Contributions

YZ conceived the idea and designed the study, has full access to all data in the study, and takes responsibility for the integrity of data and the accuracy of data analysis. MS, ZJ, ZM, YQ, GW, ZY, and YZ contributed to data acquisition, statistical analysis, and the writing of the report. WMH and ZS contributed to statistical analysis. All authors contributed to data interpretation and critical revision and reviewed and approved the final version.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

A full list of the data sources (except academic studies).

[[XLSX File \(Microsoft Excel File\), 46 KB - jmir\\_v22i10e19994\\_app1.xlsx](#)]

### Multimedia Appendix 2

A list of eligible studies included for analysis.

[[DOCX File, 20 KB - jmir\\_v22i10e19994\\_app2.docx](#)]

### Multimedia Appendix 3

The Akaike information criterion scores and shifted terms for lognormal, Weibull, and Gamma distributions.

[[XLSX File \(Microsoft Excel File\), 10 KB - jmir\\_v22i10e19994\\_app3.xlsx](#)]

### Multimedia Appendix 4

The timeline of events for an "average" infector-infectee pair in a transmission chain, according to the estimates from this study. E1: exposure of infector; E2: exposure of infectee; O1: symptom onset of infector; O2: symptom onset of infectee.

[[PNG File, 66 KB - jmir\\_v22i10e19994\\_app4.png](#)]

### Multimedia Appendix 5

Stratified analyses of incubation period, serial interval, and symptoms-to-transmission time according to the selected characteristic.

[[DOCX File, 15 KB - jmir\\_v22i10e19994\\_app5.docx](#)]

## References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 2020 Mar 26;382(13):1199-1207 [[FREE Full text](#)] [doi: [10.1056/NEJMoa2001316](#)] [Medline: [31995857](#)]
2. Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X. Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province. *J Med Internet Res* 2020 Jun 30;22(6):e19636 [[FREE Full text](#)] [doi: [10.2196/19636](#)] [Medline: [32544071](#)]
3. Luo H, Lie Y, Prinzen FW. Surveillance of COVID-19 in the General Population Using an Online Questionnaire: Report From 18,161 Respondents in China. *JMIR Public Health Surveill* 2020 Apr 27;6(2):e18576 [[FREE Full text](#)] [doi: [10.2196/18576](#)] [Medline: [32319956](#)]
4. Timmers T, Janssen L, Stohr J, Murk JL, Berrevoets MAH. Using eHealth to Support COVID-19 Education, Self-Assessment, and Symptom Monitoring in the Netherlands: Observational Study. *JMIR mHealth uHealth* 2020 Jun 23;8(6):e19822 [[FREE Full text](#)] [doi: [10.2196/19822](#)] [Medline: [32516750](#)]
5. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill* 2020 Feb;25(5):2000062 [[FREE Full text](#)] [doi: [10.2807/1560-7917.ES.2020.25.5.2000062](#)] [Medline: [32046819](#)]
6. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020 Apr 30;382(18):1708-1720. [doi: [10.1056/NEJMoa2002032](#)] [Medline: [32109013](#)]
7. Jiang X, Rayner S, Luo M. Does SARS-CoV-2 has a longer incubation period than SARS and MERS? *J Med Virol* 2020 May;92(5):476-478 [[FREE Full text](#)] [doi: [10.1002/jmv.25708](#)] [Medline: [32056235](#)]

8. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung S, et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *J Clin Med* 2020 Feb 17;9(2):538 [FREE Full text] [doi: [10.3390/jcm9020538](https://doi.org/10.3390/jcm9020538)] [Medline: [32079150](https://pubmed.ncbi.nlm.nih.gov/32079150/)]
9. Xu X, Wu X, Jiang X, Xu K, Ying L, Ma C, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-CoV-2) outside of Wuhan, China: retrospective case series. *BMJ* 2020 Feb 19;368:m606 [FREE Full text] [doi: [10.1136/bmj.m606](https://doi.org/10.1136/bmj.m606)] [Medline: [32075786](https://pubmed.ncbi.nlm.nih.gov/32075786/)]
10. Yang Y, Lu Q, Liu M, Wang Y, Zhang A, Jalali N. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. *MedRxiv* 2020:1. [doi: [10.1101/2020.02.10.20021675](https://doi.org/10.1101/2020.02.10.20021675)]
11. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerg Infect Dis* 2020 Jun;26(6):1341-1343 [FREE Full text] [doi: [10.3201/eid2606.200357](https://doi.org/10.3201/eid2606.200357)] [Medline: [32191173](https://pubmed.ncbi.nlm.nih.gov/32191173/)]
12. Tindale LJ, Stockdale JE, Coombe M, Garlock ES, Lau WYV, Saraswat M, et al. Evidence for transmission of COVID-19 prior to symptom onset. *Elife* 2020 Jun 22;9:e57149 [FREE Full text] [doi: [10.7554/eLife.57149](https://doi.org/10.7554/eLife.57149)] [Medline: [32568070](https://pubmed.ncbi.nlm.nih.gov/32568070/)]
13. You C, Deng Y, Hu W, Sun J, Lin Q, Zhou F, et al. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *Int J Hyg Environ Health* 2020 Jul;228:113555 [FREE Full text] [doi: [10.1016/j.ijheh.2020.113555](https://doi.org/10.1016/j.ijheh.2020.113555)] [Medline: [32460229](https://pubmed.ncbi.nlm.nih.gov/32460229/)]
14. Zhao S, Gao D, Zhuang Z, Chong M, Cai Y, Ran J. Estimating the serial interval of the novel coronavirus disease (COVID-19): A statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020. *Research Square* 2020:1. [doi: [10.21203/rs.3.rs-18805/v3](https://doi.org/10.21203/rs.3.rs-18805/v3)]
15. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis* 2020 Apr;93:284-286 [FREE Full text] [doi: [10.1016/j.ijid.2020.02.060](https://doi.org/10.1016/j.ijid.2020.02.060)] [Medline: [32145466](https://pubmed.ncbi.nlm.nih.gov/32145466/)]
16. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect Dis* 2020 Jul;20(7):793-802 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30230-9](https://doi.org/10.1016/S1473-3099(20)30230-9)] [Medline: [32247326](https://pubmed.ncbi.nlm.nih.gov/32247326/)]
17. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med* 2020 May 05;172(9):577-582 [FREE Full text] [doi: [10.7326/M20-0504](https://doi.org/10.7326/M20-0504)] [Medline: [32150748](https://pubmed.ncbi.nlm.nih.gov/32150748/)]
18. Milwid R, Steriu A, Arino J, Heffernan J, Hyder A, Schanzer D, et al. Toward Standardizing a Lexicon of Infectious Disease Modeling Terms. *Front Public Health* 2016 Sep 28;4:213 [FREE Full text] [doi: [10.3389/fpubh.2016.00213](https://doi.org/10.3389/fpubh.2016.00213)] [Medline: [27734014](https://pubmed.ncbi.nlm.nih.gov/27734014/)]
19. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Biol Sci* 2007 Feb 22;274(1609):599-604 [FREE Full text] [doi: [10.1098/rspb.2006.3754](https://doi.org/10.1098/rspb.2006.3754)] [Medline: [17476782](https://pubmed.ncbi.nlm.nih.gov/17476782/)]
20. Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *J Evid Based Med* 2020 Feb 12;13(1):3-7 [FREE Full text] [doi: [10.1111/jebm.12376](https://doi.org/10.1111/jebm.12376)] [Medline: [32048815](https://pubmed.ncbi.nlm.nih.gov/32048815/)]
21. Tong Z, Tang A, Li K, Li P, Wang H, Yi J, et al. Potential Presymptomatic Transmission of SARS-CoV-2, Zhejiang Province, China, 2020. *Emerg Infect Dis* 2020 May;26(5):1052-1054 [FREE Full text] [doi: [10.3201/eid2605.200198](https://doi.org/10.3201/eid2605.200198)] [Medline: [32091386](https://pubmed.ncbi.nlm.nih.gov/32091386/)]
22. Yu P, Zhu J, Zhang Z, Han Y. A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period. *J Infect Dis* 2020 May 11;221(11):1757-1761 [FREE Full text] [doi: [10.1093/infdis/jiaa077](https://doi.org/10.1093/infdis/jiaa077)] [Medline: [32067043](https://pubmed.ncbi.nlm.nih.gov/32067043/)]
23. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: [10.1126/science.abb6936](https://doi.org/10.1126/science.abb6936)] [Medline: [32234805](https://pubmed.ncbi.nlm.nih.gov/32234805/)]
24. The epidemic situation of pneumonia caused by a novel coronavirus. people.cn. URL: <http://health.people.com.cn/GB/26466/431463/431576/> [accessed 2020-10-06]
25. Coronavirus disease (COVID-2019) situation reports. World Health Organization. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/> [accessed 2020-10-06]
26. CBC News. URL: <https://www.cbc.ca/news/health/canada-coronavirus-timeline-1.5482310> [accessed 2020-10-09]
27. Bai Y, Yao L, Wei T, Tian F, Jin D, Chen L, et al. Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* 2020 Apr 14;323(14):1406-1407 [FREE Full text] [doi: [10.1001/jama.2020.2565](https://doi.org/10.1001/jama.2020.2565)] [Medline: [32083643](https://pubmed.ncbi.nlm.nih.gov/32083643/)]
28. Huang R, Xia J, Chen Y, Shan C, Wu C. A family cluster of SARS-CoV-2 infection involving 11 patients in Nanjing, China. *Lancet Infect Dis* 2020 May;20(5):534-535 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30147-X](https://doi.org/10.1016/S1473-3099(20)30147-X)] [Medline: [32119823](https://pubmed.ncbi.nlm.nih.gov/32119823/)]
29. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *Int J Infect Dis* 2020 Mar;92:214-217 [FREE Full text] [doi: [10.1016/j.ijid.2020.01.050](https://doi.org/10.1016/j.ijid.2020.01.050)] [Medline: [32007643](https://pubmed.ncbi.nlm.nih.gov/32007643/)]
30. Distanto C, Piscitelli P, Miani A. Covid-19 Outbreak Progression in Italian Regions: Approaching the Peak by the End of March in Northern Italy and First Week of April in Southern Italy. *Int J Environ Res Public Health* 2020 Apr 27;17(9):3025 [FREE Full text] [doi: [10.3390/ijerph17093025](https://doi.org/10.3390/ijerph17093025)] [Medline: [32349259](https://pubmed.ncbi.nlm.nih.gov/32349259/)]



*Edited by G Eysenbach, G Fagherazzi; submitted 08.05.20; peer-reviewed by T Liu, CM Dutra, A Birukov; comments to author 13.07.20; revised version received 02.08.20; accepted 27.08.20; published 12.10.20.*

*Please cite as:*

Ma S, Zhang J, Zeng M, Yun Q, Guo W, Zheng Y, Zhao S, Wang MH, Yang Z

Epidemiological Parameters of COVID-19: Case Series Study

J Med Internet Res 2020;22(10):e19994

URL: <http://www.jmir.org/2020/10/e19994/>

doi: [10.2196/19994](https://doi.org/10.2196/19994)

PMID: [33001833](https://pubmed.ncbi.nlm.nih.gov/33001833/)

©Shujuan Ma, Jiayue Zhang, Minyan Zeng, Qingping Yun, Wei Guo, Yixiang Zheng, Shi Zhao, Maggie H Wang, Zuyao Yang. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 12.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Application of an Artificial Intelligence Trilogy to Accelerate Processing of Suspected Patients With SARS-CoV-2 at a Smart Quarantine Station: Observational Study

Ping-Yen Liu<sup>1,2,3</sup>, MD, PhD; Yi-Shan Tsai<sup>4</sup>, MD; Po-Lin Chen<sup>5</sup>, MD, PhD; Huey-Pin Tsai<sup>6,7</sup>, MSc; Ling-Wei Hsu<sup>1,8</sup>, MSc, PhD; Chi-Shiang Wang<sup>9</sup>, MSc, PhD; Nan-Yao Lee<sup>5</sup>, MD; Mu-Shiang Huang<sup>2,9</sup>, MSc, MD; Yun-Chiao Wu<sup>3</sup>, MSc; Wen-Chien Ko<sup>5</sup>, MD; Yi-Ching Yang<sup>10</sup>, MD, MPH; Jung-Hsien Chiang<sup>9</sup>, PhD; Meng-Ru Shen<sup>11,12</sup>, MD, PhD

<sup>1</sup>Institute of Clinical Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>2</sup>Division of Cardiology, Internal Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>3</sup>Department of Clinical Medical Research, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>4</sup>Department of Medical Imaging, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>5</sup>Division of Infectious Diseases, Department of Internal Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>6</sup>Department of Pathology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>7</sup>Department of Medical Laboratory Science and Biotechnology, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>8</sup>Institute of Basic Medical Science, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>9</sup>Department of Computer Science and Information Engineering, College of Electrical Engineering and Computer Science, National Cheng Kung University, Tainan, Taiwan

<sup>10</sup>Department of Family Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>11</sup>Department of Obstetrics and Gynecology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

<sup>12</sup>Department of Pharmacology, College of Medicine, National Cheng Kung University, Tainan, Taiwan

**Corresponding Author:**

Meng-Ru Shen, MD, PhD

Department of Obstetrics and Gynecology

National Cheng Kung University Hospital, College of Medicine

National Cheng Kung University

138 Sheng-Li Rd

Tainan, 70401

Taiwan

Phone: 886 6 2353535 ext 5505

Email: [mrshen@mail.ncku.edu.tw](mailto:mrshen@mail.ncku.edu.tw)

## Abstract

**Background:** As the COVID-19 epidemic increases in severity, the burden of quarantine stations outside emergency departments (EDs) at hospitals is increasing daily. To address the high screening workload at quarantine stations, all staff members with medical licenses are required to work shifts in these stations. Therefore, it is necessary to simplify the workflow and decision-making process for physicians and surgeons from all subspecialties.

**Objective:** The aim of this paper is to demonstrate how the National Cheng Kung University Hospital artificial intelligence (AI) trilogy of diversion to a smart quarantine station, AI-assisted image interpretation, and a built-in clinical decision-making algorithm improves medical care and reduces quarantine processing times.

**Methods:** This observational study on the emerging COVID-19 pandemic included 643 patients. An “AI trilogy” of diversion to a smart quarantine station, AI-assisted image interpretation, and a built-in clinical decision-making algorithm on a tablet computer was applied to shorten the quarantine survey process and reduce processing time during the COVID-19 pandemic.

**Results:** The use of the AI trilogy facilitated the processing of suspected cases of COVID-19 with or without symptoms; also, travel, occupation, contact, and clustering histories were obtained with the tablet computer device. A separate AI-mode function

that could quickly recognize pulmonary infiltrates on chest x-rays was merged into the smart clinical assisting system (SCAS), and this model was subsequently trained with COVID-19 pneumonia cases from the GitHub open source data set. The detection rates for posteroanterior and anteroposterior chest x-rays were 55/59 (93%) and 5/11 (45%), respectively. The SCAS algorithm was continuously adjusted based on updates to the Taiwan Centers for Disease Control public safety guidelines for faster clinical decision making. Our ex vivo study demonstrated the efficiency of disinfecting the tablet computer surface by wiping it twice with 75% alcohol sanitizer. To further analyze the impact of the AI application in the quarantine station, we subdivided the station group into groups with or without AI. Compared with the conventional ED ( $n=281$ ), the survey time at the quarantine station ( $n=1520$ ) was significantly shortened; the median survey time at the ED was 153 minutes (95% CI 108.5-205.0), vs 35 minutes at the quarantine station (95% CI 24-56;  $P<.001$ ). Furthermore, the use of the AI application in the quarantine station reduced the survey time in the quarantine station; the median survey time without AI was 101 minutes (95% CI 40-153), vs 34 minutes (95% CI 24-53) with AI in the quarantine station ( $P<.001$ ).

**Conclusions:** The AI trilogy improved our medical care workflow by shortening the quarantine survey process and reducing the processing time, which is especially important during an emerging infectious disease epidemic.

(*J Med Internet Res* 2020;22(10):e19878) doi:[10.2196/19878](https://doi.org/10.2196/19878)

## KEYWORDS

SARS-CoV-2; COVID-19; artificial intelligence; smart device assisted decision making; quarantine station

## Introduction

In December 2019, a local outbreak of pneumonia was detected in Wuhan, Hubei Province, China. Initially of unknown cause, the pneumonia was quickly determined to be caused by a novel coronavirus [1,2]. The International Committee on the Taxonomy of Viruses termed this novel virus SARS-CoV-2; the disease it causes is called COVID-19 [3,4]. The COVID-19 outbreak subsequently spread to every province of mainland China and then globally, with 1,773,084 confirmed cases and 111,652 deaths as of April 14, 2020 [5].

COVID-19 is considered to be very contagious in human beings [6]. Person-to-person transmission has been documented for COVID-19 [7], and the prevention of in-hospital outbreaks has become an important issue for crowded departments with highly infectious areas, including emergency departments (EDs), outpatient clinics, and inpatient admission wards [8-10]. In response to this, Taiwan implemented a disaster preparation plan by moving to “heightened” military level status, the highest of the three security tiers during peacetime; this is similar to war preparation status, with increased alertness and enhanced caution regarding the possibility of cross-contamination among people, especially within hospitals. As the COVID-19 epidemic situation deteriorates, the burden on quarantine stations outside EDs at hospitals is increasing daily. To address the high screening workload at quarantine stations, all staff members with medical licenses are required to work shifts in these stations. Therefore, it has become necessary to simplify the workflow and decision-making process for physicians and surgeons from all subspecialties.

To address this pandemic, the government of Taiwan reacted with a swift, efficient, and precise response [11]. Taiwan prevented entry to travelers from high-risk regions outside Taiwan as the first step; also, Taiwanese citizens and alien residents were quarantined, screened, and isolated based on specific criteria and conditions announced by the Taiwan Centers for Disease Control (CDC). However, the situation was dynamic and fluid and changed daily, with new CDC announcements requiring immediate protocol changes. To

address these criteria in their screening strategy and to reduce the processing time in COVID-19 quarantine stations, National Cheng Kung University Hospital (NCKUH) designed an artificial intelligence (AI) system to address this unmet clinical need. The NCKUH AI team used previous experience from prior smart medicine research and applications to develop a specific computer-assisted technology device to accelerate medical decisions. The goal was to reduce potentially dangerous SARS-CoV-2 exposure by shortening the period of time spent by patients in the quarantine unit. In response to this increasing public health emergency, the NCKUH AI team developed a smart quarantine station for use outside the NCKUH ED to facilitate the screening and survey process. The aim of this study is to demonstrate how the NCKUH AI trilogy of diversion to the smart quarantine station, AI-assisted image interpretation, and a built-in clinical decision-making algorithm improved medical care and reduced quarantine processing times.

## Methods

### The Smart Quarantine Station AI Trilogy

To improve the workflow and efficacy of the COVID-19 infection survey at NCKUH, an AI trilogy policy was employed outside and adjacent to the ED. The three parts of the trilogy were diversion of suspected patients with COVID-19 to the quarantine station; a travel, occupation, contact, and cluster (TOCC) survey on a tablet computer; and an AI-assisted image interpretation feature with clinical strategy decision-making algorithms at the smart quarantine station. Together, these steps facilitated the screening and survey process for suspected cases with or without symptoms and the collection of TOCC histories at the NCKUH.

The NCKUH is a referral medical center and teaching hospital with the largest medical service capacity in Tainan, a traditional culture city. Tainan is the sixth largest city in Taiwan, with >1,880,000 citizens. NCKUH has three main care facilities: an outpatient building that houses short-term patient care and clinical laboratories; an inpatient building where long-term care, the intensive care unit, and surgery wards are located; and the

ED. With >5000 outpatient visits daily and 1300 inpatient beds in addition to >300 emergency critical care patients in the ED, NCKUH understood the need to prevent the emerging infectious disease epidemic from becoming a local disaster and overburdening the medical system. The lessons learned from Taiwan's previous experience with the severe acute respiratory syndrome coronavirus (SARS-CoV) outbreak in 2003 have paid substantial dividends during the current COVID-19 outbreak. To address the expected high influx of COVID-19-related hospital visits and prevent in-hospital outbreaks, a quarantine

station was established adjacent to the ED. NCKUH immediately set up eight temporary wartime quarantine tents according to the Wuhan Pneumonia Quarantine Action. Simultaneously, a more permanent structured housing facility for long-term quarantine was established near these wartime tents (Figure 1). This effective diversion of patients from the ED to the quarantine tents relieved pressure on the crowded ED. On average, each medical staff member was estimated to examine 35 to 50 quarantine patients per day.

**Figure 1.** The quarantine station at National Cheng Kung University Hospital. ED: emergency department.



### Workflow Survey and Definition of Survey Time

Because the COVID-19 outbreak was rapidly becoming an epidemic and posed a threat of high risk of exposure to our staff and patients, NCKUH urgently instituted a workflow survey. Patients who visited the quarantine station were asked to complete a questionnaire that included their TOCC history as well as any chronic illnesses, symptoms, or signs. Before they entered any private information, the patients gave signed consent allowing us to use their data and baseline characteristics for subsequent analysis of the workflow. The survey time was defined as the total of the time a patient took to complete the TOCC history with a nurse's written record and history taking; the time needed to undergo a chest x-ray examination with related inquiry and expert interpretation; and the time to revisit with the examining physician to discuss a medical decision based on the patient workflow recommended by the Taiwan CDC.

The physician would decide whether the patient should be hospitalized, be registered, or recover at home under self-care. Furthermore, the NCKUH staff could access electronic medical records using the tablet computer along with medical AI radiomics of chest x-rays to alert radiologists and pulmonary physicians to screen pneumonia patches earlier than usual. Finally, the embedded clinical decision-making algorithm assisted the staff in assessing potential cases; this could also be useful to physicians who were not familiar with the emerging virus. This process was updated daily using the most recent guidelines from the Taiwan Ministry of Health and Welfare in collaboration with the Taiwan CDC.

### Disinfection Protocol for the TOCC Tablet Computer

To assess if our disinfection process could effectively clean any possible SARS-CoV-2 contamination from a tablet computer surface, we performed an ex vivo study by spraying 20 microliters of a positive SARS-CoV-2 virus solution onto a computer in a P3 laboratory at NCKUH. After the computer



surface was air-dried for 30 minutes, it was swabbed by technicians to obtain the first sample (V0), which indicated the surface virus load after initial exposure to the virus. After V0 sample collection, the technicians used 75% alcohol to mimic the disinfection procedure that was recommended in the current Taiwan CDC guideline. Samples V1 and V2 were collected after disinfecting the surface once and twice using 75% alcohol, respectively. The crossing point (CP), which is the maximum second derivative of the amplification curve of the real time–polymerase chain reaction (RT-PCR), was measured to obtain the viral load on the tablet computer surface using the COVID-19 Genesig RT-PCR assay (Primerdesign Ltd) using a LightCycler 480 II PCR platform (Roche Holding AG). A positive viral load was defined by the CP of amplification curve <45 of the RT-PCR result. Two specific genes, RNA-dependent RNA polymerase (RdRp) and Envelope (E) genes, were simultaneously assayed to determine the results.

### AI-Assisted Reading of Chest X-Rays

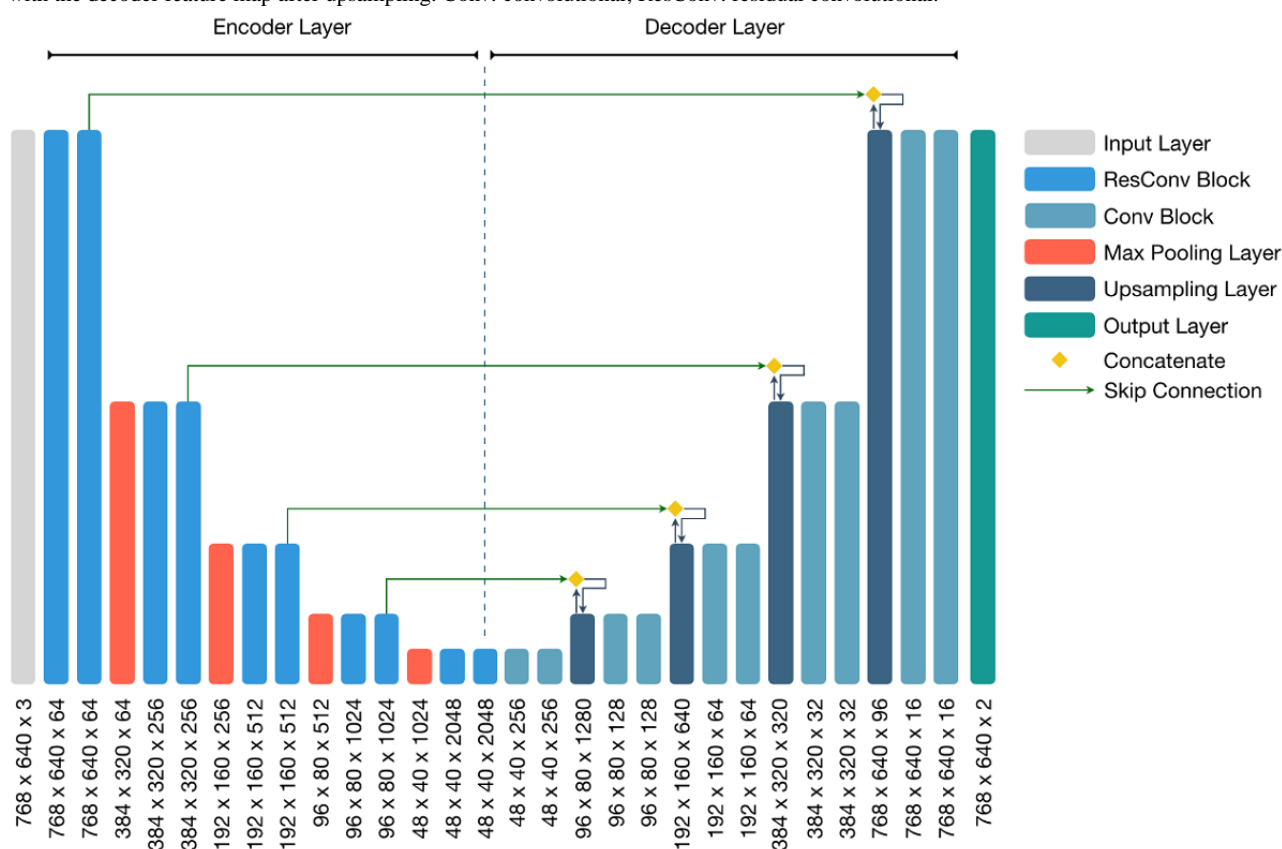
COVID-19 pneumonia exhibits a milder radiological course than severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) pneumonia; only 33% of COVID-19 pneumonia cases in Korea and only 60% of cases in China were reported to have chest x-ray abnormalities [12]. Patients with COVID-19 pneumonia can have negative chest

radiographs or computed tomography (CT) images even if their RT-PCR test is positive. According to Pan et al [13], the time course of COVID-19 pneumonia adheres to a clinical progression. The subclinical group has a CT pattern of unilateral and multifocal ground-glass opacities (GGOs) that progress immediately to bilateral and diffuse GGOs. The image may be mixed with light consolidations by the peak of disease severity, which occurs approximately 2 weeks after symptom onset [14]. In current practice, it is difficult to interpret chest radiographs showing GGOs, and this can limit the early diagnosis of COVID-19 pneumonia within the first week of symptom onset. Radiographic deterioration observed in follow-up thoracic imaging implies poor prognosis of COVID-19 pneumonia. Therefore, NCKUH set up portable imaging equipment to accelerate the characterization of chest conditions and limit the transportation of suspected cases to lower the risk of transmission [15].

### Development of AI Models for Chest X-Ray Analysis

The development of AI models for chest x-ray analysis can ease the burden on medical staff and enable more rapid triaging of patients. To detect the precise location of pneumonia sites, we adopted a segmentation model with a class attention map [16]. The pneumonia segmentation model was based on U-Net [17], as shown in Figure 2.

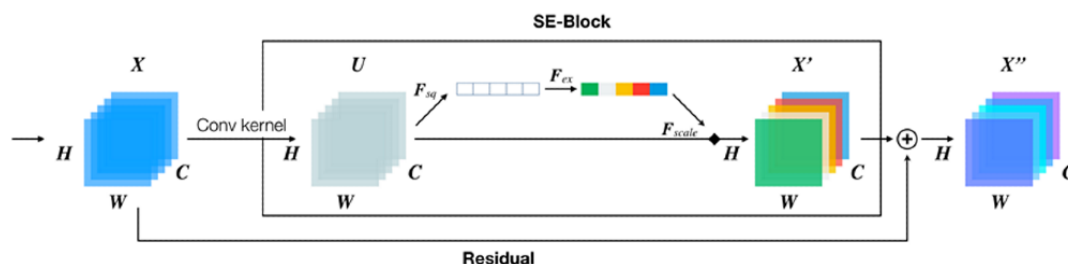
**Figure 2.** The pneumonia segmentation model based on U-Net. ResNet-50 was used as the encoder backbone model. The squeeze-and-excitation block was added to each ResConv block to enhance feature extraction from feature maps. The skip connection was used to concatenate the encoder feature map with the decoder feature map after upsampling. Conv: convolutional; ResConv: residual convolutional.



The Conv block was constructed using a  $3 \times 3$  convolutional block and batch normalization [18] and was then activated using a rectified linear unit (ReLU) function. We also adopted a residual network, ResNet-50 [19], with the pretrained weight

on the ImageNet as the backbone of this model. Moreover, we leveraged the squeeze-and-excitation block [20], as shown in Figure 3, to capture more significant information in each residual convolutional block.





**Figure 3.** Model of the addition of the SE block to the residual convolutional block. Conv: convolutional; SE: squeeze-and-excitation.

After  $U$  was established, the global average pooling in  $F_{sq}$  was used to obtain global information from each feature map. Two fully connected layers were adopted with the ReLU and sigmoid functions to apply a nonlinear transformation to earn attention information in  $F_{ex}$ . The attention weight produced by  $F_{ex}$  was multiplied to feature map  $U$  to obtain more efficient features of  $X'$  in  $F_{scale}$ . The input  $X$  was added to  $X'$  to obtain the new feature map  $X''$  to implement the residual block.

The output masks of this model contained two different predictors: pneumonia and GGO. We used multiple bounding boxes to label all lesions or as many lesions as possible. Because it is possible to observe different illnesses at the same position, we used a sigmoid activation function to leverage two  $1 \times 1$  convolutional kernels, one for detecting pneumonia and one for detecting GGO. Next, we trained this model using pixel-wise binary cross entropy and then minimized the loss function as follows:



where  $N$  and  $M$  denote the height and width of the input image, respectively, and  and  represent the gold standard and predicted value of this pixel, respectively.

### Grouping of Survey Patients

Patients could follow three routes in the NCKUH ED or quarantine station from January 31 to March 17, 2020. Route 1 was the traditional ED route, which did not include AI or the smart device app; this route was followed by patients who visited the ED between 8 PM and 8 AM, when no quarantine station service was available. Route 2 was the quarantine station without AI or the smart device app; most patients followed this route between January 31 and February 5, 2020. Route 3 was the quarantine station with AI and the smart device app, which was mainly followed by patients from February 5 to March 17, 2020. To compare baseline characteristics, we merged Route 2 and Route 3 into a general quarantine station route. Route 1 was recognized as the traditional ED track for screening. Data regarding patients' preexisting conditions, including coronary artery disease, diabetes, hypertension, chronic obstructive pulmonary disease, underlying malignancy, and chronic kidney disease with or without end-stage renal disease, were all obtained from medical records or the SCAS database. Patients'

symptoms, including dyspnea, cough, stuffed nose, fever, and body temperature, were also recorded individually. Informed patient consent was built into the front page of the tablet computer app before the start of the questionnaire due to the urgent need to collect data while avoiding paper cross-contamination.

### Patient and Public Involvement and Ethical Issues

This study was designed by the investigators. Patients or the public were enrolled in the research study when they visited the quarantine station and agreed to participate before answering the SCAS questionnaire on the tablet computer. The NCKUH Institutional Review Board approved our use of information and images obtained from the quarantine station (A-ER-109-149). Furthermore, due to the pandemic, the dissemination of results to the study participants was not applicable.

### Statistical Analysis

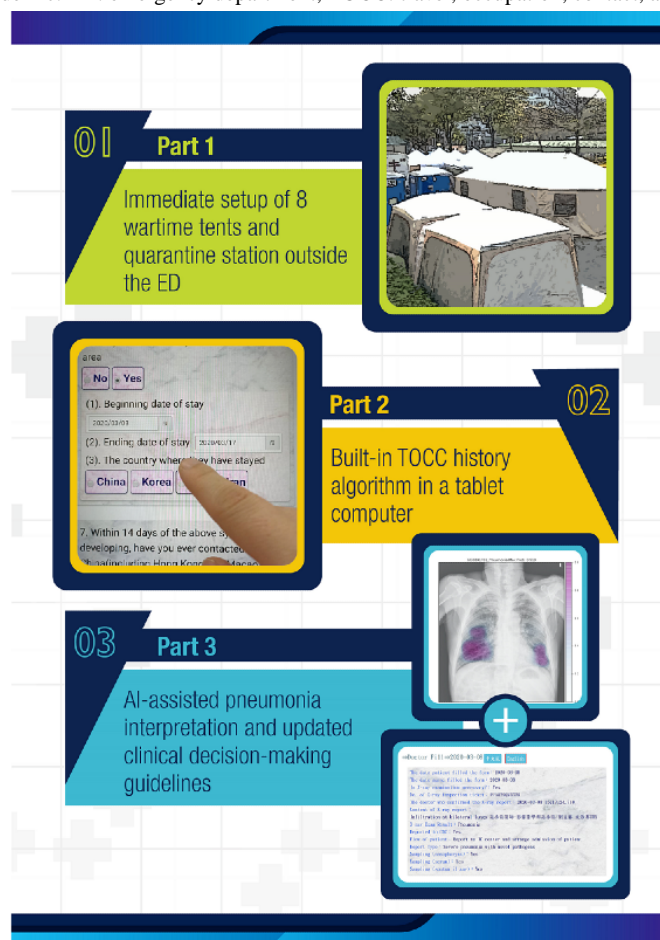
Statistical analyses were performed using SPSS Version 22.0 (IBM Corporation). Continuous data are presented as mean (SD) or median (IQR) depending on the distribution. Dichotomous data are presented as  $n$  (%). Comparisons were conducted with nonparametric statistics using the Wilcoxon rank sum test or the Mann-Whitney U test for continuous variables.  $P$  values  $< .05$  were considered significant.

## Results

### Overview

To improve the efficiency of the survey to determine COVID-19 infection risk and safely reduce crowding in the ED, NCKUH in Taiwan established an AI trilogy comprising three parts. First, we built a quarantine station with a SCAS to accelerate the workflow (Figure 4). SCAS is a clinical decision tree algorithm that integrates the structured format of the TOCC history recording, AI-assisted interpretation of chest x-rays, and the clinically recommended workflow. Patients at the quarantine station who were suspected of having COVID-19 used a tablet computer to answer the SCAS questionnaire on their TOCC history, with adequate alcohol disinfection between users. To avoid cross-infection between front-line medical staff and patients, all physicians and nurses used independent computers.

**Figure 4.** Synergetic combination of quarantine station establishment, smart patient processing, and AI to improve the efficiency and safety of patient processing during the COVID-19 pandemic. ED: emergency department; TOCC: travel, occupation, contact, and cluster.

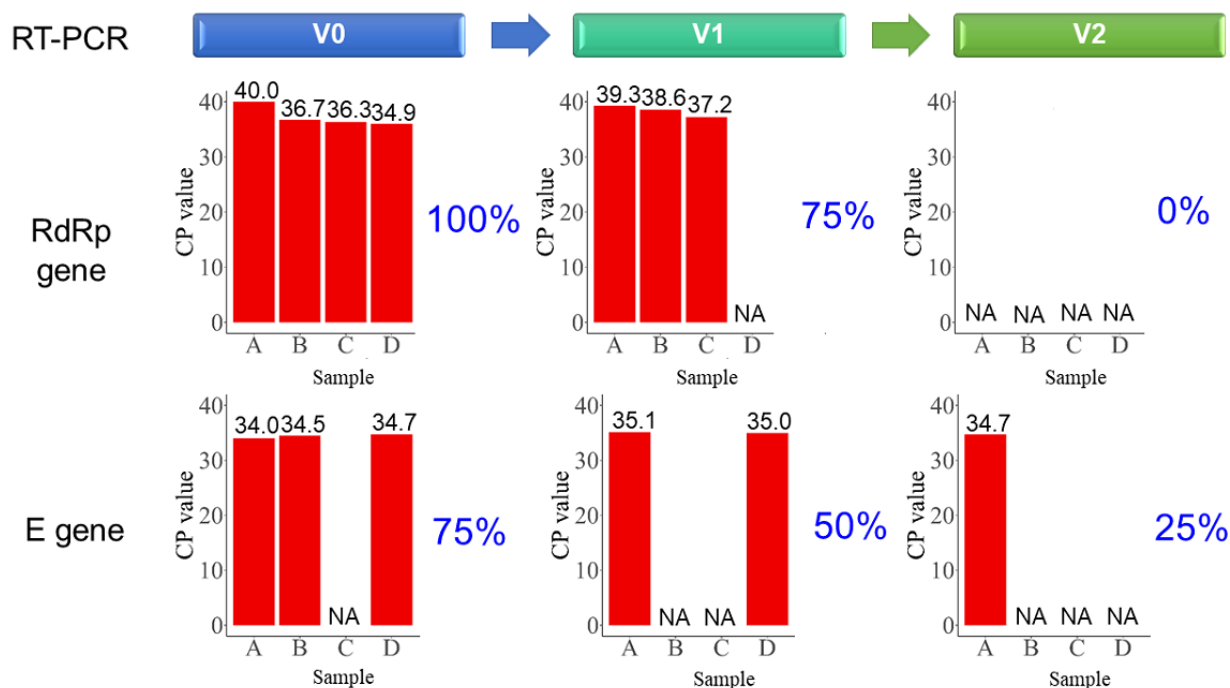


### Disinfection Process Assessment

To assess if our disinfection process of the tablet computer device using 75% alcohol sanitizer was effective and removed all residual viruses from the surface of the device, we performed an ex vivo study. First, 20 microliters of a positive SARS-CoV-2 virus solution was sprayed onto the computer surface. Subsequently, 75% alcohol was used to clean and disinfect the computer surface (Figure 5). We tested four samples without 75% alcohol treated and found that the initial results were 100%

positive tested by SARS-CoV-2 RT-PCR. Interestingly, the CP values for RdRp gene were 75% and 0% positive after the first and second disinfection procedures using 75% alcohol on the tablet computer surface, respectively. Similarly, the initial CP value for the COVID-19 E gene was 75% positive, and it decreased to 50% and 25% positive after the first and second alcohol disinfection procedures, respectively. These results supported our proposed tablet computer cleaning protocol of two careful disinfection processes using 75% alcohol between uses of the tablet.

**Figure 5.** Ex vivo study to determine the efficiency of disinfection of the tablet computer surface. Samples were collected from the tablet surface before disinfection (V0) and after the first and second disinfection processes using 75% alcohol (V1 and V2, respectively). The percentages indicate the positive rate (CP value <45 as positive) of the real time–polymerase chain reaction (RT-PCR, N=4 for each experiment). CP: crossing point. RdRp: RNA-dependent RNA polymerase gene. E: Envelope gene.



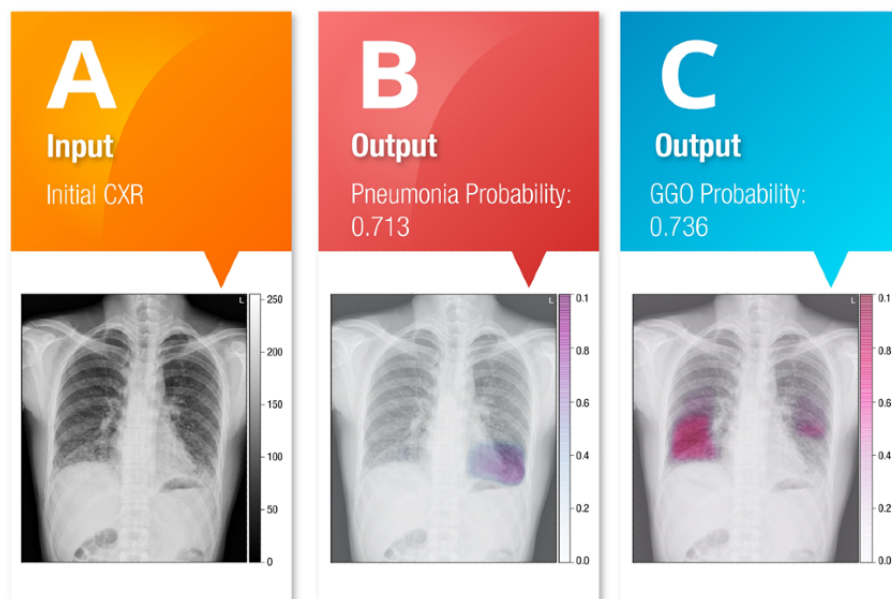
### AI Chest X-Ray Interpretation

To optimize AI chest x-ray interpretation, we retrospectively retrieved 4000 chest x-rays from our image archiving and communication system and identified 682 posteroanterior chest x-rays with pneumonia and 692 normal chest posteroanterior x-rays. All poor-quality anteroposterior chest radiographs and chest x-rays were excluded. To enhance the detection rate of light consolidations, we selected 46 cases with CT-evidenced GGOs and posteroanterior chest X-rays from the same day as the CT examination to serve as the GGO training data set (Figure 6). Our AI mode for quickly recognizing pulmonary infiltrates on chest x-rays was merged into the SCAS, with an area under

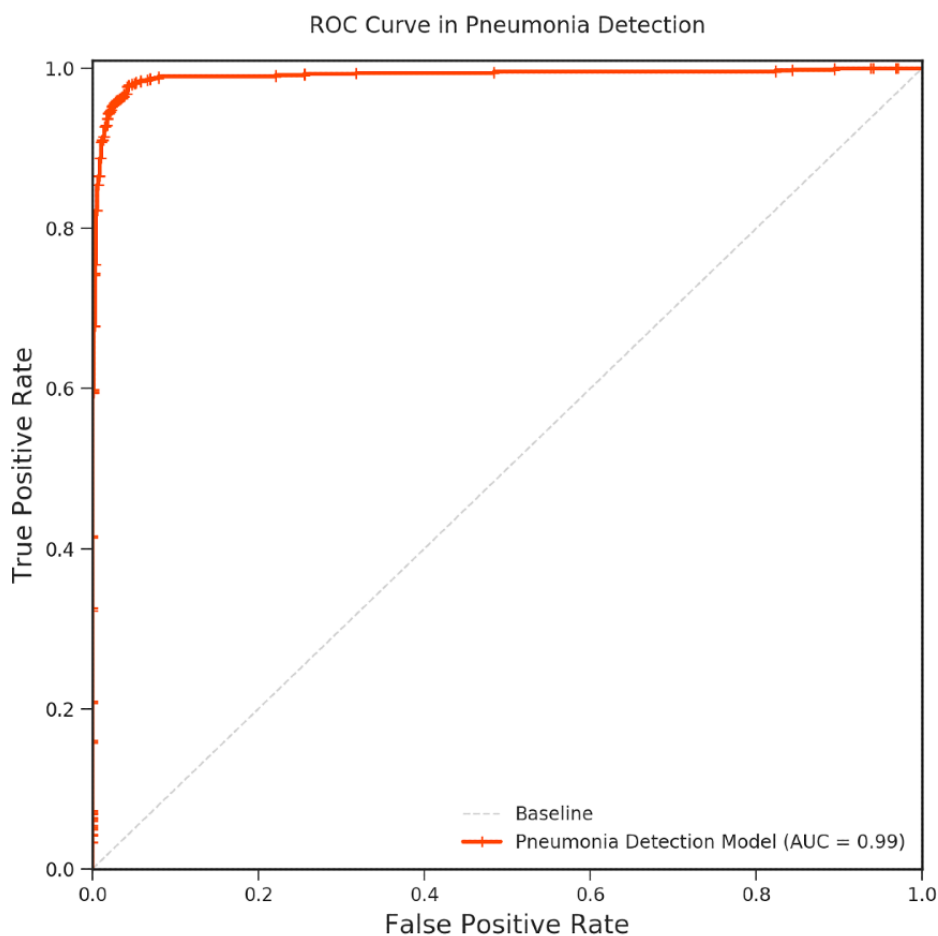
the receiver operating characteristic curve (AUC) of 0.99 (Figure 7), sensitivity of 94.1%, specificity of 95.1%, and accuracy of 94.6% using the training data set. Furthermore, we used cases of COVID-19 pneumonia from an open source data set in GitHub [21] including 59 posteroanterior and 11 anteroposterior chest x-rays to test the model. We achieved detection rates of 55/59 (93%) in posteroanterior and 5/11 (46%) in anteroposterior chest x-rays compared to interpretation by one cardiothoracic radiologist with 14 years of experience.

Our AI mode for quickly recognizing pulmonary infiltrates on chest x-rays was merged into the SCAS, with an AUC of 0.99, sensitivity of 94.1%, specificity of 95.1%, and accuracy of 94.6% using the training data set.

**Figure 6.** The artificial intelligence (AI) model for pneumonia detection incorporated into the smart clinical assisting system. (A) An original chest x-ray is automatically retrieved from the picture archiving and communication system and then interpreted by the AI model. (B) A consolidative lung is detected, and the diseased site is illustrated using a heat map. (C) A light consolidation GGO is identified by the AI model. CXR: chest x-ray; GGO: ground-glass opacity.



**Figure 7.** AUC of the AI chest x-ray interpretation system for recognizing pulmonary infiltrates on chest x-rays. AUC: area under the ROC curve; ROC: receiver operating characteristic.



## Clinical Data of Study Participants

The screening population of patients who visited the NCKUH ED and who agreed to provide their data for the study (N=643) was divided into two groups: the traditional ED group (n=281) and the smart quarantine station group (n=362). The baseline characteristics of these two groups are compared in Table 1.

The average age of the patients who visited the quarantine station was 35.6 years (SD 13.1), which was similar to the average age (34.7 years, SD 12.3) of patients visiting the traditional ED route. The most dominant trait in the TOCC was the traveling history for both groups, followed by the contact history. The screening population was a lower risk cohort, with

low prevalence rates of hypertension (9/281, 3.3%, and 19/362, 4.4%, in the ED and quarantine station groups, respectively) and chronic kidney disorder (8/281, 2.8%, and 5/362, 1.5%, in the ED and quarantine station groups, respectively); this suggests similar chronic illness histories in the two groups. Most patient symptoms manifested as a cough at the initial presentation (ED vs quarantine station: 140/281, 49.8% vs 226/362, 62.4%;  $P=.31$ ), followed by a stuffed nose (ED vs quarantine station: 88/281, 31.3%, vs 166/362, 45.8%;  $P=.31$ ) and fever (ED vs quarantine station: 78/281, 27.7%, vs 134/362, 37%;  $P=.37$ ). In general, the comparison between the two groups showed no difference in the distributions with respect to baseline age, gender, pre-existing conditions, or symptoms.

**Table 1.** Comparison of clinical data of patients seeking treatment in the traditional ED and at the smart quarantine station during the COVID-19 epidemic from January 31 to March 17, 2020 (N=643).

Characteristic	ED <sup>a</sup> (n=281)	Smart quarantine station (n=362)	P value
Age (years), mean (SD)	34.7 (12.3)	35.6 (13.1)	.41
Sex (male), n (%)	144 (51.2)	165 (45.6)	.11
<b>TOCC<sup>b</sup> history, n (%)</b>			
Traveling	10 (62.4)	308 (49.0)	.32
Occupational	2 (12.5)	33 (5.3)	.21
Clustering	1 (6.3)	30 (4.8)	.55
Contact	4 (25.0)	129 (20.5)	.75
<b>Pre-existing conditions, n (%)</b>			
CAD <sup>c</sup>	2 (0.5)	3 (0.9)	.99
Diabetes	3 (0.9)	11 (2.5)	.24
Hypertension	9 (3.3)	19 (4.4)	.67
COPD <sup>d</sup>	0 (0.0)	3 (0.7)	.56
Malignancy	8 (2.8)	7 (1.8)	.39
CKD <sup>e</sup> /ESRD <sup>f</sup>	8 (2.8)	5 (1.5)	.37
<b>Initial symptoms</b>			
Dyspnea, n (%)	53 (18.8)	60 (16.4)	.74
Cough, n (%)	140 (49.8)	226 (62.4)	.31
Stuffed nose, n (%)	88 (31.3)	166 (45.8)	.31
Fever, n (%)	78 (27.7)	134 (37.0)	.37
Body temperature (°C), mean (SD)	36.9 (0.9)	37.1 (1.7)	.69

<sup>a</sup>ED: emergency department.

<sup>b</sup>TOCC: travel, occupation, contact and cluster.

<sup>c</sup>CAD: coronary artery disease.

<sup>d</sup>COPD: chronic obstructive pulmonary disease.

<sup>e</sup>CKD: chronic kidney disease.

<sup>f</sup>ESRD: end-stage renal disease.

## Survey Times

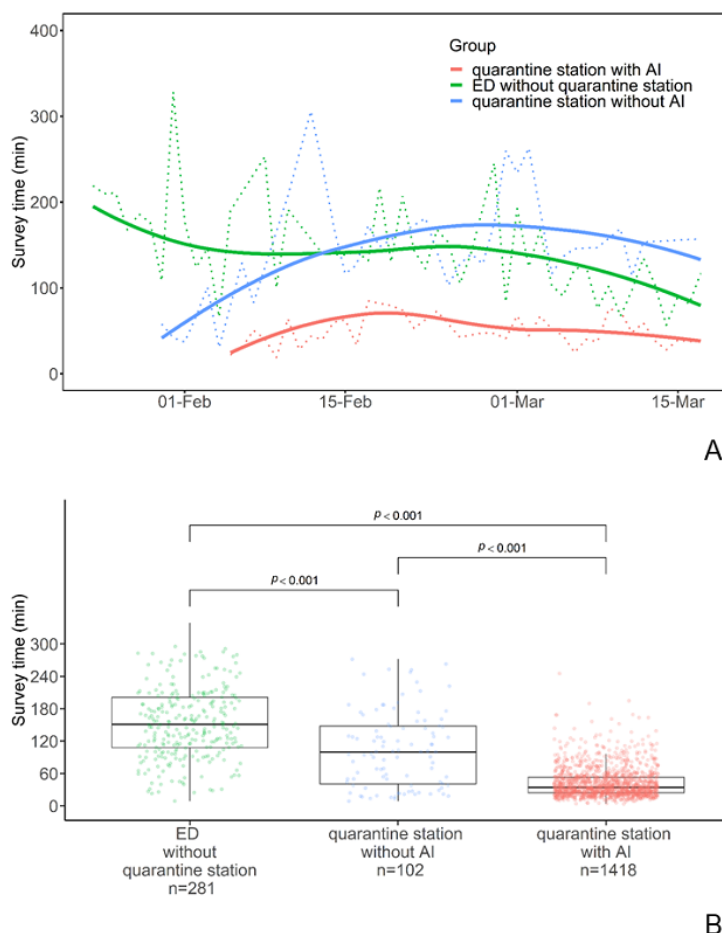
To further analyze the impact of the AI application on the quarantine station, we subdivided those in the quarantine station into groups with or without AI. As shown in Figure 8, compared with the conventional ED track (n=281), the survey time at the clinical quarantine station (n=1520) was significantly shortened,

with a median survey time at the ED of 153.0 minutes (95% CI 108.5-205.0) vs 35 minutes at the clinical quarantine station (95% CI 24-56;  $P<.001$ ). Furthermore, the use of the AI application in the quarantine station reduced the survey time in the station; the median survey time without AI in the quarantine



station was 100.5 minutes (95% CI 40.3-152.5), vs 34 minutes with AI in the quarantine station (95% CI 24-53;  $P<.001$ ).

**Figure 8.** (A) Comparison of survey processing times over the study period between the traditional ED and quarantine station groups with or without AI applications. (B) Box plots of the survey processing times for the traditional ED group, the quarantine station group without AI, and the quarantine station group with AI. AI: artificial intelligence; ED: emergency department; min: minutes.



## Discussion

### Principal Findings

In this study, we demonstrate how the NCKUH AI trilogy of diversion to a smart quarantine station incorporated with the SCAS, AI-assisted image interpretation, and a built-in clinical decision-making algorithm improved medical care and reduced the processing time at the quarantine station. In addition, our ex vivo study demonstrated the efficiency of 75% alcohol disinfection of the tablet computer after initial exposure to a positive viral load; by applying this procedure, the tablet computer is rendered safe for use by people in the quarantine station. This enhances patient quality of care and reduces risks to medical staff. Based on Taiwan's experience with SARS in 2003, the diversion of patients in the ED was crucial for early management and in-hospital infection control. Most Taiwanese hospitals needed to set up clinical quarantine stations near their EDs to survey patients who were suspected of being infected with COVID-19 due to the heavy caseloads experienced by hospitals daily. When a clinical quarantine station is set up, several factors can be detrimental to the conditions in the station. These factors include patient crowding in the restricted space within the quarantine station, high possibility of cross-contamination or coinfection, daily changes and

modifications to clinical guidelines by the Taiwan CDC, lack of familiarity of backup physicians with infectious disease control, and scarcity of radiology or pulmonary physicians in most hospitals to provide accurate and immediate pneumonia diagnoses.

To the best of our knowledge, this is the first time a smart quarantine station with a computer featuring a built-in clinical assistance system was used during an epidemic to accelerate complex workflows. Compared to the conventional ED track, the efficiency of the survey of the risk of COVID-19 infection was significantly improved by at least threefold for the NCKUH quarantine station. These results also highlight that the efficiency of the risk survey for COVID-19 infections at clinical quarantine stations is of general interest and helps address an unmet clinical need.

Triage in the ED for pandemic infectious disease screening is particularly challenging. Although telemedicine allows patients and physicians to communicate without direct contact, symptomatic or worried patients still directly visit the ED, and they can be efficiently screened by modern triage [22]. For in-person care, patients with positive signs of high-risk features should be immediately isolated to avert further contact with patients and health care workers. According to Hollander et al

[22], tablet computers can be cleaned between patients using well-defined infection control procedures. Furthermore, daily updates of the SCAS system help physicians or surgeons who may not be experts in radiologic interpretation of pneumonia patches or infectious disease to make proper decisions with little preparation before assuming their quarantine station duties.

The COVID-19 pandemic is affecting not only patients but also economies, politics, and daily life worldwide. Better management of future pandemics will reduce these human and economic burdens. All medical science personnel should use lessons learned from past and present pandemics to improve

upon past performance to prevent catastrophic effects of future disease outbreaks.

## Conclusions

We demonstrated a feasible, safe, and scientific application of a smart device with a built-in algorithm combined with an AI-based image system in a quarantine station to facilitate the survey process, avoid cross-infection, and reduce the burden on team members, including physicians, nurses, and technicians. This helpful process should be adapted into our strategies to address emerging endemic infectious diseases in the future.

## Acknowledgments

This study was supported by grants 109-2745-B-006-003, 109-2327-B-006-005, 109-2634-F-006-023, and 108-2314-B-006-098-MY3 from the Ministry of Science and Technology of Taiwan and grants D108-G2512 D109-G4803, D109-G4804, and D109-G2512 from Higher Education Sprout Project, Ministry of Education, to the Headquarters of University Advancement at National Cheng Kung University. We would like to thank all the NCKUH faculty members who participated in the quarantine station during this emerging infectious disease pandemic. We would also like to thank our technicians Chun-Sheng Yeh, I-Ting Lin, and Li-Rong Wang for their kind technical support and suggestions for the ex vivo study. Finally, we would like to thank Prof Michael W Hughes for his help with this English edition. We are grateful to Prof Pei-Fan Su for discussing and providing statistical consulting services from the Center for Quantitative Sciences, Clinical Medicine Research Center, NCKUH.

## Conflicts of Interest

None declared.

## References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020 Feb;395(10223):497-506. [doi: [10.1016/s0140-6736\(20\)30183-5](https://doi.org/10.1016/s0140-6736(20)30183-5)]
2. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)] [Medline: [32007143](https://pubmed.ncbi.nlm.nih.gov/32007143/)]
3. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020 Apr 2;5(4):536-544 [FREE Full text] [doi: [10.1038/s41564-020-0695-z](https://doi.org/10.1038/s41564-020-0695-z)] [Medline: [32123347](https://pubmed.ncbi.nlm.nih.gov/32123347/)]
4. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: [10.1056/NEJMoa2001316](https://doi.org/10.1056/NEJMoa2001316)] [Medline: [31995857](https://pubmed.ncbi.nlm.nih.gov/31995857/)]
5. Coronavirus disease 2019 (COVID-19) Situation Report 51. World Health Organization. 2020 Mar 11. URL: <https://apps.who.int/iris/bitstream/handle/10665/331475/nCoVsitrep11Mar2020-eng.pdf> [accessed 2020-10-08]
6. Glasser JW, Hupert N, McCauley MM, Hatchett R. Modeling and public health emergency responses: lessons from SARS. *Epidemics* 2011 Mar;3(1):32-37 [FREE Full text] [doi: [10.1016/j.epidem.2011.01.001](https://doi.org/10.1016/j.epidem.2011.01.001)] [Medline: [21420657](https://pubmed.ncbi.nlm.nih.gov/21420657/)]
7. Abbott S, Hellewell J, Munday J, CMMID nCoV working group, Funk S. The transmissibility of novel Coronavirus in the early stages of the 2019-20 outbreak in Wuhan: Exploring initial point-source exposure sizes and durations using scenario analysis. *Wellcome Open Res* 2020 Feb 3;5:17 [FREE Full text] [doi: [10.12688/wellcomeopenres.15718.1](https://doi.org/10.12688/wellcomeopenres.15718.1)] [Medline: [32322691](https://pubmed.ncbi.nlm.nih.gov/32322691/)]
8. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill* 2020 Jan;25(4):pii=2000058 [FREE Full text] [doi: [10.2807/1560-7917.ES.2020.25.4.2000058](https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058)] [Medline: [32019669](https://pubmed.ncbi.nlm.nih.gov/32019669/)]
9. Heymann DL, Shindo N. COVID-19: what is next for public health? *Lancet* 2020 Feb;395(10224):542-545. [doi: [10.1016/s0140-6736\(20\)30374-3](https://doi.org/10.1016/s0140-6736(20)30374-3)]
10. Majumder M, Mandl KD. Early Transmissibility Assessment of a Novel Coronavirus in Wuhan, China. SSRN Preprint posted online on January 1, 2020. [doi: [10.2139/ssrn.3524675](https://doi.org/10.2139/ssrn.3524675)] [Medline: [32714102](https://pubmed.ncbi.nlm.nih.gov/32714102/)]
11. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. *JAMA* 2020 Apr 14;323(14):1341-1342. [doi: [10.1001/jama.2020.3151](https://doi.org/10.1001/jama.2020.3151)] [Medline: [32125371](https://pubmed.ncbi.nlm.nih.gov/32125371/)]

12. Yoon SH, Lee KH, Kim JY, Lee YK, Ko H, Kim KH, et al. Chest Radiographic and CT Findings of the 2019 Novel Coronavirus Disease (COVID-19): Analysis of Nine Patients Treated in Korea. *Korean J Radiol* 2020 Apr;21(4):494-500 [FREE Full text] [doi: [10.3348/kjr.2020.0132](https://doi.org/10.3348/kjr.2020.0132)] [Medline: [32100485](https://pubmed.ncbi.nlm.nih.gov/32100485/)]
13. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). *Radiology* 2020 Jun;295(3):715-721. [doi: [10.1148/radiol.2020200370](https://doi.org/10.1148/radiol.2020200370)] [Medline: [32053470](https://pubmed.ncbi.nlm.nih.gov/32053470/)]
14. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis* 2020 Apr;20(4):425-434. [doi: [10.1016/s1473-3099\(20\)30086-4](https://doi.org/10.1016/s1473-3099(20)30086-4)]
15. Kooraki S, Hosseiny M, Myers L, Gholamrezanezhad A. Coronavirus (COVID-19) Outbreak: What the Department of Radiology Should Know. *J Am Coll Radiol* 2020 Apr;17(4):447-451 [FREE Full text] [doi: [10.1016/j.jacr.2020.02.008](https://doi.org/10.1016/j.jacr.2020.02.008)] [Medline: [32092296](https://pubmed.ncbi.nlm.nih.gov/32092296/)]
16. Zhou B, Lapedriza A, Oliva A. Learning Deep Features for Discriminative Localization. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV p. 2921-2929. [doi: [10.1109/cvpr.2016.319](https://doi.org/10.1109/cvpr.2016.319)]
17. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Lecture Notes in Computer Science, vol 9351. Cham, Switzerland: Springer; 2015.
18. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: *Proceedings of the 32nd International Conference on Machine Learning*. 2015 Presented at: 32nd International Conference on Machine Learning; July 6-11, 2015; Lille, France p. 448-456.
19. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV p. 770-778. [doi: [10.1109/cvpr.2016.90](https://doi.org/10.1109/cvpr.2016.90)]
20. Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. 2018 Presented at: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-22, 2018; Salt Lake City, UT p. 7132-7141. [doi: [10.1109/cvpr.2018.00745](https://doi.org/10.1109/cvpr.2018.00745)]
21. Covid Chest X-Ray dataset. GitHub. URL: <https://github.com/ieee8023/covid-chestxray-dataset> [accessed 2020-10-08]
22. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. *N Engl J Med* 2020 Apr 30;382(18):1679-1681. [doi: [10.1056/NEJMp2003539](https://doi.org/10.1056/NEJMp2003539)] [Medline: [32160451](https://pubmed.ncbi.nlm.nih.gov/32160451/)]

## Abbreviations

**AI:** artificial intelligence  
**CDC:** Centers for Disease Control  
**Cp:** crossing point  
**CT:** computed tomography  
**ED:** emergency department  
**GGO:** ground-glass opacity  
**MERS:** Middle East respiratory syndrome  
**NCKUH:** National Cheng Kung University Hospital  
**ReLU:** rectified linear unit  
**RdRp:** RNA-dependent RNA polymerase  
**RT-PCR:** real time–polymerase chain reaction  
**SARS:** severe acute respiratory syndrome  
**SARS-CoV:** severe acute respiratory syndrome coronavirus  
**SCAS:** smart clinical assisting system  
**TOCC:** travel, occupation, contact, and cluster

*Edited by G Eysenbach; submitted 02.07.20; peer-reviewed by D Mendes, S Lalmuanawma; comments to author 10.09.20; revised version received 15.09.20; accepted 16.09.20; published 14.10.20.*

### *Please cite as:*

Liu PY, Tsai YS, Chen PL, Tsai HP, Hsu LW, Wang CS, Lee NY, Huang MS, Wu YC, Ko WC, Yang YC, Chiang JH, Shen MR  
*Application of an Artificial Intelligence Trilogy to Accelerate Processing of Suspected Patients With SARS-CoV-2 at a Smart Quarantine Station: Observational Study*  
*J Med Internet Res* 2020;22(10):e19878  
 URL: <http://www.jmir.org/2020/10/e19878/>  
 doi:[10.2196/19878](https://doi.org/10.2196/19878)  
 PMID:[33001832](https://pubmed.ncbi.nlm.nih.gov/33001832/)

©Ping-Yen Liu, Yi-Shan Tsai, Po-Lin Chen, Huey-Pin Tsai, Ling-Wei Hsu, Chi-Shiang Wang, Nan-Yao Lee, Mu-Shiang Huang, Yun-Chiao Wu, Wen-Chien Ko, Yi-Ching Yang, Jung-Hsien Chiang, Meng-Ru Shen. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# How Data Analytics and Big Data Can Help Scientists in Managing COVID-19 Diffusion: Modeling Study to Predict the COVID-19 Diffusion in Italy and the Lombardy Region

Davide Tosi<sup>1</sup>, DPhil; Alessandro Campi<sup>2</sup>, DPhil

<sup>1</sup>Università degli Studi dell'Insubria, Varese, Italy

<sup>2</sup>Politecnico di Milano, Milano, Italy

**Corresponding Author:**

Alessandro Campi, DPhil

Politecnico di Milano

Piazza Leonardo da Vinci 32

Milano

Italy

Phone: 39 0223993644

Email: [alessandro.campi@polimi.it](mailto:alessandro.campi@polimi.it)

## Abstract

**Background:** COVID-19 is the most widely discussed topic worldwide in 2020, and at the beginning of the Italian epidemic, scientists tried to understand the virus diffusion and the epidemic curve of positive cases with controversial findings and numbers.

**Objective:** In this paper, a data analytics study on the diffusion of COVID-19 in Italy and the Lombardy Region is developed to define a predictive model tailored to forecast the evolution of the diffusion over time.

**Methods:** Starting with all available official data collected worldwide about the diffusion of COVID-19, we defined a predictive model at the beginning of March 2020 for the Italian country.

**Results:** This paper aims at showing how this predictive model was able to forecast the behavior of the COVID-19 diffusion and how it predicted the total number of positive cases in Italy over time. The predictive model forecasted, for the Italian country, the end of the COVID-19 first wave by the beginning of June.

**Conclusions:** This paper shows that big data and data analytics can help medical experts and epidemiologists in promptly designing accurate and generalized models to predict the different COVID-19 evolutionary phases in other countries and regions, and for second and third possible epidemic waves.

(*J Med Internet Res* 2020;22(10):e21081) doi:[10.2196/21081](https://doi.org/10.2196/21081)

## KEYWORDS

COVID-19; SARS-CoV-2; big data; data analytics; predictive models; prediction; modeling; Italy; diffusion

## Introduction

The unexpected pandemic diffusion of COVID-19 [1] worldwide calls for the need to study all available data to promptly understand the epidemic curve of COVID-19 contagiousness [2-4] to help medical experts, epidemiologists, and political decision makers in designing prompt reaction plans.

In this paper, we aim to develop an accurate predictive model tailored to forecast the evolution of the diffusion over time, exploiting big data and data analytics. Generally, epidemics follow an exponential curve in the spread of positive cases. This is not the case of the curve observed in Wuhan, China [5], where the official curve of confirmed positive cases of COVID-19

follows a behavior that is different from typical epidemics. More precise studies are reported in [1,6-8]. Starting from this observation, we tried at the beginning of March to correlate the Wuhan official data set with the COVID-19 data set available in Italy by applying big data and data analytics techniques [9] to all official open data available worldwide [5] to design the logistic curve for early estimation of the number of COVID-19 positive cases day-to-day and in all the phases of the pandemic over time (eg, the peak in the number of daily cases, the logistic plateau, and at the end of the pandemic).



## Methods

To design the predictive model, we exploited all the official data sets available so far. We analyzed the Wuhan official data set, available at [5]. As for the Italian perspective, we adopted the official data set that is daily published and updated by the Department of the Italian Civil Protection [10]. All these data sets are freely available. Additional statistics (eg, percentage of people that have severe COVID-19, percentage of people that need intensive care, COVID-19 mortality index) have been imported from the World Health Organization (WHO) website [11].

Exploiting the similarity between the behavior of the COVID-19 contagion in Wuhan and the starting Italian scenario, we designed our predictive curve adapting well-known mathematical methods to this particular context: the Pearson correlation index to formally evaluate this similarity in the contagion, the logistic curve (sigmoid) to design the cumulative

number of COVID-19 positive cases, regression models to evaluate the best correlation fit for our predictions, and the power law to model the initial ascent of the pandemic.

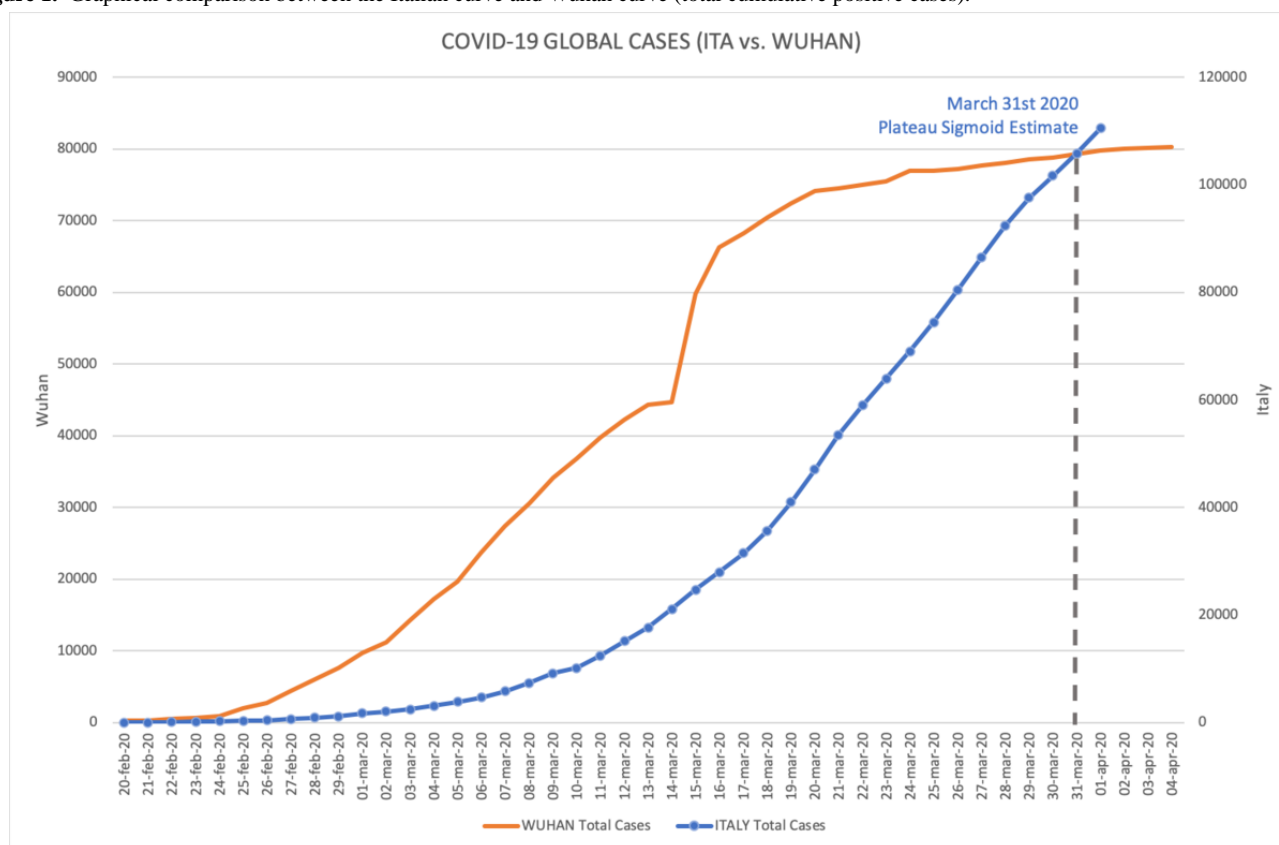
## Results

In the first step of our method, we computed the Pearson correlation coefficient applied to a sample between the two data sets (Italy vs Wuhan) with reference to the daily new positive cases as:



Starting from the data set available from March 2, 2020, it was possible to observe a strong Pearson correlation between the data set related to Italy and the data set related to Wuhan with a *Correlation Index of 0.9944*, as depicted in Figure 1 where the Wuhan curve is compared with the Italian one.

**Figure 1.** Graphical comparison between the Italian curve and Wuhan curve (total cumulative positive cases).



Starting from this assumption, it was then possible to use the Wuhan data set (with a strong statistical significance) to try to predict the logistic curve of total positive cases to COVID-19 for Italy and the Lombardy Region. We started designing our model by the basic assumption that every pandemic phenomenon follows a logistic (sigmoid) curve. We then searched for a curve that best fit the initial growing part of the logistic, and we found that the best fit is a power law curve in the form of  $y = m \cdot x^b$ . To compute the coefficients  $m$  and  $b$  of the power curve, we started our elaborations with two other assumptions in mind:

1. We based our initial estimation on the number of swab tests analyzed in the initial days of the pandemic period, where an average of 8000 swab tests were performed daily (in May, the number of swab tests was increased significantly to an average of 50,000 tests/day).
2. We assumed that the Italian Government would act promptly with restrictions and lockdowns on the Italian population and that the Italian citizens would follow these restrictions with a sense of responsibility.

Moreover, we used the stabilized data set for Wuhan City, and we adopted additional official statistics (published by the WHO

[11]) about the COVID-19 spread in China (eg, percentage of severe disease cases, percentage of critical disease cases, crude fatality ratio, the number of days passed in each phase of the China pandemic) to adjust the multiplier coefficient of our predictive model. In this way, we were able to estimate the position of the first inflections of the curve (to estimate the date for the peak of new daily positive cases), the second curve inflection (to estimate the date of the curve plateau), and then to determine the day-to-day number of positive cases during the second part of the pandemic (ie, during the descent.)

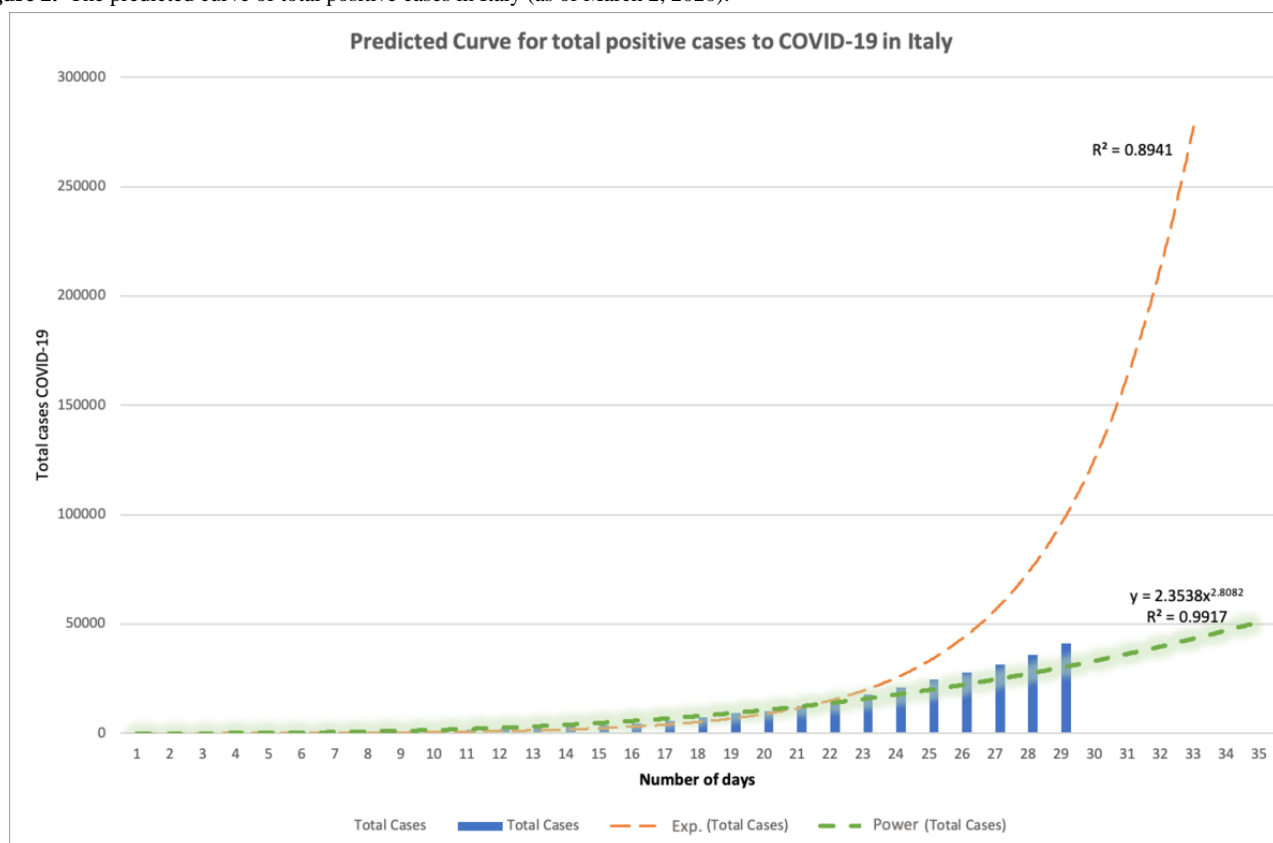
Figure 2 shows the number of new daily cases in Italy starting from the first identification of a patient in the town of Codogno

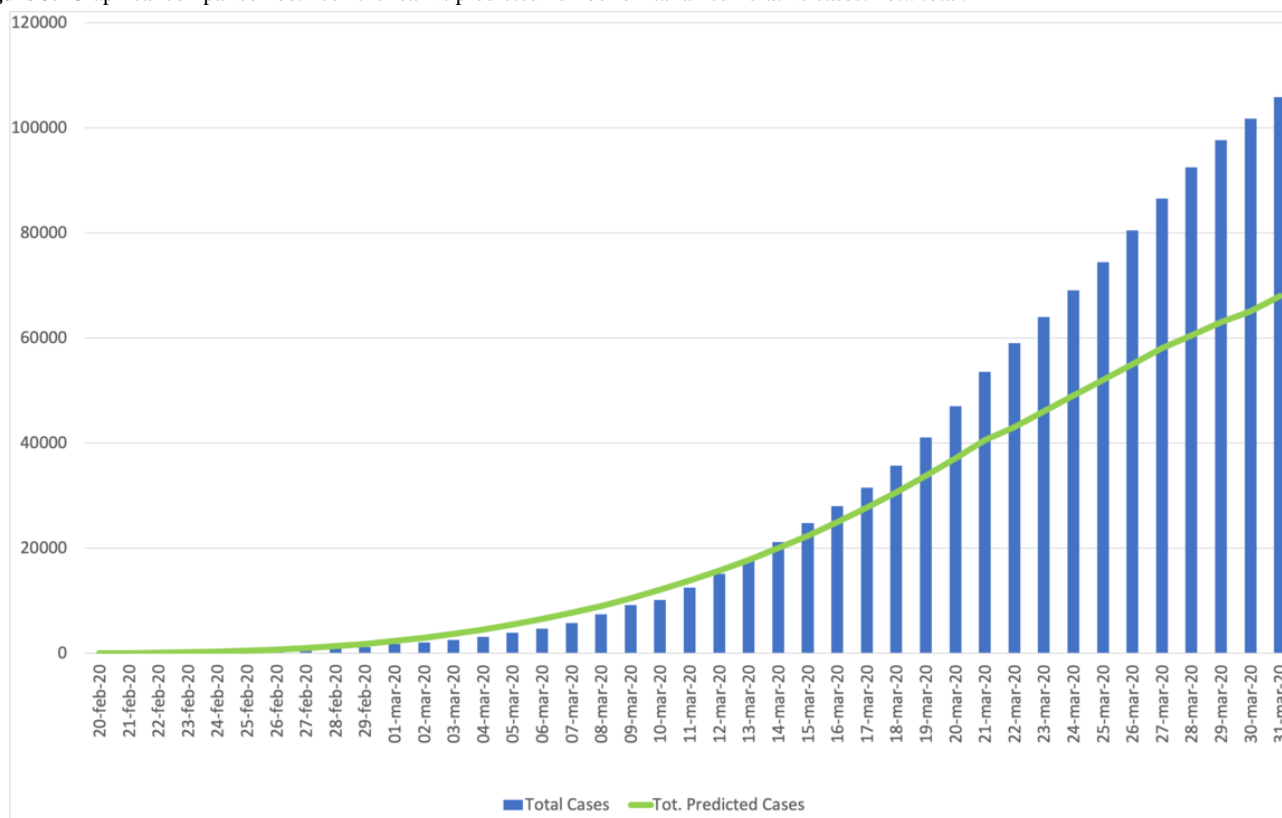
(Lodi, Lombardy) on February 24, 2020. The dotted orange line is the classical exponential curve typical of epidemics, while the green line is the best fit for the data set available for Italy until March 2, that is a power curve in the form of  $y = 3.00 * x^{2.77}$ .

The goodness of fit for the detected model is high, with a high coefficient of determination ( $R^2=0.9917$ ).

Figure 3 compares the actual behavior of cumulative COVID-19 positive cases with the predictions of our model over time until the end of March (where our model predicted the logistic plateau).

**Figure 2.** The predicted curve of total positive cases in Italy (as of March 2, 2020).



**Figure 3.** Graphical comparison between the real vs predicted number of Italian cumulative cases. Tot.: total.

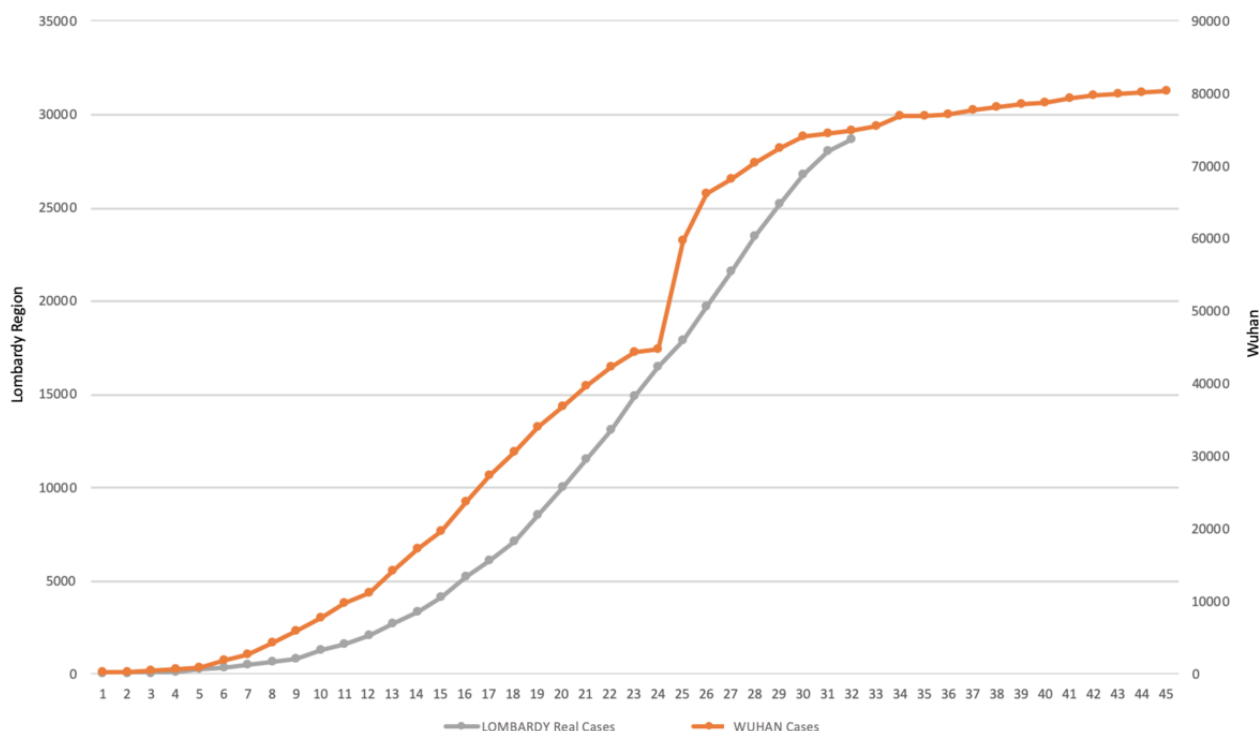
Our model forecasted (18 days in advance) the peak in the number of daily new cases on March 21, 2020, with a total of 42,000 positive cases against the official datum of 53,500 total positive cases, with a confidence level of 95%. The model significantly outperformed other predictions based on exponential models that forecasted more than 180,000 positive cases.

As highlighted in Figure 3, the model underestimated some of the real official values for several reasons:

- March 21, 2020, is registered as the date with the absolute peak in the number of daily new cases, hence the error (that is a cumulative error) is increasing over time.
- The Lombardy Region has the same consideration (where the peak has been forecasted and actually observed on March 17). The Lombardy Region accounts for 50% of the overall national value.

- The predictive model is strongly related to the curve of Wuhan. Although the increase of daily cases is similar between the Italian curve and the Wuhan curve, the decrease is a bit different. The Italian one is less steep than the Wuhan decrease since the restrictions implemented by the Italian government are less stringent than the Wuhan restrictions, and the Italian population reacted to the restrictions with less determination.

Moreover, it is important to highlight that (as depicted in Figure 4) the plateau for the Lombardy Region was predicted and confirmed to start at the end of March, a few days in advance of the plateau for the Italian curve, hence the absolute contribution of the Lombardy Region on the total national value was decreasing over time. It is important to observe that the Lombardy Region anticipated all the national restrictions by 1 week with the Decree to Manage the COVID-19 Emergency: DPCM-9 March 4, 2020.

**Figure 4.** Graphical comparison between the Lombardy Region curve and the Wuhan curve (total cumulative positive cases).

The model elaborated from the beginning of March 2020 (Figures 1-3) predicted the plateau (as confirmed by Prof S Brusaferrò from Istituto Superiore Sanità during a national live TV broadcast [12]) for the logistic Italian curve at the end of March with an estimate on the total positive cases to be 68,000 (while the actual value was 100,000 total cases). It is important to highlight that the model focused on the prediction of the pandemic timeline and not on the exact number of people officially testing positive for COVID-19, a number that is well known to be dependent on the number of analyzed swab tests. We based our initial estimation on the number of swab tests analyzed in the initial days of the pandemic period. The increasing amount of swab tests moved our curves toward higher values but did not modify the prediction regarding the starting day of the logistic curve plateau and the prediction about the different phases of the COVID-19 pandemic. As previously stated, another factor that contributes to the larger number of affected people is the different enforcement and timeliness applied to the restrictions and lockdowns by the Italian government and the Italian population compared to China, which we adopted in our initial assumptions. In any case, in the forecast of the total positive cases in a pandemic phenomena, the magnitude of the values is the real discriminant on the quality of the prediction. Exponential curves (adopted by several scientists at the beginning of the COVID-19 pandemic in Italy) predicted 1 million cases for the end of March, with a ten-fold overestimation, while our model underestimated for a value less than 40%. Ricolfi [13] estimated for March 8, 2020, 60,000 total cases, while the official value was 7400, thus an overestimation of 800%. In Italy, we have 60 million inhabitants; a difference of 32,000 cases between the estimated value and the actual value on this population introduces an error that is

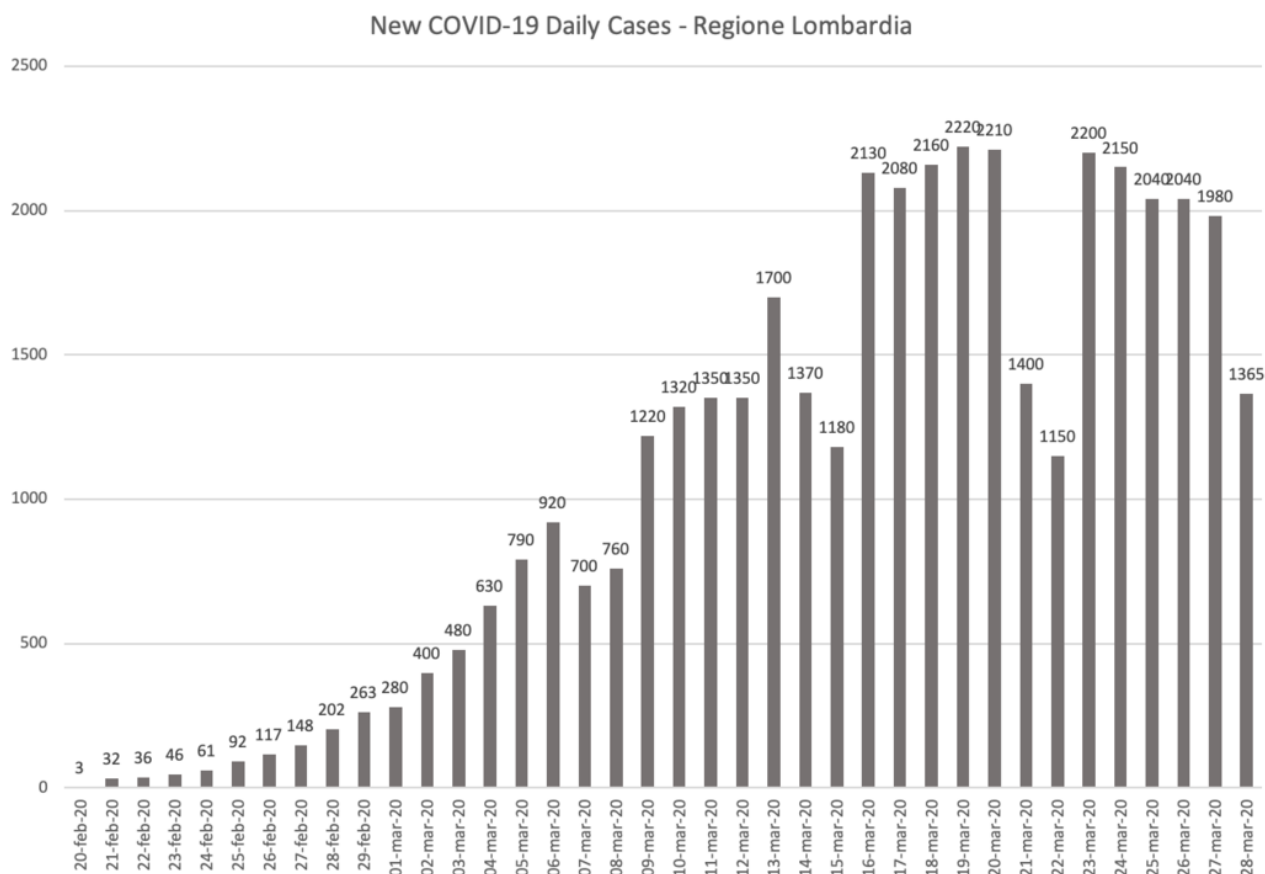
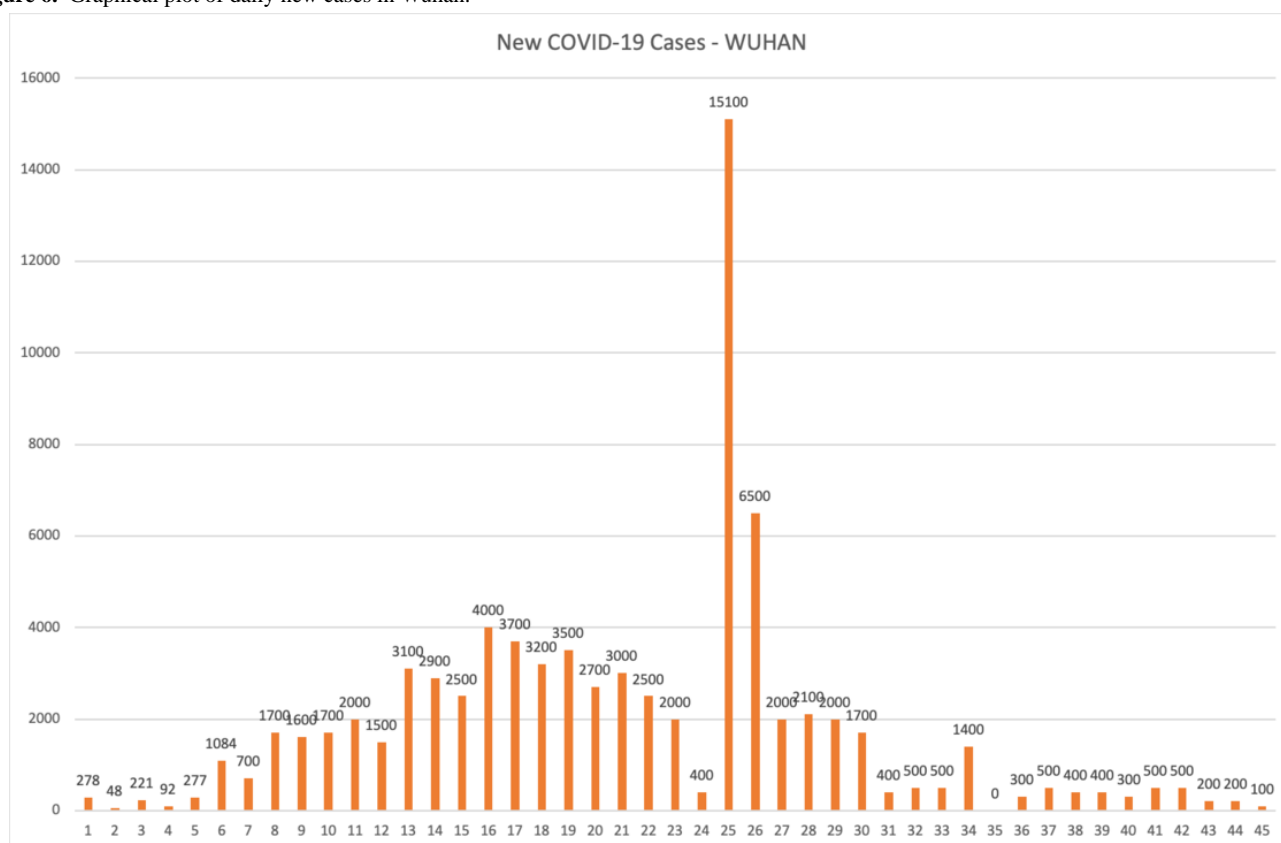
less than 0.08%, and this does not impact the governmental policies and actions to preserve the national health system.

It is important to observe that reaching the plateau does not indicate that the COVID-19 epidemic has been solved, but it means that the cumulative number of COVID-19 positive cases is slowing down in its ascent. The WHO guidelines suggest waiting for the contagion to be +0 (ie, the new daily cases are reduced to a few dozen positive cases per day) and then taking restrictions for two additional cycles of COVID-19 incubation (mean incubation period 5-6 days, range 1-14 days.) Hence, considering the national plateau on March 31, 2020, we estimated the actual containment of the COVID-19 epidemic at the beginning of June for the following reasons:

- One additional month, after the beginning of the plateau, to reach a small and contained number of new daily cases (end of April)
- One additional month in waiting for the two COVID-19 incubation cycles (end of May)
- Gradual return to normal life starting from the beginning of June

In Figure 4, the two curves for total cases in Wuhan (orange line) and the Lombardy Region (gray line) are depicted. It is interesting to observe that in this case the two behaviors are similar and comparable. Figure 4 also shows that we are approaching the plateau for the Lombardy Region.

In Figures 5 and 6, we show the new COVID-19 cases per day in the Lombardy Region and Wuhan, respectively. As previously mentioned, the peak in the Lombardy Region was registered on March 17, 2020, with 2200 new daily cases; then, the number of daily cases decreased over time (with some exception as you can see in Figure 5).

**Figure 5.** Graphical plot of daily new cases in the Lombardy Region.**Figure 6.** Graphical plot of daily new cases in Wuhan.

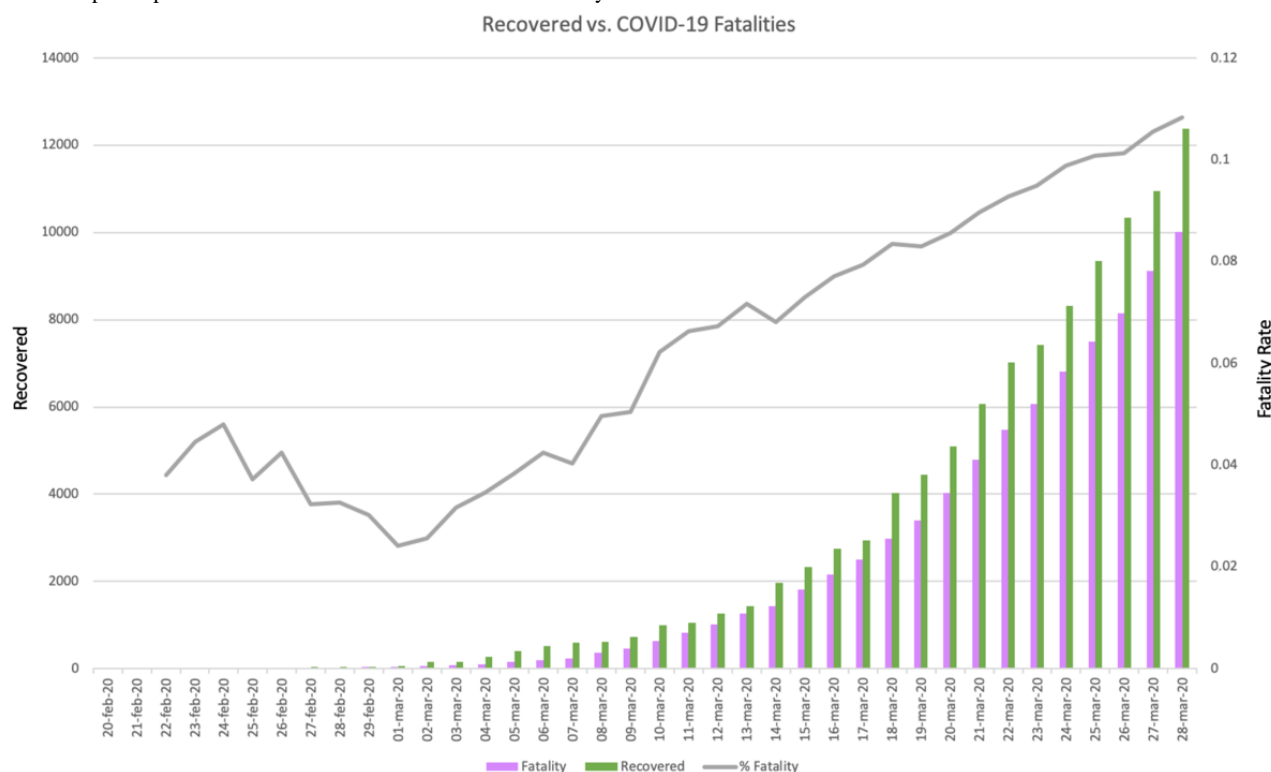


This is a good indicator (the peak matches the inflection point in the logistic curve) that we had started the descent toward the plateau in the Lombardy Region (the last inflection point in the logistic curve).

Additionally, in that case, it is important to observe that the curve of fatalities and recoveries (as depicted in Figure 7) were misaligned temporally with the curve of new cases, since

fatalities occur on average after 8 days, while recoveries are an additional 14 days, and new cases are observed after 8 days of COVID-19 incubation (on average). This explains why the peak in the number of new cases had been reached, while the peak on daily fatalities was approaching, and we expected it at the end of March. The same is true for the peak of recoveries that we expected to appear in 2 more weeks (middle or end of April).

**Figure 7.** Graphical plot of COVID-19 fatalities and recovered in Italy.



## Discussion

This study was conducted in the early days of the pandemic in Italy to promptly define a model able to predict the curve of total positive cases in Italy and the Lombardy Region.

The model predicted the real data published daily by the Department of the Italian Civil Protection, estimating in a precise manner and several months in advance the plateau for both the logistic curves for the Lombardy Region and Italy, and the end of this first COVID-19 pandemic wave. This suggests

the possibility to generalize the model for other countries, which will follow the restrictions imposed by the Italian government, to have a clear picture on the evolution of the number of new cases and to act promptly with policies and restrictions that can maximize care and treatments offered to patients with COVID-19. Moreover, this paper shows that big data and data analytics can help medical experts and epidemiologists in promptly designing accurate models to predict the different COVID-19 evolutionary phases in other countries and regions, and for second and third possible epidemic waves.

## Acknowledgments

This research is partially funded by the European Research Council Advanced Grant project 693174 GeCo (Data-Driven Genomic Computing), 2016-2021.

## Conflicts of Interest

None declared.

## References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020 Feb 20;382(8):727-733. [doi: [10.1056/nejmoa2001017](https://doi.org/10.1056/nejmoa2001017)]

2. Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19 — studies needed. *N Engl J Med* 2020 Mar 26;382(13):1194-1196. [doi: [10.1056/nejmp2002125](https://doi.org/10.1056/nejmp2002125)]
3. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg Infect Dis* 2020 Jul;26(7):1470-1477. [doi: [10.3201/eid2607.200282](https://doi.org/10.3201/eid2607.200282)] [Medline: [32255761](https://pubmed.ncbi.nlm.nih.gov/32255761/)]
4. Tosi D, Verde A, Verde M. Clarification of misleading perceptions of COVID-19 fatality and testing rates in Italy: data analysis. *J Med Internet Res* 2020 Jun 17;22(6):e19825 [FREE Full text] [doi: [10.2196/19825](https://doi.org/10.2196/19825)] [Medline: [32490842](https://pubmed.ncbi.nlm.nih.gov/32490842/)]
5. COVID-19 coronavirus pandemic. Worldometer. 2020. URL: <https://www.worldometers.info/coronavirus/>
6. Tsang TK, Wu P, Lin Y, Lau EHY, Leung GM, Cowling BJ. Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *Lancet Public Health* 2020 May;5(5):e289-e296. [doi: [10.1016/s2468-2667\(20\)30089-x](https://doi.org/10.1016/s2468-2667(20)30089-x)]
7. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020 Mar 26;382(13):1199-1207. [doi: [10.1056/nejmoa2001316](https://doi.org/10.1056/nejmoa2001316)]
8. Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020 Feb;395(10225):689-697. [doi: [10.1016/s0140-6736\(20\)30260-9](https://doi.org/10.1016/s0140-6736(20)30260-9)]
9. Tosi D. Cell phone big data to compute mobility scenarios for future smart cities. *Int J Data Sci Anal* 2017 Jun 28;4(4):265-284. [doi: [10.1007/s41060-017-0061-2](https://doi.org/10.1007/s41060-017-0061-2)]
10. Department of the Italian Civil Protection. URL: <http://www.protezionecivile.gov.it/>
11. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). World Health Organization. 2020 Feb. URL: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>
12. Del Frate C. Coronavirus, Brusaferrò: «Stiamo arrivando al plateau dei contagi, indice contagiosità intorno a 1». *Corriere della Sera*. 2020 Mar 31. URL: [https://www.corriere.it/cronache/20\\_marzo\\_31/coronavirus-brusaferr%C3%B2-stiamo-arrivando-plateau-contagi-82cbb51c-734e-11ea-bc49-338bb9c7b205\\_preview.shtml](https://www.corriere.it/cronache/20_marzo_31/coronavirus-brusaferr%C3%B2-stiamo-arrivando-plateau-contagi-82cbb51c-734e-11ea-bc49-338bb9c7b205_preview.shtml)
13. Ricolfi L. Coronavirus, calcoli sbagliati: le gravi responsabilità del governo. *Il Messaggero*. 2020 Mar 05. URL: [https://www.ilmessaggero.it/editoriali/luca\\_ricolfi/coronavirus\\_italia\\_errori\\_governo\\_calcoli\\_morti\\_contagi-5091141.html](https://www.ilmessaggero.it/editoriali/luca_ricolfi/coronavirus_italia_errori_governo_calcoli_morti_contagi-5091141.html)

## Abbreviations

**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 04.06.20; peer-reviewed by G Aiello, MC De Bernardi; comments to author 08.07.20; revised version received 10.07.20; accepted 23.07.20; published 14.10.20.*

*Please cite as:*

*Tosi D, Campi A*

*How Data Analytics and Big Data Can Help Scientists in Managing COVID-19 Diffusion: Modeling Study to Predict the COVID-19 Diffusion in Italy and the Lombardy Region*

*J Med Internet Res* 2020;22(10):e21081

URL: <http://www.jmir.org/2020/10/e21081/>

doi: [10.2196/21081](https://doi.org/10.2196/21081)

PMID: [33027038](https://pubmed.ncbi.nlm.nih.gov/33027038/)

©Davide Tosi, Alessandro Campi. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org>, as well as this copyright and license information must be included.

## Original Paper

# Online Health Information Seeking Using “#COVID-19 Patient Seeking Help” on Weibo in Wuhan, China: Descriptive Study

Xiaoman Zhao<sup>1,2</sup>, PhD; Ju Fan<sup>3</sup>, PhD; Iccha Basnyat<sup>4</sup>, PhD; Baijing Hu<sup>2</sup>, PhD

<sup>1</sup>Research Center of Journalism and Social Development, Renmin University of China, Beijing, China

<sup>2</sup>School of Journalism and Communication, Renmin University of China, Beijing, China

<sup>3</sup>Data Engineering and Knowledge Engineering Lab & School of Information, Renmin University of China, Beijing, China

<sup>4</sup>School of Communication Studies, James Madison University, Harrisonburg, VA, United States

**Corresponding Author:**

Baijing Hu, PhD

School of Journalism and Communication

Renmin University of China

59 Zhongguancun Street

Beijing

China

Phone: 86 01062514835

Email: [hubaijing@vip.sina.com](mailto:hubaijing@vip.sina.com)

## Abstract

**Background:** First detected in Wuhan, China in December 2019, the COVID-19 pandemic stretched the medical system in Wuhan and posed a challenge to the state's risk communication efforts. Timely access to quality health care information during outbreaks of infectious diseases can be effective to curtail the spread of disease and feelings of anxiety. Although existing studies have extended our knowledge about online health information-seeking behavior, processes, and motivations, rarely have the findings been applied to an outbreak. Moreover, there is relatively little recent research on how people in China are using the internet for seeking health information during a pandemic.

**Objective:** The aim of this study is to explore how people in China are using the internet for seeking health information during a pandemic. Drawing on previous research of online health information seeking, this study asks the following research questions: how was the “#COVID-19 Patient Seeking Help” hashtag being used by patients in Wuhan seeking health information on Weibo at the peak of the outbreak? and what kinds of health information were patients in Wuhan seeking on Weibo at the peak of the outbreak?

**Methods:** Using entity identification and textual analysis on 10,908 posts on Weibo, we identified 1496 patients with COVID-19 using “#COVID-19 Patient Seeking Help” and explored their online health information-seeking behavior.

**Results:** The curve of the hashtag posting provided a dynamic picture of public attention to the COVID-19 pandemic. Many patients faced difficulties accessing offline health care services. In general, our findings confirmed that the internet is used by the Chinese public as an important source of health information. The lockdown policy was found to cut off the patients' social support network, preventing them from seeking help from family members. The ability to seek information and help online, especially for those with young children or older adult members during the pandemic. A high proportion of female users were seeking health information and help for their parents or for older adults at home. The most searched information included accessing medical treatment, managing self-quarantine, and offline to online support.

**Conclusions:** Overall, the findings contribute to our understanding of health information-seeking behaviors during an outbreak and highlight the importance of paying attention to the information needs of vulnerable groups and the role social media may play.

(*J Med Internet Res* 2020;22(10):e22910) doi:[10.2196/22910](https://doi.org/10.2196/22910)

**KEYWORDS**

COVID-19; coronavirus; information seeking; social media; Wuhan

## Introduction

### Background

On March 11, 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a pandemic [1]. First detected in Wuhan, China in December 2019, the disease rapidly spread into more than 100 locations internationally, including Japan, Korea, and the United States [1]. In the Report of the WHO-China Joint Mission on COVID-19, the Joint Mission reminded the public that the virus is unique in its ability to cause societal and economic disruption [2]. In China, the disease has caused 3176 deaths out of the 80,813 total confirmed cases by March 12, 2020 [3]. The rapid increase in coronavirus patients stretched the medical system in Wuhan, which accounts for 60% of mainland China's total confirmed cases and about 77% of the deaths [4]. In an attempt to limit the spread of the virus, Chinese government has enforced an unprecedented lockdown on Wuhan by suspending transport both within the city and leaving the city [5]. During the quarantine, each household was allowed to have only one person go out for necessities every 2 days [6]. The traffic ban within the city has made it hard for patients to seek health care [7]. It was also difficult to organize medical care, frequently monitor disease progression, and refer hospital care for patients in home isolation in a timely manner [8]. The lockdowns caused public panic and resulted in many cross-infections at the crowded, overwhelmed hospitals [9], posing a challenge to the state's risk communication efforts. The dramatic increase in the number of infected individuals was causing a burden on the medical system [10].

Timely access to quality health care information during outbreaks of infectious diseases can be effective to curtail the spread of disease and feelings of anxiety [11,12]. Up-to-date information about specific threats and necessary precautionary measures was found to mitigate public anxiety, reduce morbidity and mortality, and contribute to minimizing negative mental impacts [13,14]. Furthermore, previous studies relate health anxiety, internet literacy, and chronic conditions to the willingness of individuals to engage in health information-seeking behaviors [15-17]. Health information-seeking behavior is purposeful activities such as searching for condition-specific information as well as disease prevention and treatment information to fulfill specific health information need [18,19]. An information need can arise when a patient experiences health-related uncertainty originating from an inaccurate, inconsistent, ambiguous, or excessive provision of information about the diagnosis, treatment, or aspects of medical decision making [16,20,21]. However, relatively little is known about what happens to the information-seeking behavior of patients during an outbreak, especially in the context of China. Therefore, this study explores the information-seeking behaviors during the coronavirus outbreak in China.

The internet, especially social media, has also been identified as a significant source for information searching and decision making [22]. Scholars suggest potential in the internet to supplement traditional sources of health information and to support patients' decision making [22,23]. In fact, health information seeking has been found to be a popular online

activity [24]. Thus, researchers have suggested that an analysis of web behaviors can provide insights into individuals' information seeking during an outbreak, as public reactions are visible more quickly online [12,25]. For instance, by monitoring web activities, previous studies identified large increases in activities on social networking websites, including posting and searching, that are closely associated with the outbreaks of pandemic diseases [26,27]. Text analysis programs that are developed for measuring emotional expression in natural language are also found to produce reliable results that are congruent with human ratings [28,29]. Using web-based text analysis to monitor public emotions is also suggested by researchers to avoid self-report biases from social desirability effects or memory distortions [12]. Therefore, using entity identification and text analysis on a total of 10,908 posts on Weibo, the largest social networking platform in China, we identified 1496 patients with COVID-19 living in or with family in Wuhan, China and explored their online health information-seeking behavior during a pandemic.

### Online Health Information Seeking

Health information seeking has been found to be a popular online activity. Studies in the United States and Europe have reported more than 70% of internet users have looked online for health information of one kind or another [24], or having used the internet for health purposes [30]. The motivations for seeking online health information are diverse, including self-diagnosing, coping with uncertainty, staying informed on preventing diseases, and looking for others with a similar health concern [24,31]. Through query construction and information source selection, information seekers can enjoy greater control over information acquisition processes and achieve desired levels of uncertainty [32]. In view of the features such as convenience, cost effectiveness, and private sharing, scholars suggest potential in the internet to supplement traditional sources of health information and support patients' decision making [22,23].

Previous studies have identified multiple factors that may influence patients' motivation to seek health information online, including biological sex, income, age, chronic illness, and travel time to offline sources of health care [16,17,33,34]. For instance, biological sex was found to significantly predict online health information seeking, with females more likely to seek out online health information than males—perhaps because females often take on primary caregiving roles in families and are more cautious in risk contexts [16,35]. Increased age is frequently associated with decreased levels of motivation for health-related information seeking online, as older adults are always found to have lower levels of internet literacy and experience more difficulties navigating websites [36,37]. Further, the digital divide, a gap between individuals from different socioeconomic backgrounds with regard to their access to and use of digital equipment and services, can increase challenges associated with online health information seeking, such as inequality of accessibility and difficulties to differentiate between high and low quality resources [17,21].

Additional factors, such as efficacious feelings about using the internet and health anxiety, may also influence one's motivation



of seeking out health information online [16,34,38]. User experience online was also found to influence one's feelings of efficacy and, thus, be linked to their likelihood to use online search strategies for health-related information [34]. Adding on to self-efficacy and health anxiety is chronic illness, as individuals with a chronic illness were more likely to use the internet to search for health information compared to those without a chronic illness [39]. A long travel time to offline sources of health care was also associated with a stronger likelihood of using the internet to find health information [17].

Although these existing studies have extended our knowledge about online health information-seeking behavior, processes, and motivations, rarely have the findings been applied to an outbreak. Therefore, in this study, we examine the ways in which patients with COVID-19 and families living in Wuhan, China used the internet to seek health information on the social media platform Weibo. The findings can help build an understanding of how the internet can be used to better serve the needs of the public, especially the patients at the peak of the outbreak, given the high level of uncertainty and risks.

### Online Health Information Seeking in China

In China, people are found to face challenges in accessing health care resources, among which are the difficulty in making medical appointments, short consultation times, and a significant socioeconomic disparity in health literacy [40,41]. Despite the governmental attempt to reform its health care system, patients still expressed unsatisfied needs for various kinds of health-related information, including the treatment of diseases and the effect, etiology, and risk factors, as well as use, of drugs and medication [42]. The barriers to accessing accurate information and subsequent health care also lead to people's lack of trust in doctors and their unwillingness to visit them [42]. In this context, the internet has been increasingly used to access health information, supplementing the traditional sources of health information [43,44].

According to statistics from the government-run China Internet Network Information Centre, the number of internet users in the country had skyrocketed to 854 million at the end of June 2019, with the internet penetration rate reaching 61.2% [45]. Although 18.2% of the total Chinese population are 60 years or older [46], users older than 60 years accounted for only 6.9% of all users on the internet [45]. A survey study found that 36.7% of the Chinese participants had sought online health information at least one or two times, citing saving money and easing the privacy concern as the two major reasons for seeking out health information online [43]. Despite the increase in health-related internet use in China, there is relatively little recent research on how people in China are using the internet for seeking health information. Almost all the studies were limited to individuals that were young and educated with certain levels of online health information-seeking experience, mostly in Hong Kong [42-44,47]. According to these studies, the overall health literacy level in China is lower than that in Western countries [48].

The majority of online health information was found to be of poor quality, and the functions of health websites were ineffective and hard to navigate [43]. However, information seekers in China have still been found to consider the internet

a highly reliable source of information [49]. Although some studies reported no correlations between education or gender on Chinese patients' online health information-seeking behaviors [42,50], some other studies have identified digital inequalities associated with education level, household income, and socioeconomic status, further leading to variations in personal health condition and family well-being [44,51]. These disagreements in the findings point to the need to conduct further research on how people in China, especially the general population in mainland China, are using the internet for health information, particularly during a pandemic. Therefore, drawing on previous research of online health information seeking, this study asks the following research questions: How was the "#COVID-19 Patient Seeking Help" hashtag being used by patients in Wuhan seeking health information on Weibo at the peak of the outbreak? What kinds of health information were patients in Wuhan seeking on Weibo at the peak of the outbreak?

## Methods

### Data Collection

During the COVID-19 outbreak, Weibo, the largest social networking platform in China, created a hashtag named #COVID-19 Patient Seeking Help ("Feiyan Huanzhe Qiuizhu Chaohua") for the patients and their families to leave their name, age, city, neighborhood, address, time of sickness, health condition, additional description, and contact information, making the posts structured. For the purpose of identity verification, patients were also asked to upload pictures of a medical examination, if any, which further improved the credibility of the data. We crawled and analyzed Weibo posts with this hashtag published from January 29, 2020, when the hashtag was first created, to February 17, 2020, to examine the online health information-seeking behaviors of patients with COVID-19. This period of 20 days was chosen because, by the end of this period, the number of patients posting with this hashtag fell to zero. In total, 10,908 Weibo post entries with the #COVID-19 Patient Seeking Help hashtag were collected.

For each post, in addition to the structured patient information previously noted, we further extracted the following items: *the date and time of posting, user ID, user gender, and URL of the thread* (for further referring back to the entry online). We excluded retweets and general comments about the outbreak. After this step, we obtained 4983 entries of patients with COVID-19 in Wuhan using the hashtag #COVID-19 Patient Seeking Help. Data consolidation was further carried out by patient name and detailed address, which resulted in 1496 unique patient cases.

We further standardized the patients' information that was crawled, as the language used on the social media platform was flexible. In particular, to standardize the patients' detailed address, we crawled a full list of housing estates in Wuhan, from the largest housing estates website in China, Lianjia [52], and mapped the original Weibo texts with patient's address to the housing estate names. We further obtained the longitudes and latitudes of the patients' locations through the Baidu map application programming interface (API) [53].



## Data Analysis

Previous studies identified multiple factors that may influence patients' motivation to seek out health information online, including biological sex, age, chronic illness, and travel time to offline sources of health care [16,17,33,34]. To answer the first research question, we examined the age of the patients. The original posts did not include patient gender, but we examined the gender of the posting users, as previous studies have found that about half of individuals' health information searches are on behalf of someone else's health situation [24,54]. We also examined the patients' underlying condition by extracting the health condition description in each entry. Specifically, we conducted word segmentation using the Chinese word segmentation module, Jieba [55]; computed document frequency for every single term that appeared in the content; and identified the terms indicating their underlying diseases.

We also examined the patients' shortest walking distance to offline sources of health care. Specifically, we extracted the list of 42 fever clinics and the 28 designated hospitals that was first published by the Health Commission of Hubei Province [56] and further updated by the Hubei Provincial People's Government [57]. Through Baidu map API, we further obtained the longitudes and latitudes of these clinics and hospitals, and calculated the shortest walking distance of each patient's location to the nearest fever clinic or designated hospital, as public transportation and private car driving in the city were prohibited during the period of our study, and patients may have faced difficulties seeking medical care due to the traffic ban [7].

To answer our second research question, we examined their information-seeking behavior, as indicated by the number of entries posted by each patient or user. We also examined the specific information they were seeking out by analyzing the content of health condition and additional description in each entry, where the patients had given more details about their needs. We carried out textual analysis on every post to identify the information needs of the patients. Specifically, thematic analysis was carried out. Open coding, the first step of the coding process entailed reading each entry and its messages, highlighting salient phrases and words [58,59]. At this stage, one of the authors conducted the open coding, reading the entries

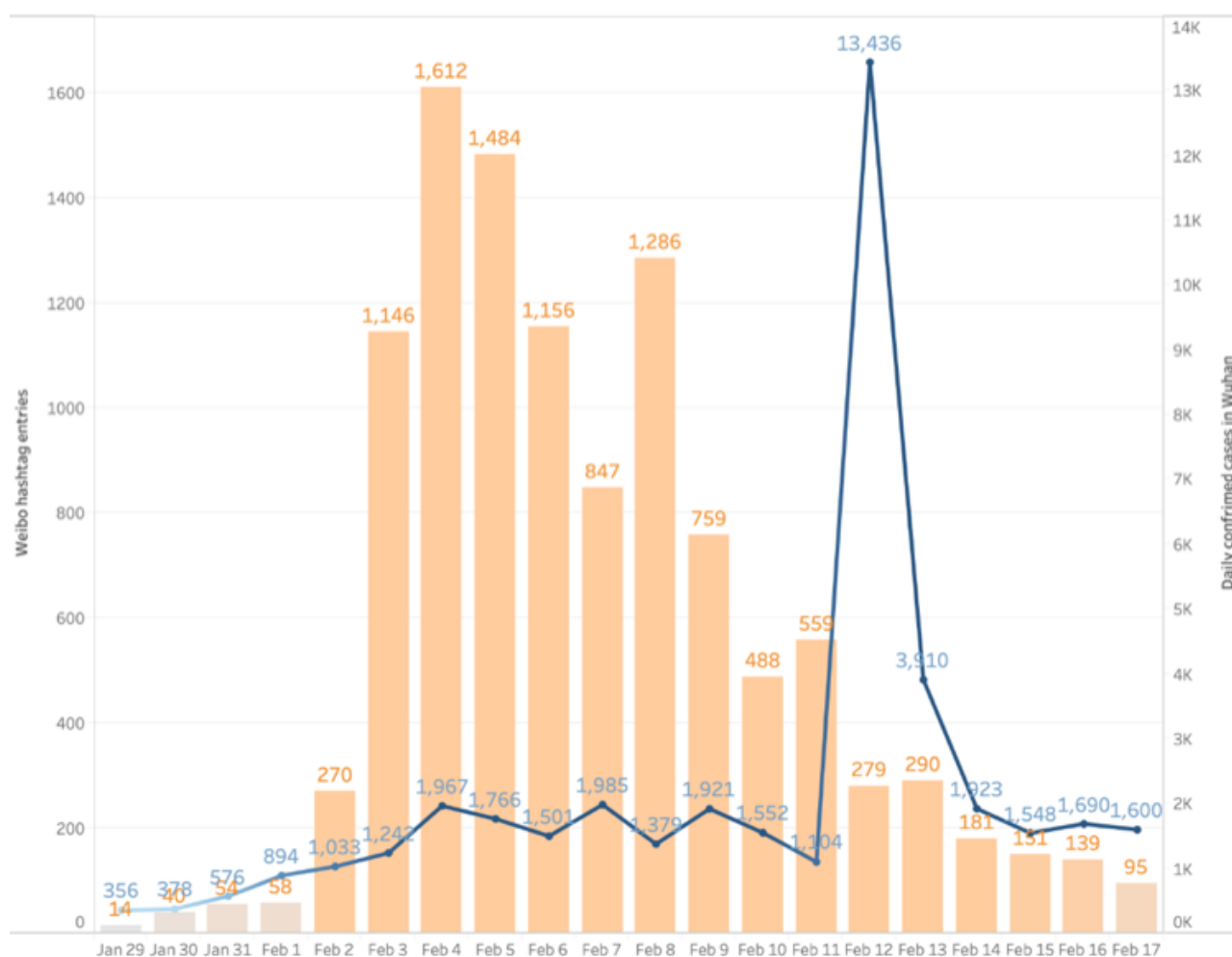
in Chinese. More than 200 open codes were generated in this process. Some examples of open codes that emerged were queuing for test, called every hospital for bed, no foreseeable treatment, staying overnight for injection, the hospital is full of patients, waiting for confirmatory testing result, rushing between different hospitals, hundreds of people in the waiting list, reported to the neighborhood committee with no response, and have to find solutions by ourselves. Next, the open codes and phrases were conceptually clustered into more than 30 different axial code groupings through discussion among the authors [59]. Keeping the research question in mind, through additional discussion and data refinement, the authors agreed that the following themes best answered the research question: accessing medical treatment, managing self-quarantine, and accessing tangible support.

## Results

Figure 1 presents the numbers of Weibo posts with the #COVID-19 Patient Seeking Help hashtag and the daily number of confirmed cases in Wuhan reported by the Health Commission of Hubei Province. Figure 1 illustrates that the number of entries rapidly grew from February 3, 2020, and maintained a high level until February 12. On February 12, the daily confirmed cases peaked at 13,436, at which time the central government promised to admit all the patients with COVID-19 [60]. The number of the hashtag entries have steadily declined since then.

Table 1 is an age comparison between our sample and that of the WHO-China Joint Mission [2]. The WHO-China Joint Mission identified the median age of patients as 51 years based on a total of 55,924 confirmed cases. However, for our sample, the median age was 61 (IQR 50-70) years, with an average age of 59 years. According to the report published by the WHO-China Joint Mission, individuals 60 years or older are at highest risk for severe disease and death. Previous studies have also shown an age-related digital divide in China. Although 18.2% of the total Chinese population were 60 years or older by the end of 2019 [46], users older than 60 years accounted for only 6.9% of all users of the internet [45]. The age-related digital divide might prevent the patients from seeking information and help online.

**Figure 1.** Daily numbers of #COVID-19 Patient Seeking Help hashtag entries (the orange bar plot) and daily confirmed cases in Wuhan (the blue line plot).



**Table 1.** Age comparison between our sample and that of the WHO-China Joint Mission.<sup>a</sup>

Sample	Participants, n	Age range	Age IQR (years)	Age median (years)
WHO <sup>b</sup> -China Joint Mission	55,924	2 days-100 years	39-63	51
Our sample	1454	2-99 years	50-70	61

<sup>a</sup>Age was missing in 42 Weibo entries.

<sup>b</sup>WHO: World Health Organization.

Table 2 shows the document frequency of family members' names that appeared in the posts. Out of 883 entries, words such as "mum (mother)," "dad (father)," and "elder at home" appeared in 35% (n=308), 29% (n=255), and 24% (n=209) of the entries, respectively. Further examination of these entries showed that most of the entries were posted by the younger generation for their parents or an older adult at home. We examined the gender of the users seeking out health information.

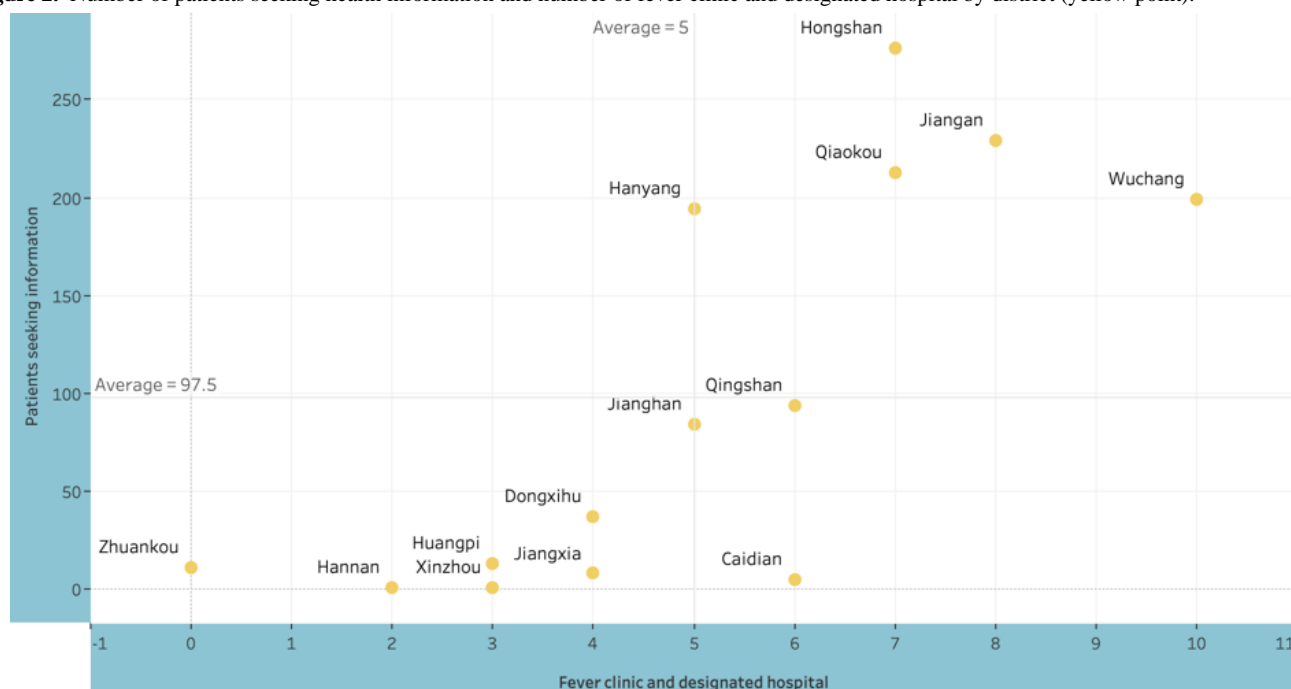
Among the 2405 unique users, 69% (n=1660) were female and only 31% (n=745) were male. The median number of posting times by female users was 2 (IQR 1-2) times, which was higher than that of male users, 1 (IQR 1-2) time. It shows that there were more female users than male users trying to seek out health information online, and their internet use frequency might be higher than their male counterpart.

**Table 2.** Document frequency of names of family members.

Word	Frequency (n=883), n (%)	Quote
Mother (mum)	308 (35)	"My <i>mother</i> is highly suspected with COVID-19. She could not even get up right now...She has been running a fever for more than ten days..."
Father (dad)	255 (29)	"I am the son of the patient. My <i>father</i> has been infected with pneumonia. And the lung lesions are quite serious. I have reported to the neighborhood committee for many days, but they have not arranged hospital bed for us..."
Elder at home	209 (24)	"The <i>elder at home</i> have been diagnosed with the pneumonia...Please contact his daughter as the <i>elder at home</i> do not use the Internet."
Grandma	88 (10)	"Now my <i>grandma</i> is already in incontinence, but the neighborhood committee still asks us to wait."
Parents	80 (9)	"My <i>parents</i> are both confirmed and in dangerous condition. But we haven't received hospitalization notification. Please save my <i>parents</i> ."
Grandpa	62 (7)	"The patient is my <i>grandpa</i> , who is in dangerous condition. He has emphysema and threatening myocardial infarction."
Aunt	57 (6)	"The <i>aunt</i> is currently in recurring fever. We are all in desperation."
Uncle	52 (6)	"My <i>uncle</i> 's condition is worsening in self-quarantine, with eating and breath difficulties."

We further examined the patients' distance from their residential locations to offline health care. We first extracted the district of each patient's residential location to see the distribution of these patients across different districts of Wuhan. [Figure 2](#) shows

the number of patients in our sample by district. Districts of Hongshan, Jiangnan, and Qiaokou were found to have the most patients seeking information online.

**Figure 2.** Number of patients seeking health information and number of fever clinic and designated hospital by district (yellow point).

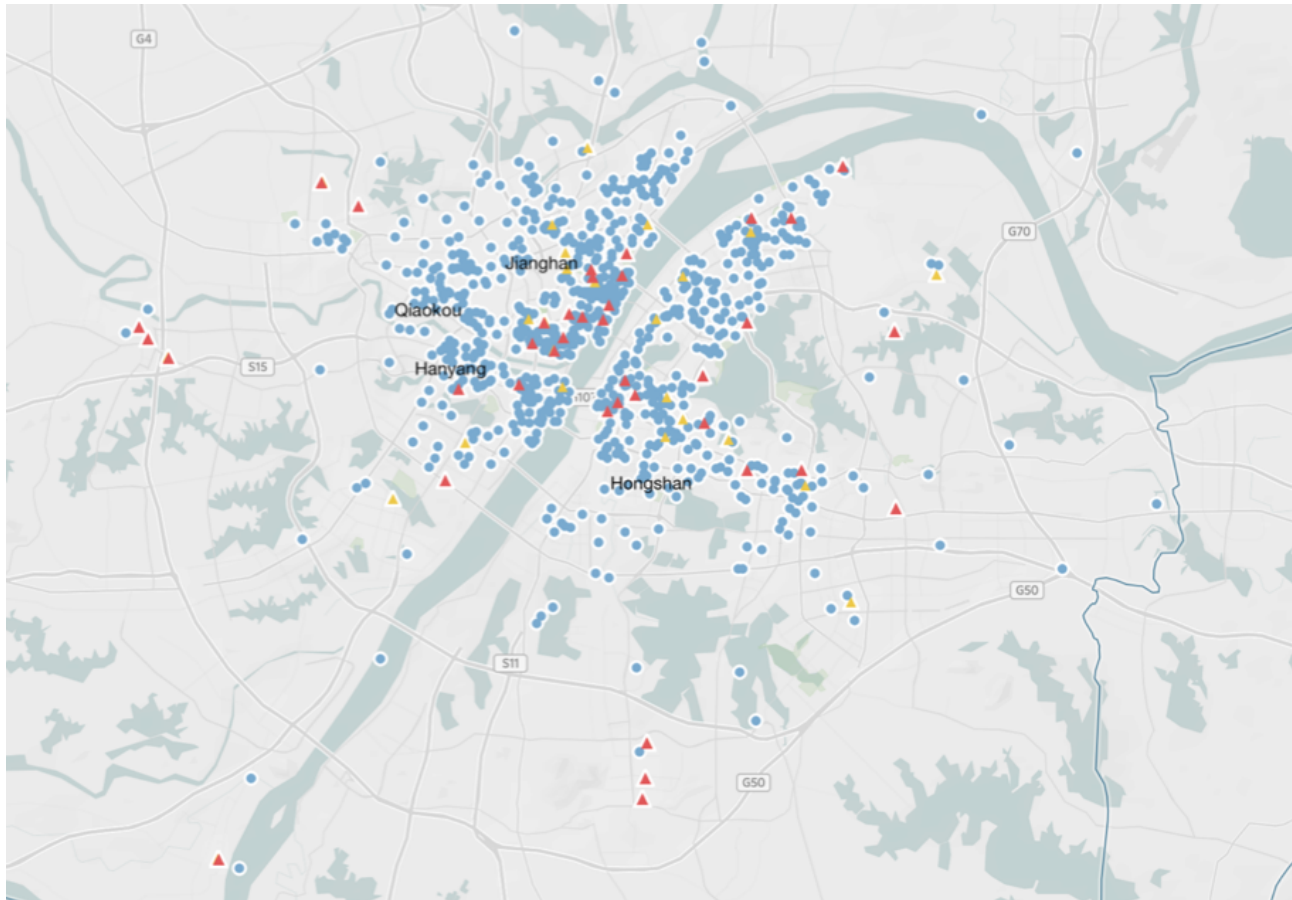
We also examined the geographic distribution of the fever clinics and designated hospitals in relation with the patients' residential locations. [Figure 3](#) is an overview of all the patients' locations, the fever clinics, and the designated hospitals on a map of Wuhan City. Seen from the map, in the districts of Hongshan, Jiangnan, and Qiaokou, some patients were far from the fever clinics and designated hospitals.

To further examine the exact distance of the patients to offline health care resources, we computed the distance between every residential location to its nearest fever clinic or designated hospital. [Table 3](#) maps the patients' distance by district to the

nearest fever clinic or designated hospital, and the average distance was 2.67 (SD 2.88) km. According to estimation by the Baidu map API, it takes 50 minutes to walk 3 km by an adult. Considering the average age of our sample was older, their actual walking time could be even longer. This means that 27.6% (n=413) of the sample had to walk for around 1 hour to access offline health care. In the context of our study, distance to offline health care is especially important. Due to the transport prohibition and quarantine enforced, patients could only walk to access health care offline. The difficulties to travel were also

cited by many users as a major reason for going online to seek health information as identified in textual analysis.

**Figure 3.** Overview of the locations of patient residential addresses (blue points), fever clinics (red triangle), and designated hospitals (yellow triangle).



**Table 3.** Patients' distance to the nearest fever clinic or designated hospital.

District	<1 km distance, n	1-2 km distance, n	2-3 km distance, n	>3 km distance, n
Qiaokou	68	17	4	124
Hongshan	28	66	67	116
Jiangnan	68	58	35	68
Hanyang	28	77	52	38
Dongxihu	3	0	7	27
Jiangnan	36	22	11	15
Wuchang	64	91	38	6
Huangpi	5	3	0	5
Zhuankou	0	6	1	4
Jiangxia	0	0	2	4
Caidian	0	2	0	3
Xinzhou	0	0	0	1
Qingshan	39	39	15	1
Hannan	0	0	0	1
All districts	339	381	232	413

Table 4 shows the patients' underlying conditions, which were extracted from the health condition description in each entry. Document frequency analysis shows that "hypertension,"

"diabetes," "heart disease," and "underlying disease" were mentioned in 12% (n=110), 9% (n=82), 9% (n=79), and 3% (n=30) of the 883 posts, respectively. According to the report

published by the WHO-China Joint Mission, individuals with highest risk for severe disease and death [2]. underlying conditions such as hypertension and diabetes are at

**Table 4.** Document frequency of terms indicating patients' underlying condition.<sup>a</sup>

Word	Frequency (n=883), n (%)	Quote
Hypertension	110 (12)	"The patient has <i>hypertension</i> , diabetes and other underlying conditions. He has had diarrhea for five days with continuous wheezing and breathing difficulties..."
Diabetes	82 (9)	"My father has multiple underlying diseases, including <i>diabetes</i> and hypertension. The CT <sup>b</sup> scan shows ground glass opacity. We have reported to the community hospital and they said that they could do nothing..."
Heart disease	79 (9)	"My grandpa has a history of <i>heart disease</i> for years and received emergency treatment for several times before."
Underlying disease	30 (3)	"The CT scan shows ground glass opacity in both lungs. My father has serious <i>underlying disease</i> of cardiomegaly. He needs to be hospitalized immediately..."

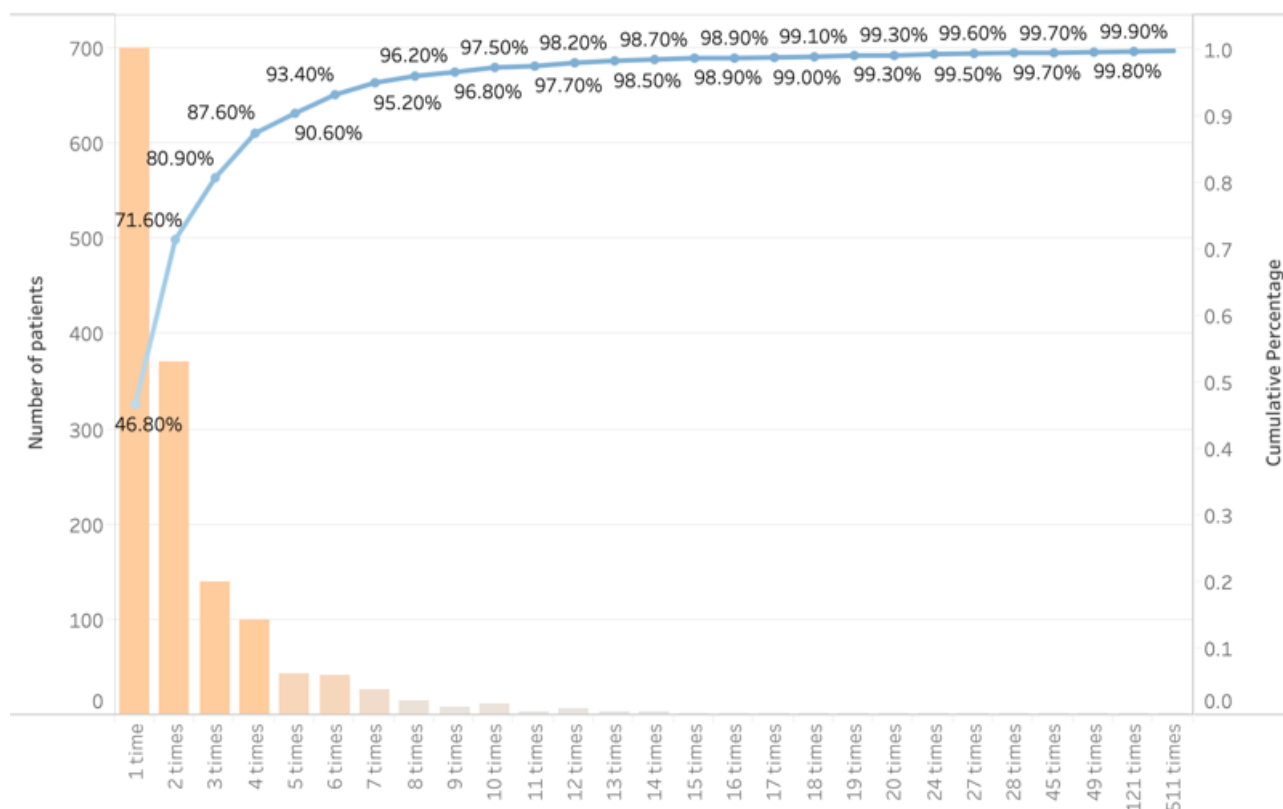
<sup>a</sup>Health condition was missing in 613 Weibo entries.

<sup>b</sup>CT: computed tomography.

Our second research question asks what kinds of health information patients in Wuhan were seeking on Weibo at the peak of the outbreak. We first examined the frequency of posting per patient. Seen from Figure 4, most of the patients have posted

at least 3 times, which accounted for 80.9% (n=1211) of the sample. Still, another 14.2% (n=213) of the sample posted 4-7 times. Although some patients did post more than 7 times, it was rare, which accounted for only 4.8% (n=72) of the sample.

**Figure 4.** Posting times per patient.



We also examined the specific information that these patients or their families were seeking. Textual analysis showed that information about accessing medical treatment, especially hospital beds and confirmatory testing, was most sought out by the patients. In those posts, the users shared their experience of "running around for medical treatment," "being denied service and/or a bed," and for "not knowing how to proceed." Turning to the hashtag, posters were desperately seeking information

on health care. Feelings of desperation and loss of desire for survival were prevalent, such as "my mother is very desperate, telling me no need for treatment any more. She has lost confidence to survive" and "my mom's desire for survival is waning."

The posters also sought information about managing self-quarantine. Due to the shortage of offline health care, the neighborhood committees and hospitals usually suggested the



patients to stay at home to conduct self-quarantine but did not help patients make strategic decisions about self-treatment. The posting highlighted that the patient's "illness deteriorated" due to "a lack of proper treatment and nutritious food." Many users recounted experience of "infection among family members" living under the same roof.

In addition to information and advice, the patients were also requesting for possible ways to provide tangible resources such as "childcare" and "transportation to seek offline health care." Particularly, many requested for social support to "check in on elderly parents" living in Wuhan, to "take them to the hospital," or to "help them with their medication," as they were isolated in a different province, different city, or different district. Although the lockdown policy cut off their offline social support network, the online platform provided a possibility to seek help.

Overall, the social media platform played a vital role connecting health information seekers with a reliable and timely source of information, tangible support, as well as a more empathetic crowd. This source provided not only informal advices but also important, timely discussion and caring interactions. Posts such as "I could do nothing other than following the updates on Weibo everyday," "thank all the kind-hearted people sending me clues and suggestions. Thank you for your empathy," and "my grandpa has been hospitalized. Hope every patient seeking help could be as lucky as me. Thank you all for your kind help" highlight the intangible benefits of using the online platform during the pandemic to seek health information.

## Discussion

### Principal Findings

In this study, we examined how patients with COVID-19 living in Wuhan, China used the "#COVID-19 Patient Seeking Help" hashtag to seek health information. Our findings provide important insights into health information-seeking behaviors during pandemic outbreaks. The curve of the hashtag posting provided a dynamic picture of public attention to the COVID-19 pandemic. Previous studies suggest that an analysis of web behaviors can provide insights into individuals' information seeking during an outbreak, as public reactions are visible more quickly online [12,25]. In our study, we identified a rapid increase in posting under the #COVID-19 Patient Seeking Help hashtag at the onset of the pandemic outbreak and a decrease following the government's effort to admit every patient. The steep curve of the hashtag indicates that online information-seeking behaviors such as posting, commenting, and reposting are useful markers of public reaction and draws attention to the need for public health practitioners to pay attention to online space in their responses. This finding is consistent with research identifying an increase in activities on social networking websites following the outbreaks of pandemic diseases [26,27].

In general, our findings confirmed that the internet is used by the Chinese public as an important source of health information. Previous studies have associated increased age with decreased levels of motivation for health-related information seeking online [36,37]. Similarly, our findings highlighted that younger

family members primarily sought information online for parents or for older adult patients at home [61]. Furthermore, our findings highlight the ability to seek information and help online, especially for those with young children or older adult members during the pandemic. This finding deserves consideration in the context of China, considering the age-related digital divide and the decline in health information-searching behaviors among older adults [50]. Although timely access to quality health care information during outbreaks is vital for reducing morbidity and mortality [13,14], it is equally important to pay attention to group-specific health information needs and their ability to act upon the information.

In our sample, we also identified a high proportion of female information seekers, which was consistent with previous studies that found females more likely to seek online health information [16,35]. However, posting frequency was comparable between female and male seekers in our sample. Another factor that may have contributed to the patients' use of the internet for health information is the long travel time to access offline health care resources. Previous studies indicated that the cost associated with time to visit health care providers in traditional settings has influence on patients' motivation for seeking out health information on the internet [17]. In our sample, around 30% of the patients with COVID-19 lived in a distance more than 3 km from their nearest clinic or designated hospital. The suspension of transportation in Wuhan meant that patients had to walk for at least an hour one way to access an offline health care source. The difficulties in travelling to clinics or hospitals were also cited by many as a major reason to seek health information and help from online platforms. The lockdown policy was also found to cut off the patients' social support network, preventing them from seeking help from family members. Social support was sought on social media to check in on older adults, to take them to the hospital, or to help them with medication, which highlights the vulnerability of this population despite the effectiveness of the policy in containing the disease.

Our findings give insight into the issues that patients and their families were most concerned about during the peak of the outbreak, including where and how to seek medical treatment and confirmatory testing, decision making on self-quarantine, and experience of infection among family members. Previous work indicated that an information need can arise when a patient experiences health-related uncertainty and, in turn, engages in health information-seeking behavior to get reassurance, to manage uncertainty, and to reconcile oneself with a new health situation [16,21]. Our findings highlight a need for information originating from the stretched condition in the health care system and the anxiety over the lack of access to proper treatment. To the patients and their families, the act of searching for information online is a help-seeking step so that they can manage their own health with the affordance of the internet. Scholars suggest this kind of behavior should be encouraged as an integral and positive part of the patients' journey because online health information seeking enables patients to accumulate more social support, which is associated with better health outcomes and health decision making [17,48].

Methodologically, our study also indicates the usefulness of using a computational method to explore individuals' responses

to public health crises in real time. For example, the increase and then decrease in public anxiety eased by the communication effort in response to the H1N1 epidemic was hard to capture by traditional survey methods [12]. Consistent with previous studies, our study shows that the number of entries with the #COVID-19 Patient Seeking Help hashtag rapidly grew and was kept at a high level within a period of more than 1 week and then steadily declined following the government's effort to admit every patient [60]. By monitoring and analyzing the patients' online data, our method enables a possible advantage over traditional approaches to offer a dynamic picture of changes in public response to the pandemic in real time.

Our study also helps build an understanding of how the internet can be used to better serve the needs of the public, especially the patients in the time of an outbreak. In general, our findings confirmed that the internet is used by the Chinese public as an important source of information and help. Although previous studies mainly focused on the online health information-seeking experience of the young and educated [43,44,47,51], our findings highlight the needs of older adults, who may have equal motivations but lack the ability for searching and comprehending online health-related information. Therefore, in addition to making relevant and high-quality information available online, it is vital to motivate social support to facilitate their information needs.

## Limitations and Conclusion

Using a nonprobability and convenience sample, this study focused on basic descriptive analyses of how people in China are using the internet for seeking health information during a pandemic. Nonprobability sampling means there lacks a sound theoretical basis for statistical inference [62]. Future studies using random sampling are needed to allow valid statistical analysis so that informed judgments can be made. Another limitation of our study is the lack of our ability to establish direct links with the patients' health outcomes. More in-depth discussion is needed to explore whether and how the information-seeking behaviors on social media aids in better health outcomes. Future studies should explore the link between information need and patients' health outcomes.

The COVID-19 pandemic has been found to stretch the local medical system and poses a challenge to the state's risk communication efforts. Social media is used by the patients to seek health information relevant to the outbreak. Some factors may contribute to their online information-seeking motivation including age, gender, underlying conditions, and travel time to offline health care service providers. Overall, the findings contribute to our understanding of health information-seeking behaviors during an outbreak and highlight the importance of paying attention to the information need of vulnerable groups and the role social media may play.

## Acknowledgments

The authors disclosed receipt of the following financial support for the research, authorship, or publication of this article: This work was supported by the MOE Key Project of Key Research Institute of Humanities and Social Sciences (Project No. 16JJD860008).

## Conflicts of Interest

None declared.

## References

1. United States COVID-19 cases and deaths by state. Centers for Disease Control and Prevention. 2020. URL: <https://www.cdc.gov/coronavirus/2019-ncov/summary.html> [accessed 2020-03-12]
2. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). World Health Organization. 2020 Feb. URL: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf> [accessed 2020-03-10]
3. Updates on coronavirus disease until 24, March 12. National Health Commission of the People's Republic of China. 2020 Mar 13. URL: <http://www.nhc.gov.cn/xcs/yqtb/202003/816e6f71236b4dca96378df5f6f4ae53.shtml> [accessed 2020-08-26]
4. Coronavirus disease situation summary of Hubei province on March 12, 2020. Health Commission of Hubei Province. 2020. URL: [http://wjw.hubei.gov.cn/bmdt/ztlz/fkxgzbdgrfyyq/xxfb/202003/t20200313\\_2180674.shtml](http://wjw.hubei.gov.cn/bmdt/ztlz/fkxgzbdgrfyyq/xxfb/202003/t20200313_2180674.shtml) [accessed 2020-08-26]
5. Tian H, Liu Y, Li Y, Wu C, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 2020 May 08;368(6491):638-642 [FREE Full text] [doi: [10.1126/science.abb6105](https://doi.org/10.1126/science.abb6105)] [Medline: [32234804](https://pubmed.ncbi.nlm.nih.gov/32234804/)]
6. Pan SL, Cui M, Qian J. Information resource orchestration during the COVID-19 pandemic: a study of community lockdowns in China. *Int J Inf Manage* 2020 Oct;54:102143 [FREE Full text] [doi: [10.1016/j.ijinfomgt.2020.102143](https://doi.org/10.1016/j.ijinfomgt.2020.102143)] [Medline: [32394997](https://pubmed.ncbi.nlm.nih.gov/32394997/)]
7. Lin X. Analysis of 25,000 lab-confirmed COVID-19 cases in Wuhan: epidemiological characteristics and nonpharmaceutical intervention effects. Harvard University. 2020. URL: <https://cdn1.sph.harvard.edu/wp-content/uploads/sites/21/2020/03/COVID-19-03-16-2020-Lin.pdf> [accessed 2020-08-26]
8. Chen S, Zhang Z, Yang J, Wang J, Zhai X, Bärnighausen T, et al. Fangcang shelter hospitals: a novel concept for responding to public health emergencies. *Lancet* 2020 Apr;395(10232):1305-1314. [doi: [10.1016/s0140-6736\(20\)30744-3](https://doi.org/10.1016/s0140-6736(20)30744-3)]

9. Liu L. Sustainable COVID-19 mitigation: Wuhan lockdowns, health inequities, and patient evacuation. *Int J Health Policy Manag* 2020 Apr 28;1. [doi: [10.34172/ijhpm.2020.63](https://doi.org/10.34172/ijhpm.2020.63)] [Medline: [32610729](https://pubmed.ncbi.nlm.nih.gov/32610729/)]
10. Khan S, Nabi G, Han G, Siddique R, Lian S, Shi H, et al. Novel coronavirus: how things are in Wuhan. *Clin Microbiol Infect* 2020 Apr;26(4):399-400 [FREE Full text] [doi: [10.1016/j.cmi.2020.02.005](https://doi.org/10.1016/j.cmi.2020.02.005)] [Medline: [32058086](https://pubmed.ncbi.nlm.nih.gov/32058086/)]
11. Holmes BJ. Communicating about emerging infectious disease: the importance of research. *Health Risk Soc* 2008 Aug;10(4):349-360. [doi: [10.1080/13698570802166431](https://doi.org/10.1080/13698570802166431)]
12. Tausczik Y, Faasse K, Pennebaker JW, Petrie KJ. Public anxiety and information seeking following the H1N1 outbreak: blogs, newspaper articles, and Wikipedia visits. *Health Commun* 2012;27(2):179-185. [doi: [10.1080/10410236.2011.571759](https://doi.org/10.1080/10410236.2011.571759)] [Medline: [21827326](https://pubmed.ncbi.nlm.nih.gov/21827326/)]
13. Barry JM. Pandemics: avoiding the mistakes of 1918. *Nature* 2009 May 21;459(7245):324-325 [FREE Full text] [doi: [10.1038/459324a](https://doi.org/10.1038/459324a)] [Medline: [19458695](https://pubmed.ncbi.nlm.nih.gov/19458695/)]
14. Kumar S, Quinn SC, Kim KH, Musa D, Hilyard KM, Freimuth VS. The social ecological model as a framework for determinants of 2009 H1N1 influenza vaccine uptake in the United States. *Health Educ Behav* 2012 Apr;39(2):229-243 [FREE Full text] [doi: [10.1177/1090198111415105](https://doi.org/10.1177/1090198111415105)] [Medline: [21984692](https://pubmed.ncbi.nlm.nih.gov/21984692/)]
15. Kammerer Y, Amann DG, Gerjets P. When adults without university education search the Internet for health information: the roles of internet-specific epistemic beliefs and a source evaluation intervention. *Comput Hum Behav* 2015 Jul;48:297-309. [doi: [10.1016/j.chb.2015.01.045](https://doi.org/10.1016/j.chb.2015.01.045)]
16. Lagoe C, Atkin D. Health anxiety in the digital age: An exploration of psychological determinants of online health information seeking. *Comput Hum Behav* 2015 Nov;52:484-491. [doi: [10.1016/j.chb.2015.06.003](https://doi.org/10.1016/j.chb.2015.06.003)]
17. Peddie KA, Kelly-Campbell RJ. How people with hearing impairment in New Zealand use the Internet to obtain information about their hearing health. *Comput Hum Behav* 2017 Aug;73:141-151. [doi: [10.1016/j.chb.2017.03.037](https://doi.org/10.1016/j.chb.2017.03.037)]
18. Niederdeppe J, Hornik RC, Kelly BJ, Frosch DL, Romantan A, Stevens RS, et al. Examining the dimensions of cancer-related information seeking and scanning behavior. *Health Commun* 2007;22(2):153-167. [doi: [10.1080/10410230701454189](https://doi.org/10.1080/10410230701454189)] [Medline: [17668995](https://pubmed.ncbi.nlm.nih.gov/17668995/)]
19. Wong LP, Sam I. Public sources of information and information needs for pandemic influenza A(H1N1). *J Community Health* 2010 Apr 22;35(6):676-682. [doi: [10.1007/s10900-010-9271-4](https://doi.org/10.1007/s10900-010-9271-4)]
20. DeLorme DE, Huh J. Seniors' uncertainty management of direct-to-consumer prescription drug advertising usefulness. *Health Commun* 2009 Sep;24(6):494-503. [doi: [10.1080/10410230903104277](https://doi.org/10.1080/10410230903104277)] [Medline: [19735027](https://pubmed.ncbi.nlm.nih.gov/19735027/)]
21. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. *Comput Hum Behav* 2017 Apr;69:256-267. [doi: [10.1016/j.chb.2016.12.032](https://doi.org/10.1016/j.chb.2016.12.032)]
22. Miller LMS, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. *J Aging Health* 2012 Apr;24(3):525-541. [doi: [10.1177/0898264311428167](https://doi.org/10.1177/0898264311428167)] [Medline: [22187092](https://pubmed.ncbi.nlm.nih.gov/22187092/)]
23. Kitchens B, Harle CA, Li S. Quality of health-related online search results. *Decision Support Syst* 2014 Jan;57:454-462. [doi: [10.1016/j.dss.2012.10.050](https://doi.org/10.1016/j.dss.2012.10.050)]
24. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013 Jan 15. URL: <https://www.ordinedeimedici.com/documenti/Docs7-cybercondria-PIP-HealthOnline.pdf> [accessed 2020-03-14]
25. Weeks B, Southwell B. The symbiosis of news coverage and aggregate online search behavior: Obama, rumors, and presidential politics. *Mass Commun Soc* 2010 Aug 23;13(4):341-360. [doi: [10.1080/15205430903470532](https://doi.org/10.1080/15205430903470532)]
26. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Feb 19;457(7232):1012-1014. [doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634)] [Medline: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)]
27. Petrie KJ, Faasse K. Monitoring public anxiety about flu weblog message. *The BMJ Opinion*. 2009 Jun 11. URL: <https://blogs.bmj.com/bmj/2009/06/11/keith-j-petrie-and-kate-faasse-monitoring-public-anxiety-about-flu/> [accessed 2020-03-14]
28. Bantum EO, Owen JE. Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychol Assess* 2009 Mar;21(1):79-88. [doi: [10.1037/a0014643](https://doi.org/10.1037/a0014643)] [Medline: [19290768](https://pubmed.ncbi.nlm.nih.gov/19290768/)]
29. Kahn JH, Tobin RM, Massey AE, Anderson JA. Measuring emotional expression with linguistic inquiry and word count. *Am J Psychol* 2007;120:286. [doi: [10.2307/20445398](https://doi.org/10.2307/20445398)]
30. Andreassen H, Bujnowska-Fedak M, Chronaki C, Dumitru R, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. *BMC Public Health* 2007 Apr 10;7:53 [FREE Full text] [doi: [10.1186/1471-2458-7-53](https://doi.org/10.1186/1471-2458-7-53)] [Medline: [17425798](https://pubmed.ncbi.nlm.nih.gov/17425798/)]
31. Rains SA, Tukachinsky R. An examination of the relationships among uncertainty, appraisal, and information-seeking behavior proposed in uncertainty management theory. *Health Commun* 2015;30(4):339-349. [doi: [10.1080/10410236.2013.858285](https://doi.org/10.1080/10410236.2013.858285)] [Medline: [24905910](https://pubmed.ncbi.nlm.nih.gov/24905910/)]
32. Rains SA. Health information seeking and the World Wide Web: an uncertainty management perspective. *J Health Commun* 2014;19(11):1296-1307. [doi: [10.1080/10810730.2013.872731](https://doi.org/10.1080/10810730.2013.872731)] [Medline: [24730566](https://pubmed.ncbi.nlm.nih.gov/24730566/)]
33. Anker AE, Reinhart AM, Feeley TH. Health information seeking: a review of measures and methods. *Patient Educ Couns* 2011 Mar;82(3):346-354. [doi: [10.1016/j.pec.2010.12.008](https://doi.org/10.1016/j.pec.2010.12.008)] [Medline: [21239134](https://pubmed.ncbi.nlm.nih.gov/21239134/)]
34. Lin CA, Atkin DJ, Cappotto C, Davis C, Dean J, Eisenbaum J, et al. Ethnicity, digital divides and uses of the Internet for health information. *Comput Hum Behav* 2015 Oct;51:216-223. [doi: [10.1016/j.chb.2015.04.054](https://doi.org/10.1016/j.chb.2015.04.054)]



35. Booth A, Cardona-Sosa L, Nolen P. Gender differences in risk aversion: do single-sex environments affect their development? *J Econ Behav Organ* 2014 Mar;99:126-154. [doi: [10.1016/j.jebo.2013.12.017](https://doi.org/10.1016/j.jebo.2013.12.017)]
36. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of "Dr Google": a survey on health information-seeking behaviors and navigational needs. *J Med Internet Res* 2015 Dec 29;17(12):e288 [FREE Full text] [doi: [10.2196/jmir.4345](https://doi.org/10.2196/jmir.4345)] [Medline: [26715363](https://pubmed.ncbi.nlm.nih.gov/26715363/)]
37. van Deursen A, van Dijk J. Internet skills and the digital divide. *N Media Soc* 2010 Dec 06;13(6):893-911. [doi: [10.1177/1461444810386774](https://doi.org/10.1177/1461444810386774)]
38. Baumgartner SE, Hartmann T. The role of health anxiety in online health information search. *Cyberpsychol Behav Soc Netw* 2011 Oct;14(10):613-618. [doi: [10.1089/cyber.2010.0425](https://doi.org/10.1089/cyber.2010.0425)] [Medline: [21548797](https://pubmed.ncbi.nlm.nih.gov/21548797/)]
39. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? *Health Serv Res* 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: [10.1111/j.1475-6773.2006.00510.x](https://doi.org/10.1111/j.1475-6773.2006.00510.x)] [Medline: [16704514](https://pubmed.ncbi.nlm.nih.gov/16704514/)]
40. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. *J Med Internet Res* 2015 Jun 01;17(6):e134 [FREE Full text] [doi: [10.2196/jmir.4365](https://doi.org/10.2196/jmir.4365)] [Medline: [26032933](https://pubmed.ncbi.nlm.nih.gov/26032933/)]
41. Tang C, Wu X, Chen X, Pan B, Yang X. Examining income-related inequality in health literacy and health-information seeking among urban population in China. *BMC Public Health* 2019 Feb 21;19(1):221 [FREE Full text] [doi: [10.1186/s12889-019-6538-2](https://doi.org/10.1186/s12889-019-6538-2)] [Medline: [30791882](https://pubmed.ncbi.nlm.nih.gov/30791882/)]
42. Yang Y, Zhang S, Xia L. Preventive health information seeking behavior among general population in China. 2012 Presented at: International Symposium on Information Technologies in Medicine and Education; 2012; Hokodate, Hokkaido, Japan p. 608-612. [doi: [10.1109/itime.2012.6291380](https://doi.org/10.1109/itime.2012.6291380)]
43. Cao W, Zhang X, Xu K, Wang Y. Modeling online health information-seeking behavior in China: the roles of source characteristics, reward assessment, and internet self-efficacy. *Health Commun* 2016 Sep;31(9):1105-1114. [doi: [10.1080/10410236.2015.1045236](https://doi.org/10.1080/10410236.2015.1045236)] [Medline: [26861963](https://pubmed.ncbi.nlm.nih.gov/26861963/)]
44. Wang MP, Wang X, Viswanath K, Wan A, Lam TH, Chan SS. Digital inequalities of family life information seeking and family well-being among Chinese adults in Hong Kong: a population survey. *J Med Internet Res* 2014 Oct 03;16(10):e227 [FREE Full text] [doi: [10.2196/jmir.3386](https://doi.org/10.2196/jmir.3386)] [Medline: [25281831](https://pubmed.ncbi.nlm.nih.gov/25281831/)]
45. The 44th statistical report on Internet development in China. China Internet Network Information Center (CNNIC). 2019 Aug 30. URL: <http://www.cnnic.net.cn/hlwfzyj/hlwxyzbg/hlwztjbg/201908/P020190830356787490958.pdf> [accessed 2020-03-14]
46. Statistical communiqué of the People's Republic of China on the 2019 national economic and social development. National Bureau of Statistics of China. 2020 Feb 28. URL: [http://www.stats.gov.cn/english/PressRelease/202002/t20200228\\_1728917.html](http://www.stats.gov.cn/english/PressRelease/202002/t20200228_1728917.html) [accessed 2020-03-14]
47. Wang MP, Viswanath K, Lam TH, Wang X, Chan SS. Social determinants of health information seeking among Chinese adults in Hong Kong. *PLoS One* 2013;8(8):e73049 [FREE Full text] [doi: [10.1371/journal.pone.0073049](https://doi.org/10.1371/journal.pone.0073049)] [Medline: [24009729](https://pubmed.ncbi.nlm.nih.gov/24009729/)]
48. Jiang S, Street RL. Pathway linking internet health information seeking to better health: a moderated mediation study. *Health Commun* 2017 Aug;32(8):1024-1031. [doi: [10.1080/10410236.2016.1196514](https://doi.org/10.1080/10410236.2016.1196514)] [Medline: [27464036](https://pubmed.ncbi.nlm.nih.gov/27464036/)]
49. Gao L, Larsson M, Luo S. Internet use by Chinese women seeking pregnancy-related information. *Midwifery* 2013 Jul;29(7):730-735. [doi: [10.1016/j.midw.2012.07.003](https://doi.org/10.1016/j.midw.2012.07.003)] [Medline: [22958935](https://pubmed.ncbi.nlm.nih.gov/22958935/)]
50. Wu D, Li Y. Online health information seeking behaviors among Chinese elderly. *Libr Inf Sci Res* 2016 Jul;38(3):272-279. [doi: [10.1016/j.lisr.2016.08.011](https://doi.org/10.1016/j.lisr.2016.08.011)]
51. Yan YY. Online health information seeking behavior in Hong Kong: an exploratory study. *J Med Syst* 2010 Apr;34(2):147-153. [doi: [10.1007/s10916-008-9226-9](https://doi.org/10.1007/s10916-008-9226-9)] [Medline: [20433053](https://pubmed.ncbi.nlm.nih.gov/20433053/)]
52. Lianjia. URL: <https://bj.lianjia.com/>
53. Baidu Map Open Platform. URL: <http://lbsyun.baidu.com/>
54. Fox S. Online health search 2006. Pew Research Center. 2006. URL: [https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2006/PIP\\_Online\\_Health\\_2006.pdf.pdf](https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf) [accessed 2020-03-14]
55. fxsjy / jieba. GitHub. URL: <https://github.com/fxsjy/jieba>
56. The list of fever clinics and designated hospitals. Health Commission of Hubei Province. 2020 Jan 20. URL: [http://wjw.hubei.gov.cn/bmdt/ztlz/fkxxgzbdgrfyyq/fkdt/202001/t20200122\\_2014148.shtml](http://wjw.hubei.gov.cn/bmdt/ztlz/fkxxgzbdgrfyyq/fkdt/202001/t20200122_2014148.shtml) [accessed 2020-08-26]
57. Summary of hospital beds availability in 28 designated hospitals in Wuhan. Hubei Provincial People's Government. 2020 Feb 05. URL: [http://www.hubei.gov.cn/zhuanti/2020/gzxxgzbd/zxtb/202002/t20200205\\_2019615.shtml](http://www.hubei.gov.cn/zhuanti/2020/gzxxgzbd/zxtb/202002/t20200205_2019615.shtml) [accessed 2020-08-26]
58. Charmaz K. Constructing grounded theory (introducing qualitative methods series). London: SAGE Publications, Inc; 2014:1-416.
59. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: SAGE Publications, Inc; 2014:1-456.
60. Wuhan launches the final attack: admit every patient with no time to delay. People's Daily. 2020 Feb 11. URL: <http://society.people.com.cn/n1/2020/0211/c1008-31580458.html> [accessed 2020-03-14]
61. Zhao X, Mao Q, Kreps GL, Yu G, Li Y, Chou SW, et al. Cancer information seekers in china: a preliminary profile. *J Health Commun* 2015;20(5):616-626. [doi: [10.1080/10810730.2015.1012244](https://doi.org/10.1080/10810730.2015.1012244)] [Medline: [25845031](https://pubmed.ncbi.nlm.nih.gov/25845031/)]

62. Pierce M, McManus S, Jessop C, John A, Hotopf M, Ford T, et al. Says who? The significance of sampling in mental health surveys during COVID-19. *Lancet Psychiatry* 2020 Jul;7(7):567-568. [doi: [10.1016/s2215-0366\(20\)30237-6](https://doi.org/10.1016/s2215-0366(20)30237-6)]

## Abbreviations

**API:** application programming interface

**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 27.07.20; peer-reviewed by T Liyan, M Barbosa; comments to author 18.08.20; revised version received 01.09.20; accepted 15.09.20; published 15.10.20.*

*Please cite as:*

Zhao X, Fan J, Basnyat I, Hu B

Online Health Information Seeking Using “#COVID-19 Patient Seeking Help” on Weibo in Wuhan, China: Descriptive Study  
*J Med Internet Res* 2020;22(10):e22910

URL: <http://www.jmir.org/2020/10/e22910/>

doi: [10.2196/22910](https://doi.org/10.2196/22910)

PMID: [33001838](https://pubmed.ncbi.nlm.nih.gov/33001838/)

©Xiaoman Zhao, Ju Fan, Iccha Basnyat, Baijing Hu. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 15.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



## Original Paper

# Knowledge, Awareness, and Attitudes Relating to the COVID-19 Pandemic Among Different Populations in Central China: Cross-Sectional Survey

Huifang Xu<sup>1\*</sup>, MD, PhD; Maria Jose Gonzalez Mendez<sup>2\*</sup>, MSc; Lanwei Guo<sup>1</sup>, MD, PhD; Qiong Chen<sup>1</sup>, MD, PhD; Liyang Zheng<sup>1</sup>, BSc; Peipei Chen<sup>1</sup>, BSc; Xiaoqin Cao<sup>1</sup>, MD, PhD; Shuzheng Liu<sup>1</sup>, MD, PhD; Xibin Sun<sup>1</sup>, BSc; Shaokai Zhang<sup>1</sup>, MD, PhD; Youlin Qiao<sup>1,3</sup>, MD, PhD

<sup>1</sup>Department of Cancer Epidemiology, Affiliated Cancer Hospital of Zhengzhou University/Henan Cancer Hospital, Zhengzhou, China

<sup>2</sup>School of Public Health, Dalian Medical University, Dalian, China

<sup>3</sup>Department of Epidemiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

\*these authors contributed equally

**Corresponding Author:**

Shaokai Zhang, MD, PhD

Department of Cancer Epidemiology

Affiliated Cancer Hospital of Zhengzhou University/Henan Cancer Hospital

127 Dongming Rd

Jinshui District

Zhengzhou

China

Phone: 86 037165587346

Email: [shaokaizhang@126.com](mailto:shaokaizhang@126.com)

## Abstract

**Background:** The COVID-19 pandemic has threatened the health systems of many countries worldwide. Several studies have suggested that the pandemic affects not only physical health but also all aspects of society. A lot of information has been reported about the disease since the beginning of the outbreak. For that reason, it is essential to investigate the attitudes and level of knowledge and awareness that different populations had regarding COVID-19 during the critical period of the outbreak.

**Objective:** This study aimed to assess the knowledge and awareness of and attitudes toward the COVID-19 pandemic among different populations in Central China during the critical period of the outbreak.

**Methods:** A cross-sectional web-based survey was conducted in Central China from February to March 2020. The study participants included three different populations: medical workers, students, and those with other occupations. In this study, a questionnaire was designed to collect information on the following four aspects: sociodemographic information, knowledge related to COVID-19, awareness of COVID-19, and attitude toward COVID-19. The chi-square test and Fisher test were used for comparison among groups. The level of significance was set at  $P<.05$ .

**Results:** This study enrolled a total of 508 participants. Among them, there were 380 students (74.8%), 39 medical workers (7.7%), and 89 people with other occupations (17.5%). Most of the participants were female ( $n=272$ , 53.5%), lived in rural areas ( $n=258$ , 50.8%), and were single ( $n=423$ , 86.9%). The majority of the respondents had attended college ( $n=454$ , 89.4%). Most of the participants said they had heard about COVID-19 by January, and most of them looked for information on social media (Sina Weibo, 84.7%), and WeChat and QQ groups (74.2%). The participants showed an adequate level of knowledge about COVID-19 with no significant differences among the groups. However, medical workers demonstrated a slightly advanced knowledge in their responses to professional questions such as the potential susceptible population, possible host, treatment of COVID-19, and disease category. A higher proportion of medical workers (71.8%) and those in the other occupations group (52.8%) were highly concerned about the COVID-19 pandemic. More than 43% of the participants stated that the lockdown of their village/city had a significant impact on their lives. Nevertheless, the majority of respondents had an overall optimistic attitude toward the control of the disease (92.1% of students [ $n=350$ ], 94.9% of medical workers [ $n=37$ ], and 92.3% of those in other occupations [ $n=83$ ]).

**Conclusions:** All three groups reported an adequate background knowledge about COVID-19 but medical workers showed a slightly advanced knowledge in their responses to professional questions. Most of the participants were highly concerned about COVID-19 during the critical period of the outbreak. The majority of respondents declared that the village/city lockdown policy had a significant impact on their daily life but most of them held an optimistic attitude toward the control of COVID-19.

(*J Med Internet Res* 2020;22(10):e22628) doi:[10.2196/22628](https://doi.org/10.2196/22628)

## KEYWORDS

COVID-19; knowledge; awareness; attitude

## Introduction

Coronavirus disease (COVID-19) rapidly spread around the globe and it has led to an economic crisis alongside a health care crisis in many countries and regions across the world [1]. It was declared a pandemic by the World Health Organization (WHO) on March 11, 2020, affecting 114 countries by that time [2]. Even today, COVID-19 is still a threat to the health systems of nearly 150 countries and regions, especially those dealing with an international health emergency for the first time. In China, authorities and citizens acted quickly and effectively to contain the spread of COVID-19. The measures taken included several lockdown policies and strict preventive measures across the country [3], offering a model of COVID-19 pandemic control to other countries and regions.

The increasing use of computers, tablets, and smartphones enables the rapid dissemination of information through the internet and social media, but such information lacks an effective guarantee of quality. As with previous epidemics such as Ebola or Zika virus disease, the internet has been used to spread misinformation regarding the COVID-19 pandemic [4]. It has been reported that social distancing promotes the spread of misinformation and has a profound impact on psychological well-being in vulnerable populations. Misinformation may lead to uncertainty, which is increased due to the novelty of the disease [5]. Therefore, it is essential to raise awareness and distribute accurate information about COVID-19 to avoid misinformation and decrease unnecessary stress in the population. Studies have described the effects that the pandemic can have on specific groups, such as medical workers [6], children and teenagers [7], older adults [8], and students [9]. In China, a study conducted by Yulan Lin and colleagues [10] determined the knowledge, attitudes, and anxiety levels of the general population in relation to the COVID-19 outbreak, as well as its impact on them. However, the knowledge- and attitude-related items in their study were limited to the symptoms and modes of transmission of COVID-19 and did not explore the differences between different population groups. To avoid the above limitations, this study was designed and conducted to evaluate the knowledge and awareness of as well as attitude toward COVID-19 of different populations groups in Central China during the critical period of the outbreak. This information might be necessary for policy makers to promote health education campaigns to better control the disease before the availability of a preventive vaccine. As there is no available vaccine or effective antiviral treatment against COVID-19, the active engagement of the population in preventive behaviors is necessary. For that reason, an adequate knowledge and high

level of awareness of COVID-19, as well as an optimistic attitude, are crucial for controlling the disease.

## Methods

### Study Design and Participants

This study presents data from the Central China part of a population-based multicenter cross-sectional online survey, conducted in seven geographical regions across China (Northeast China, East China, South China, North China, Northwest China, Southwest China, and Central China) in February and March 2020. The anonymous survey was conducted via Wenjuanxing, an online crowdsourcing platform in mainland China that provides professional online surveys, voting, testing, and comments. The survey was developed to investigate the knowledge and awareness of and attitude toward the COVID-19 pandemic in different populations (medical workers, students, and those in other occupations). The survey contained items related to participants' demographic background, knowledge, awareness, and attitudes.

The inclusion criterion was that the respondents were at least 18 years old when they filled out the survey. Thus, anyone that had the ability to complete this survey and met the age requirement was qualified to take part in the study. Respondents in this study were divided into three different population groups: students, medical workers, and those in other occupations.

Before starting the online questionnaire, a brief introduction was displayed to each participant, and electronic informed consent was obtained if they agreed to complete the questionnaire.

To confirm the quality of the online survey, two control items were included in the questionnaire. The first one was gender; although gender information was recorded initially, it appeared one more time in the questionnaire with confusing options. Another control item was a question about whether the influenza vaccine can prevent COVID-19. Two different descriptions were displayed, and the answers for the control items were supposed to match. Moreover, the members of the research team had been trained to ensure the quality of the questionnaire. Data cleaning and checking were done once the questionnaire was submitted.

### Sociodemographic Information

In the questionnaire, the following sociodemographic information was collected from all participants: gender, ethnicity, date of birth, marital status, household type, education background, monthly income, occupation, and smoking and drinking status.

In this study, respondents were identified as a current smoker or drinker if they reported being an active smoker or drinker.

### Knowledge and Awareness

The participants' knowledge was assessed using a series of questions related to COVID-19. The questions included the following items: "Did you hear about COVID-19?" "When did you hear about COVID-19?" "Do you agree that COVID-19 is the same as influenza virus?" "Do you agree that COVID-19 is the same as SARS?" "Do you agree that influenza vaccine can prevent COVID-19?" "Which level is the category and treatment of COVID-19 classified in China?" "What is the possible host for COVID-19?" "Which population group is susceptible to COVID-19?" "How long is the incubation period for COVID-19?" "Do you think that COVID-19 is contagious during the incubation period?" "Is there any effective treatment for COVID-19?" "When do you think that it is necessary to wear a mask during the COVID-19 pandemic?" and "Do you know that COVID-19 has been declared to be a Public Health Emergency of International Concern (PHEIC)?"

### Attitude Toward the COVID-19 Pandemic

In this section, the following items were designed to determine the participants' attitude toward COVID-19: "What is your level of concern about COVID-19?" "How often do you check updates about COVID-19?" "What impact does the village/city lockdown policy have on your daily life?" "What is your attitude toward COVID-19 pandemic control?" and "Do you think that COVID-19 could be a global outbreak?"

### Ethical Considerations

This study protocol was approved by the Ethics Committee of Jining Medical College (JNMC-2020-KY-001).

### Statistical Analysis

We calculated counts and proportions for countable data in the questionnaire. Bar plots were used if necessary. The chi-square test was used to compare the differences in the countable data of groups. The Fisher test was applied to data that was not qualified for the chi-square test due to the small size of the sample. The differences between groups were considered statistically significant if the *P* value was  $<.05$ . All statistics were completed with SAS (Version 9.5; SAS Institute Inc).

## Results

### Sociodemographic Characteristics

Table 1 summarizes the sociodemographic characteristics of respondents by group. In this study, a total of 822 questionnaires were collected, 508 were used in analysis, and 314 were excluded due to the age limitation or inconsistencies in control item responses.

Of the 508 respondents, there were 380 students (74.8%), 39 medical workers (7.7%), and 89 people in other occupations (17.5%). The mean age was 24.1 years. The mean age for each group was 21.5 years for students, 29.6 years for medical workers, and 32.8 years for those in other occupations. In total, 272 participants (53.5%) were female. A slightly higher percentage of participants lived in rural areas ( $n=258$ , 50.8%) than urban areas ( $n=250$ , 49.2%), and most of them were single ( $n=423$ , 86.9%). The education level of the majority of the respondents was college and above ( $n=454$ , 89.4%). For monthly income, 58.7% had no income; this answer was especially high among students (76.9%). Most of the participants said they were not currently smoking ( $n=452$ , 89%) or drinking ( $n=344$ , 67.7%).

**Table 1.** Sociodemographic characteristic distributions of participants by occupation group.

Characteristics	Students (n=380)	Medical workers (n=39)	Other occupations (n=89)	Total (N=508)
Age (years), mean (SD)	21.5 (SD 3.48)	29.6 (SD 4.03)	32.8 (SD 9.95)	24.1 (SD 6.95)
<b>Gender, n (%)</b>				
Male	190 (50.0)	12 (30.78)	34 (38.2)	236 (46.5)
Female	190 (50.0)	27 (69.2)	55 (61.8)	272 (53.5)
<b>Ethnicity, n (%)</b>				
Han	367 (96.6)	39 (100)	88 (98.9)	494 (97.2)
Other	13 (3.4)	0 (0.0)	1 (1.1)	14 (2.8)
<b>Household type, n (%)</b>				
Urban	161 (42.4)	30 (76.9)	59 (66.9)	250 (49.2)
Rural	219 (57.6)	9 (23.1)	30 (33.7)	258 (50.8)
<b>Marital status, n (%)</b>				
Single	372 (100)	16 (61.5)	35 (39.3)	423 (86.9)
Married	0 (0.0)	8 (30.8)	51 (57.3)	59 (12.1)
Others	0 (0.0)	2 (7.7)	3 (3.4)	5 (1.0)
<b>Education level, n (%)</b>				
≤High school	2 (0.5)	5 (12.8)	47 (42.8)	54 (10.6)
College and above	378 (99.5)	34 (87.2)	42 (47.2)	454 (89.4)
<b>Monthly income (¥), n (%)</b>				
No income	292 (76.9)	0 (0.0)	6 (6.7)	298 (58.7)
<4000 (<US \$585)	86 (22.6)	8 (20.5)	23 (25.9)	117 (23.0)
≥4000 (≥US \$585)	2 (0.5)	31 (79.5)	60 (67.4)	93 (18.3)
<b>Current smoker, n (%)</b>				
No	341 (89.7)	34 (87.2)	77 (86.5)	452 (89.0)
Yes	39 (10.3)	5 (12.8)	12 (13.5)	56 (11.0)
<b>Current drinker, n (%)</b>				
No	257 (67.6)	27 (69.2)	60 (67.4)	344 (67.7)
Yes	123 (32.4)	12 (30.8)	29 (32.6)	164 (32.3)

## Knowledge and Awareness About COVID-19

Almost all of the respondents said they had heard about COVID-19 (99% of students, 100% of medical workers, and 98.9% of those in other occupations). A higher proportion of medical workers (n=18, 46.2%) and students (n=155, 40.8%) heard about COVID-19 by December 2019. However, a sizeable percentage of the respondents said they found out about COVID-19 between January 1 and 20, especially those in other occupations (52.8%) and students (49.5%).

Regarding the host of COVID-19, all medical workers reported that it was possibly a wild animal (eg, bats), and 98.7% of students (n=375) and 95.6% of those in other occupations (n=85) made the same choice. When the participants were asked about the populations most susceptible to COVID-19, 44.5% of the students (n=169) and 41.6% of those in other occupations (n=37) thought that middle-aged and older adults were more vulnerable. In contrast, 66.7% of medical workers answered that people of all ages were susceptible. The majority of the participants agreed

that the incubation period of COVID-19 was between 1 and 14 days (86.5% to 92.3%), and over 98% of them agreed that the virus was contagious during the incubation period. Concerning treatment, most of the participants knew that there was no effective available treatment against COVID-19 (97.4% of medical workers [n=38] and 93.3% of those in other occupations [n=83]). In addition, 97.4% of medical workers (n=38) knew that COVID-19 was declared a PHEIC by the end of January, followed by students (n=319, 89%) and those in other occupations (n=78, 87.6%).

Regarding the use of masks, most of the participants agreed that it was necessary to wear a mask when going outside, with 90.5% (n=344), 87.2% (n=34), and 85.4% (n=76) of students, medical workers, and those in other occupations groups, respectively. This proportion was significantly higher than that of people that agreed to wear a mask in a crowded place.

As to the comparison between COVID-19 and other virus diseases, most of the students (n=321, 84.5%), those in the other

occupations group (n=77, 86%), and all medical workers (n=39, 100%), knew that COVID-19 and influenza virus were not the same. Correspondingly, a high proportion of students (n=319, 84%), medical workers (n=37, 94.9%), and those in other occupations (n=76, 85.4%) answered that COVID-19 was different from severe acute respiratory syndrome (SARS). When participants were asked if the influenza vaccine can prevent COVID-19, a low proportion of students (n=33, 8.7%), medical workers (n=1, 2.6%), and those in other occupations (n=7, 7.9%) responded “Yes.”

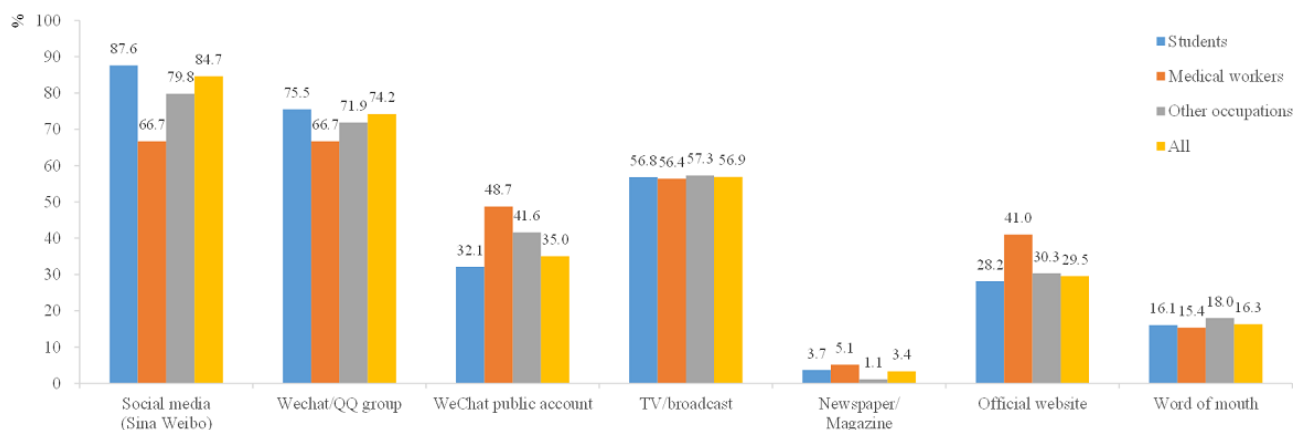
Although the majority of information about COVID-19 was well known by the participants, with no significant differences observed among the groups, medical workers showed a significantly advanced knowledge in their responses to professional questions (Table 2).

Regarding the sources used to search for information about COVID-19, most of the participants obtained information from social media (Sina Weibo, 84.7%), as well as WeChat and QQ groups (74.7%). The least used sources of information for all three groups were newspapers and magazines (3.4%; Figure 1).



**Table 2.** Participants' knowledge and awareness of COVID-19 by occupation group.

Questions and responses	Students, n (%)	Medical workers, n (%)	Other occupations, n (%)	P value
<b>Have you heard about COVID-19?</b>				>.99
Yes	376 (99.0)	39 (100)	88 (98.9)	
No	4 (1.0)	0 (0.0)	1 (1.1)	
<b>When did you hear about COVID-19?</b>				.72
December 2019	155 (40.8)	18 (46.2)	30 (33.7)	
January 1-20, 2020	188 (49.5)	17 (43.6)	47 (52.8)	
January 21-23, 2020	31 (8.2)	4 (10.3)	10 (11.2)	
January 24, 2020, and after	6 (1.6)	0 (0.0)	2 (2.3)	
<b>Do you know that COVID-19 has been declared to be a Public Health Emergency of International Concern?</b>				.06
Yes	319 (89.0)	38 (97.4)	78 (87.6)	
No	61 (16.1)	1 (2.6)	11 (12.4)	
<b>Do you agree that COVID-19 is the same as influenza virus?</b>				.03
Yes	59 (15.5)	0 (0.0)	12 (14.0)	
No	321 (84.5)	39 (100)	77 (86.0)	
<b>Do you agree that COVID-19 is the same as severe acute respiratory syndrome (SARS)?</b>				.19
Yes	61 (16.1)	2 (5.1)	13 (14.6)	
No	319 (84.0)	37 (94.9)	76 (85.4)	
<b>Do you agree that the influenza vaccine can prevent COVID-19?</b>				.41
Yes	33 (8.7)	1 (2.6)	7 (7.9)	
No	345 (91.3)	38 (97.4)	82 (92.1)	
<b>Which level is the category and treatment of COVID-19 classified in China?</b>				<.001
Class B infectious disease and treated as Class A	189 (49.7)	33 (84.6)	34 (38.2)	
Other	191 (50.3)	6 (15.4)	55 (61.8)	
<b>What is the possible host of COVID-19</b>				.09
Wildlife (eg, bat)	375 (98.7)	39 (100.0)	85 (95.6)	
Other	5 (1.3)	0 (0.0)	4 (4.4)	
<b>Which population group is susceptible to COVID-19?</b>				.02
Middle-aged and older adults	169 (44.5)	7 (18.0)	37 (41.6)	
Older adults and children	62 (16.3)	6 (15.4)	14 (15.7)	
All ages	141 (37.1)	26 (66.7)	36 (40.4)	
Young adults	8 (2.1)	0 (0.0)	2 (2.2)	
<b>How long is the incubation period of COVID-19?</b>				.63
1-14 days	332 (87.4)	36 (92.3)	77 (86.5)	
Other	48 (12.6)	3 (7.7)	12 (13.5)	
<b>Do you think that COVID-19 is contagious during the incubation period?</b>				.75
Yes	374 (98.4)	39 (100)	89 (100)	
No	6 (1.6)	0 (0.0)	0 (0.0)	
<b>Is there any effective treatment for COVID-19?</b>				.02
Available	55 (14.5)	1 (2.6)	6 (6.7)	
Unavailable	325 (85.5)	38 (97.4)	83 (93.3)	
<b>When do you think that it is necessary to wear a mask during the COVID-19 pandemic?</b>				.33
When going outside	344 (90.5)	34 (87.2)	76 (85.4)	

**Figure 1.** Ways by which participants obtained COVID-19 information, by occupation group.

### Attitude Toward the COVID-19 Pandemic

The three groups showed significant differences in their level of concern. A higher proportion of medical workers stated they were highly concerned ( $n=28$ , 71.8%), followed by those in other occupations ( $n=47$ , 52.8%) and students ( $n=149$ , 39.2%). The frequency of checking updates was also significantly different among the groups. Regarding the number of times per day participants reported checking for updates about COVID-19, 48.2% of students ( $n=183$ ) checked once per day, as did 46.1% of those in other occupations ( $n=41$ ). On the other hand, the proportion of medical workers that checked updates once per day ( $n=18$ , 46.2%) was slightly lower than the proportion that checked updates more than once per day ( $n=19$ , 48.7%). The

proportion of participants that thought that COVID-19 would not be a global outbreak was 71.3% ( $n=271$ ), 79.5% ( $n=31$ ), and 69.7% ( $n=62$ ) for students, medical workers, and those in other occupations, respectively.

Regarding the village/city lockdown policy, most of the participants from the three groups stated that the policy had a significant impact on their daily life (55.3% of students [ $n=210$ ], 43.6% of medical workers [ $n=17$ ], and 43.8% of those in other occupations [ $n=39$ ]). Nevertheless, most of them had an optimistic attitude toward COVID-19 control (92.1% of students [ $n=350$ ], 94.9% of medical workers [ $n=37$ ], and 92.3% of those in other occupations [ $n=83$ ]). Table 3 displays the attitude toward COVID-19 of participants by group.

**Table 3.** Participants' attitude toward COVID-19 by occupation group.

Questions and responses	Students, n (%)	Medical workers, n (%)	Other occupations, n (%)	P value
<b>How concerned are you about COVID-19?</b>				.002
I don't care	1 (0.3)	0 (0.0)	0 (0.0)	
Low concern	56 (14.7)	2 (5.1)	7 (7.9)	
Medium concern	174 (45.8)	9 (23.1)	36 (39.3)	
High concern	149 (39.2)	28 (71.8)	47 (52.8)	
<b>How often do you check updates about COVID-19?</b>				.007
More than once per day	104 (27.4)	19 (48.7)	33 (37.1)	
Once per day	183 (48.2)	18 (46.2)	41 (46.1)	
Less than once per day	93 (24.5)	2 (5.1)	15 (16.8)	
<b>What impact does the village/city lockdown policy have on your daily life?</b>				.27
Minimal	59 (15.5)	8 (20.5)	17 (19.1)	
Moderate	111 (29.2)	14 (35.9)	33 (37.1)	
Significant	210 (55.3)	17 (43.6)	39 (43.8)	
<b>What is your attitude toward the control of the COVID-19 pandemic?</b>				>.99
Optimistic	350 (92.1)	37 (94.9)	83 (92.3)	
Neutral	9 (2.4)	0 (0.0)	2 (3.3)	
Pessimistic	21 (5.5)	2 (5.1)	4 (4.5)	
<b>Do you think COVID-19 could be a global outbreak?</b>				.50
Yes	109 (28.7)	8 (20.5)	27 (30.3)	
No	271 (71.3)	31 (79.5)	62 (69.7)	

## Discussion

### Principal Findings

This study determined the knowledge and awareness of and attitude toward COVID-19 among students, medical workers, and those in other occupations during the critical period of the outbreak. It demonstrated that participants from different population groups had similar knowledge and awareness of COVID-19. However, medical workers performed significantly better on questions related to professional knowledge. It was confirmed that the lockdown policy had a substantial impact on the daily life of the analyzed groups. Nevertheless, most of the participants believed that COVID-19 would not be a global outbreak and they held an optimistic attitude regarding the control of the COVID-19 pandemic.

In this study, 314 respondents were excluded from the analysis as they did not answer the control items correctly. Of 314 unqualified participants, students accounted for 77.4% ( $n=226$ ), while 17.5% ( $n=51$ ) were in the other occupations group, and 5.1% ( $n=15$ ) were medical workers. No significant differences were found in the occupation distribution between the qualified and unqualified groups. However, the mean age of the unqualified group was significantly lower than that of the qualified group because of the high proportion of students. The difference in distribution of ethnicity and marriage status was not statistically significant between the two groups. Thus, it can be inferred that the participants in the qualified group are representative.

Our study revealed that most of the participants had heard about COVID-19, and a higher proportion of them had heard about it by January, which corresponded with other studies that found that the peak of information seeking about COVID-19 occurred by the end of January in various countries [11,12]. It is worth noting that a certain proportion of medical workers and students had heard about the virus by the end of December 2019.

Regarding general knowledge about COVID-19, our findings showed that almost all participants knew that the host of the virus was probably a wild animal. In addition, most of the participants knew that COVID-19 was different from influenza and SARS, and also had a good understanding of the incubation time of COVID-19, as well as the most susceptible populations. Medical workers showed a more advanced knowledge in their responses to professional questions than the other two groups. Compared with another similar study conducted in India, the level of knowledge and awareness among the participants in our study is higher; most of the respondents in India were passably aware of the basic elements of the disease [13]. Findings similar to ours were reported in another study performed in China that demonstrated that most of the participants were knowledgeable about COVID-19 [14].

Participants of this study used social media (eg, Sina Weibo) most often to obtain information about COVID-19, as well as other apps (WeChat and QQ). Traditional media such as newspapers and magazines were used less often to get updates. Similar findings were found in other studies that analyzed where people searched for information about COVID-19 [15-17]. A

possible explanation could be that social media apps have become popular and easily accessible compared with traditional media sources, thus speeding up the spread of information.

Regarding preventive measures, our study found that most of the participants would wear a mask when outside and in crowded places. However, there is controversial information about the effectiveness of wearing a mask among the general population [18]. In East Asia, especially in China, wearing a mask has been one of the main preventive measures recommended by the government since the beginning of the outbreak. This was possibly one of the main differences in response compared with other places in the world where the use of masks was not compulsory for the general population [19]. In fact, wearing a mask has contributed much to the effective control of COVID-19 in China. Furthermore, a wide range of countries and regions have recommended their citizens wear a mask if social distancing cannot be ensured.

Most of the study participants, especially medical workers, reported that they were highly concerned about COVID-19. Similarly, another web-based study conducted in China suggested that 97.1% of participants paid close attention to COVID-19 by checking updates more than once per day [20]. A possible reason for frequent update checking is the psychological stress associated with the pandemic that the public has experienced. Another study in China showed that more than 90% of the study population experienced concern and stress [21]. The high level of concern among medical workers could be due to their higher risk of exposure to COVID-19 and their fear of getting infected. A similar study performed in Henan, China, demonstrated that 85% of health care workers were afraid of being infected at work [22].

The lockdown policy has been an important preventive measure to deter the spread of COVID-19 in several countries [23,24], and it has been effective in China [25]. In our study, most of the respondents reported that the lockdown policy had a significant impact on their daily life. In this regard, other studies also confirmed that lockdown measures can have a substantial impact on the mental health of the population and affect people's daily lives due to social distancing and decreased physical activity [26]. Moreover, another study that reviewed the impact of COVID-19 on university students showed that most of them were concerned about the effects that the outbreak would have on their academic performance. Most universities closed and classes were online [27]; this may be why students considered the impact of lockdown significant. Nevertheless, as a positive outcome, the lockdown policy also provided the opportunity and time for families to stay close, especially if they learned to cope with tensions that occurred due to changes in the family routine. However, several countries, including China, reported an increase in domestic violence during lockdown [28]. There was also a general concern about parenting during the lockdown and how families could deal with the stress of the crisis [29].

Despite being highly concerned about COVID-19, more than 90% of the respondents declared to have an overall optimistic attitude toward COVID-19 control. Similar findings were reported in another cross-sectional web-based study among health care workers globally, especially in Asia, where 78% of

the participants held a positive perception of COVID-19 [30]. In contrast, a study performed in Uganda showed that only 21% of the participants had a good attitude toward the disease [31]. A possible cause of these differences in attitude toward the COVID-19 pandemic may be the social and economic gap among countries, which might influence the availability of medical services and sanitary conditions in houses and workplaces. The Human Development Index (HDI) might be an adequate measurement of this gap. In Asia, the HDI ranges from 0.647 to 0.866, while in Uganda it is 0.528. Similarly, more than half of the respondents said that COVID-19 could not become a global outbreak. This trend might be related to the low international incidence of cases when the survey was conducted.

This study evaluated the knowledge and awareness of and attitude toward COVID-19 among different populations in Central China, providing meaningful findings that might be useful for other countries and regions facing the COVID-19 pandemic. Nevertheless, limitations still exist. First, the sample size was small and unbalanced, especially for medical workers

and the other occupations group, which might limit the generalizability of findings. Therefore, the interpretation of the results should be cautious. Second, the findings related to mental health were not included in the current analysis. Finally, the sample representing the target population could be limited to some extent because all participants came from the same geographical region.

## Conclusions

In general, the findings show that the participants have an adequate background knowledge of COVID-19 and hold an optimistic attitude toward the control of the disease. However, most of the participants were highly concerned about the pandemic and said that the lockdown policy had a significant impact on their daily life, suggesting a need for effective preventive measures to relieve emotional and psychological stress. Finally, this research offers essential evidence for COVID-19 pandemic control prior to the availability of a prophylactic vaccine. Further studies are still needed to identify the specific impact that the COVID-19 pandemic has had on the lives and mental health of the population.

## Conflicts of Interest

None declared.

## References

1. Cutler D. How Will COVID-19 Affect the Health Care Economy? JAMA 2020 Jun 09;323(22):2237. [doi: [10.1001/jama.2020.7308](https://doi.org/10.1001/jama.2020.7308)]
2. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020. URL: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> [accessed 2020-05-26]
3. Azman AS, Luquero FJ. From China: hope and lessons for COVID-19 control. The Lancet Infectious Diseases 2020 Jul;20(7):756-757. [doi: [10.1016/s1473-3099\(20\)30264-4](https://doi.org/10.1016/s1473-3099(20)30264-4)]
4. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? BMJ 2014 Oct 14;349:g6178. [doi: [10.1136/bmj.g6178](https://doi.org/10.1136/bmj.g6178)] [Medline: [25315514](https://pubmed.ncbi.nlm.nih.gov/25315514/)]
5. Kamara S, Walder A, Duncan J, Kabbedijk A, Hughes P, Muana A. Mental health care during the Ebola virus disease outbreak in Sierra Leone. Bull World Health Organ 2017 Oct 31;95(12):842-847. [doi: [10.2471/blt.16.190470](https://doi.org/10.2471/blt.16.190470)]
6. Chen Q, Liang M, Li Y, Guo J, Fei D, Wang L, et al. Mental health care for medical staff in China during the COVID-19 outbreak. The Lancet Psychiatry 2020 Apr;7(4):e15-e16. [doi: [10.1016/s2215-0366\(20\)30078-x](https://doi.org/10.1016/s2215-0366(20)30078-x)]
7. Li SW, Wang Y, Yang YY, Lei XM, Yang YF. Investigation on the influencing factors for anxiety related emotional disorders of children and adolescents with home quarantine during the prevalence of coronavirus disease 2019. Chinese Journal of Child Health Care 2020;28(4):407-410. [doi: [10.11852/zgetbjzz2020-0169](https://doi.org/10.11852/zgetbjzz2020-0169)]
8. Yang Y, Li W, Zhang Q, Zhang L, Cheung T, Xiang Y. Mental health services for older adults in China during the COVID-19 outbreak. The Lancet Psychiatry 2020 Apr;7(4):e19. [doi: [10.1016/s2215-0366\(20\)30079-1](https://doi.org/10.1016/s2215-0366(20)30079-1)]
9. Zhai Y, Du X. Mental health care for international Chinese students affected by the COVID-19 outbreak. The Lancet Psychiatry 2020 Apr;7(4):e22. [doi: [10.1016/s2215-0366\(20\)30089-4](https://doi.org/10.1016/s2215-0366(20)30089-4)]
10. Lin Y, Hu Z, Alias H, Wong LP. Knowledge, Attitudes, Impact, and Anxiety Regarding COVID-19 Infection Among the Public in China. Front Public Health 2020 May 27;8:236 [FREE Full text] [doi: [10.3389/fpubh.2020.00236](https://doi.org/10.3389/fpubh.2020.00236)] [Medline: [32574305](https://pubmed.ncbi.nlm.nih.gov/32574305/)]
11. Mesa Vieira C, Franco OH, Gómez Restrepo C, Abel T. COVID-19: The forgotten priorities of the pandemic. Maturitas 2020 Jun;136:38-41 [FREE Full text] [doi: [10.1016/j.maturitas.2020.04.004](https://doi.org/10.1016/j.maturitas.2020.04.004)] [Medline: [32386664](https://pubmed.ncbi.nlm.nih.gov/32386664/)]
12. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the Internet: Infodemiology Study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: [10.2196/18444](https://doi.org/10.2196/18444)] [Medline: [32250960](https://pubmed.ncbi.nlm.nih.gov/32250960/)]
13. Roy D, Tripathy S, Kar SK, Sharma N, Verma SK, Kaushal V. Study of knowledge, attitude, anxiety & perceived mental healthcare need in Indian population during COVID-19 pandemic. Asian J Psychiatr 2020 Jun;51:102083 [FREE Full text] [doi: [10.1016/j.ajp.2020.102083](https://doi.org/10.1016/j.ajp.2020.102083)] [Medline: [32283510](https://pubmed.ncbi.nlm.nih.gov/32283510/)]



14. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. *Int J Biol Sci* 2020;16(10):1745-1752 [[FREE Full text](#)] [doi: [10.7150/ijbs.45221](#)] [Medline: [32226294](#)]
15. Hu D, Lou X, Xu Z, Meng N, Xie Q, Zhang M, et al. More effective strategies are required to strengthen public awareness of COVID-19: Evidence from Google Trends. *J Glob Health* 2020 Jun;10(1):011003 [[FREE Full text](#)] [doi: [10.7189/jogh.10.011003](#)] [Medline: [32373339](#)]
16. Liu HY, Tang F, Deng L. Analysis of public anxiety status and influencing factors during the novel coronavirus pneumonia epidemic. *Chinese Nursing Research* 2020;34(9):1646-1648. [doi: [10.12102/j.issn.1009-6493.2020.09.035](#)]
17. Hu YK, Zhang G, Li Z, Yang J, Mo L, Zhang X, et al. [Knowledge, attitudes, and practices related to COVID-19 pandemic among residents in Hubei and Henan Provinces]. *Journal of Southern Medical University* 2020 May 30;40(5):733-740 [[FREE Full text](#)] [doi: [10.12122/j.issn.1673-4254.2020.05.20](#)] [Medline: [32897216](#)]
18. Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ. Rational use of face masks in the COVID-19 pandemic. *The Lancet Respiratory Medicine* 2020 May;8(5):434-436. [doi: [10.1016/s2213-2600\(20\)30134-x](#)]
19. Javid B, Weekes MP, Matheson NJ. Covid-19: should the public wear face masks? *BMJ* 2020 Apr 09;369:m1442. [doi: [10.1136/bmj.m1442](#)] [Medline: [32273278](#)]
20. Chen Y, Jin YL, Zhu LJ, Fang ZM, Wu N, Du MX, et al. [The network investigation on knowledge, attitude and practice about COVID-19 of the residents in Anhui Province]. *Chinese Journal of Preventive Medicine* 2020 Apr 06;54(4):367-373. [doi: [10.3760/cma.j.cn112150-20200205-00069](#)] [Medline: [32268643](#)]
21. Wang L, Xu SY, Huang Y, Z Y, Xu Y, Yao D, et al. Knowledge, attitude and practice about novel coronavirus disease 2019 among residents in Zhejiang province during epidemic period: an online survey. *Chinese Journal of Public Health* 2020;36(5):661-664. [doi: [10.11847/zgggws1128820](#)]
22. Zhang M, Zhou M, Tang F, Wang Y, Nie H, Zhang L, et al. Knowledge, attitude, and practice regarding COVID-19 among healthcare workers in Henan, China. *J Hosp Infect* 2020 Jun;105(2):183-187 [[FREE Full text](#)] [doi: [10.1016/j.jhin.2020.04.012](#)] [Medline: [32278701](#)]
23. The Lancet. India under COVID-19 lockdown. *Lancet* 2020 Apr 25;395(10233):1315 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(20\)30938-7](#)] [Medline: [32334687](#)]
24. Dzobo M, Chitungo I, Dzinamarira T. COVID-19: a perspective for lifting lockdown in Zimbabwe. *Pan Afr Med J* 2020 Apr 30;35(Suppl 2):13 [[FREE Full text](#)] [doi: [10.11604/pamj.2020.35.2.23059](#)] [Medline: [32528624](#)]
25. Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Schubert J, Bania J, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *J Travel Med* 2020 May 18;27(3):taaa037 [[FREE Full text](#)] [doi: [10.1093/jtm/taaa037](#)] [Medline: [32181488](#)]
26. Zhang SX, Wang Y, Rauch A, Wei F. Unprecedented disruption of lives and work: Health, distress and life satisfaction of working adults in China one month into the COVID-19 outbreak. *Psychiatry Res* 2020 Jun;288:112958 [[FREE Full text](#)] [doi: [10.1016/j.psychres.2020.112958](#)] [Medline: [32283450](#)]
27. Sahu P. Closure of Universities Due to Coronavirus Disease 2019 (COVID-19): Impact on Education and Mental Health of Students and Academic Staff. *Cureus* 2020 Apr 04;12(4):e7541 [[FREE Full text](#)] [doi: [10.7759/cureus.7541](#)] [Medline: [32377489](#)]
28. Vieira PR, Garcia LP, Maciel ELN. [The increase in domestic violence during the social isolation: what does it reveals?]. *Rev Bras Epidemiol* 2020;23:e200033 [[FREE Full text](#)] [doi: [10.1590/1980-549720200033](#)] [Medline: [32321005](#)]
29. Cluver L, Lachman JM, Sherr L, Wessels I, Krug E, Rakotomalala S, et al. Parenting in a time of COVID-19. *Lancet* 2020 Apr 11;395(10231):e64 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(20\)30736-4](#)] [Medline: [32220657](#)]
30. Bhagavathula AS, Aldhaleei WA, Rahmani J, Mahabadi MA, Bandari DK. Knowledge and Perceptions of COVID-19 Among Health Care Workers: Cross-Sectional Study. *JMIR Public Health Surveill* 2020 Apr 30;6(2):e19160 [[FREE Full text](#)] [doi: [10.2196/19160](#)] [Medline: [32320381](#)]
31. Olum R, Chekwech G, Wekha G, Nassozi DR, Bongomin F. Coronavirus Disease-2019: Knowledge, Attitude, and Practices of Health Care Workers at Makerere University Teaching Hospitals, Uganda. *Front Public Health* 2020 Apr 30;8:181 [[FREE Full text](#)] [doi: [10.3389/fpubh.2020.00181](#)] [Medline: [32426320](#)]

## Abbreviations

**HDI:** Human Development Index  
**PHEIC:** Public Health Emergency of International Concern  
**SARS:** severe acute respiratory syndrome  
**WHO:** World Health Organization



*Edited by G Eysenbach; submitted 18.07.20; peer-reviewed by J Li, J Wang; comments to author 30.07.20; revised version received 20.08.20; accepted 01.09.20; published 15.10.20.*

*Please cite as:*

Xu H, Gonzalez Mendez MJ, Guo L, Chen Q, Zheng L, Chen P, Cao X, Liu S, Sun X, Zhang S, Qiao Y

Knowledge, Awareness, and Attitudes Relating to the COVID-19 Pandemic Among Different Populations in Central China: Cross-Sectional Survey

*J Med Internet Res* 2020;22(10):e22628

URL: <https://www.jmir.org/2020/10/e22628>

doi: [10.2196/22628](https://doi.org/10.2196/22628)

PMID: [32886623](https://pubmed.ncbi.nlm.nih.gov/32886623/)

©Huifang Xu, Maria Jose Gonzalez Mendez, Lanwei Guo, Qiong Chen, Liyang Zheng, Peipei Chen, Xiaoqin Cao, Shuzheng Liu, Xibin Sun, Shaokai Zhang, Youlin Qiao. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 15.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# COVID-19 Self-Reported Symptom Tracking Programs in the United States: Framework Synthesis

Tracey Pérez Koehlmoos<sup>1</sup>, PhD, MHA; Miranda Lynn Janvrin<sup>1,2</sup>, MPH; Jessica Korona-Bailey<sup>1,2</sup>, MPH; Cathaleen Madsen<sup>1,2</sup>, PhD; Rodney Sturdivant<sup>1,2</sup>, PhD, MSc

<sup>1</sup>Uniformed Services University, Bethesda, MD, United States

<sup>2</sup>Health Services Research Program, Henry M Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, United States

**Corresponding Author:**

Miranda Lynn Janvrin, MPH

Health Services Research Program

Henry M Jackson Foundation for the Advancement of Military Medicine

6720B Rockledge Dr

Suite 605

Bethesda, MD

United States

Phone: 1 6035403059

Email: [miranda.janvrin.ctr@usuhs.edu](mailto:miranda.janvrin.ctr@usuhs.edu)

## Abstract

**Background:** With the continued spread of COVID-19 in the United States, identifying potential outbreaks before infected individuals cross the clinical threshold is key to allowing public health officials time to ensure local health care institutions are adequately prepared. In response to this need, researchers have developed participatory surveillance technologies that allow individuals to report emerging symptoms daily so that their data can be extrapolated and disseminated to local health care authorities.

**Objective:** This study uses a framework synthesis to evaluate existing self-reported symptom tracking programs in the United States for COVID-19 as an early-warning tool for probable clusters of infection. This in turn will inform decision makers and health care planners about these technologies and the usefulness of their information to aid in federal, state, and local efforts to mobilize effective current and future pandemic responses.

**Methods:** Programs were identified through keyword searches and snowball sampling, then screened for inclusion. A best fit framework was constructed for all programs that met the inclusion criteria by collating information collected from each into a table for easy comparison.

**Results:** We screened 8 programs; 6 were included in our final framework synthesis. We identified multiple common data elements, including demographic information like race, age, gender, and affiliation (all were associated with universities, medical schools, or schools of public health). Dissimilarities included collection of data regarding smoking status, mental well-being, and suspected exposure to COVID-19.

**Conclusions:** Several programs currently exist that track COVID-19 symptoms from participants on a semiregular basis. Coordination between symptom tracking program research teams and local and state authorities is currently lacking, presenting an opportunity for collaboration to avoid duplication of efforts and more comprehensive knowledge dissemination.

(*J Med Internet Res* 2020;22(10):e23297) doi:[10.2196/23297](https://doi.org/10.2196/23297)

**KEYWORDS**

COVID-19; coronavirus; framework analysis; information resources; patient-reported outcome measures; self-reported; surveillance; monitoring; symptom tracking; synthesis

## Introduction

### Background

A 2019 outbreak of febrile respiratory illness in Wuhan, China, quickly evolved into the COVID-19 pandemic [1]. The disease has affected over 200 countries and territories worldwide. Globally, there are more than 18 million confirmed cases and over 700,000 deaths attributed to this flu-like illness, as of August 6, 2020 [2]. In the United States alone, there are more than 4.5 million confirmed cases and over 150,000 deaths [3]. The true number of those affected may be much higher due to the slow rollout and lack of availability of testing in the United States compared to other countries [4].

The United States, as well as other countries, has combatted this pandemic and sought to flatten the curve via social distancing, testing, isolation, and contact tracing [5]. Despite best efforts, the virus spread quickly with serious implications. In the first month of testing, the hospitalization rate was 4.6 per 100,000 people in the United States. Hospitalization rates were highest among adults over 65 years as well as those with underlying conditions [6]. At the present time, there is no specific antiviral treatment for COVID-19. Management of symptoms focuses on supportive care and oxygen therapy, both of which involve a plethora of hospital resources [5]. Modeling of COVID-19 shows that the pandemic has the potential to cause regional shortages of hospital beds, intensive care unit (ICU) beds, ventilators, and medical staff, which could lead to difficult ethical decisions [7]. A recent study suggests that COVID-19 will likely become endemic like cold and flu viruses [8]. There is a need to predict where resources should be distributed before potential patients with COVID-19 enter the hospital setting to alleviate strain on medical staff and facilities.

Epidemiological surveillance is fundamental in coordinating both immediate and long-term strategies for the detection and prevention of infectious disease outbreaks [9,10]. However, since collecting and disseminating these data take several weeks, during highly transmissible outbreaks, it may not be entirely reflective of the current prevalence of the disease. As these data are used to inform health authorities and prompt a public response, the resulting time delay can lead to inappropriate or inadequate response to actual need. Additionally, the collected data may be incomplete or insufficient to discern regional demographics that may impact effective intervention and treatment [11,12].

To overcome the limitations of epidemiological surveillance, internet-based technologies have been developed to estimate and monitor real-time changes in population, soliciting participation from the public at large [12,13]. One approach that has been introduced is self-reported symptom tracking. Symptom tracking is a form of crowd-sourced participatory surveillance that solicits individuals to report their health status on a daily or weekly basis, often with emails or notifications to prompt timely response, allowing researchers to see potential changes in the population before seeing changes in clinical presentation at hospitals and medical centers. Symptom tracking is used primarily to track and forecast influenza activity throughout the country; however, researchers have been looking

to apply this technology to other diseases, such as COVID-19 [12,13]. Participatory surveillance such as this may prove vital to complement epidemiological surveillance during highly transmissible epidemics, as it allows for the detection of outbreaks before they reach the clinical threshold, affording more time for logistical support and appropriate allocation of resources [11]. Research by Baltrusaitis et al [13] indicates that collected participatory surveillance of influenza later correlated with confirmed epidemiological surveillance data. With the current highly transmissible and deadly COVID-19 pandemic putting a strain on portions of the United States health care system, participatory surveillance is more important than ever to bolster local prevention efforts [14].

### Research Purpose

This study uses a framework synthesis to inform decision making about the utility of existing self-reported symptom tracking programs for COVID-19, with a focus on the US population, as an early-warning tool for probable clusters of infection. Due to the rapidly changing nature of both the pandemic and work in this area, this research will be updated at 6- and 12-month intervals.

### Objective

The purpose of this framework analysis is to assess the number and scope of self-reported symptom tracker programs focused on the United States and COVID-19. An innovative best fit framework analysis was chosen because of its strength, utility, and appropriateness in drawing conclusions for an evolving subject [15,16]. According to Booth and Carroll [17], the best fit framework approach is considered a highly structured and pragmatic methodology for research synthesis suited for qualitative research with specific questions, a limited time frame, and issues that have been previously identified; this served the purpose of our research well [18]. The outcomes of this synthesis and its updates should inform decision makers and health care planners about these technologies and the information they can ascertain from them in order to aid in federal, state, and local efforts to combat the pandemic both now and in the future.

## Methods

A framework analysis was conducted to assess symptom tracking programs. A best fit framework was constructed by collating information collected from each program into a table for easy comparison between programs.

### Target Population

This framework synthesis sought to identify programs that track COVID-19 symptoms in the US population for all ages, genders, and ethnicities. Inclusion and exclusion criteria are shown below:

- Inclusion criteria: programs were included if they aimed to capture and geographically collate self-reported potential symptoms of COVID-19 and if they were available for use in the United States. For our purpose, a symptom tracking tool is defined as a program that allows individuals to report symptoms of COVID-19 to identify geographic areas with emerging or changes in progression of disease.

- Exclusion criteria: programs were excluded if they did not track specific symptoms for COVID-19, were symptom checkers for individual use only, or were not targeting the US population.

### Program Identification

Programs were identified using Google search for keywords (“symptom trackers covid,” “symptom trackers coronavirus,” “symptom tracking covid,” “symptom tracking coronavirus,” “daily symptom tracking covid,” “daily symptom tracking coronavirus,” “self-reporting covid,” “self-reporting coronavirus”). The time frame for the search for programs ranges from April 7, 2020, to May 9, 2020. Further, we used snowball sampling to identify other symptom tracker programs for COVID-19.

### Screening Method

Reviewers (JK, MJ, TK) screened programs to determine if inclusion criteria were met. Reviewers (MJ, JK, TK) then extracted data from program websites using a standardized form. To complete the collection of information not available via the program webpages, we contacted the managers of the programs via email.

### Synthesis Method

Data relating to program characteristics were extracted from all included programs and organized into a table format, which was used to guide data collection and build the framework for analysis. Data were then synthesized in order to form meaningful statements about the programs.

## Results

We identified 6 programs that met the inclusion criteria. Information was gathered from the public webpages of all

eligible symptom trackers (BeatCOVID19Now, COVIDcast, COVIDNearYou, COVID Symptom Tracker, HelpBeatCOVID19, and HowWeFeel) (Table 1). Two programs, C19Check and the Department of Defense’s MySymptoms.mil, were excluded from our synthesis since they are symptom checkers that do not identify probable clusters of emerging infection.

All of the included programs were affiliated with a university, school of medicine, or school of public health. Half of the programs (n=3) included were based in Boston, Massachusetts, and affiliated with Harvard University (COVIDNearYou, COVID Symptom Tracker, and HowWeFeel), with COVIDNearYou also collecting data from participants in Canada and Mexico. Two other programs are based elsewhere within the United States (COVIDcast and HelpBeatCOVID19), and one is based in Australia, designed for international use (BeatCOVID19Now).

The number of responses, defined as unique symptom entries by an individual, to each program varied widely, with the lowest being ~27,000 (BeatCOVID19Now) and the highest being 2,573,240 (COVIDcast). COVID Symptom Tracker collected data from patients currently enrolled in large cohort studies and clinical trials not related to COVID-19 and had obtained much of their initial influx of responses through that mechanism. Two-thirds of the programs had fewer than 100,000 responses. Three programs utilized a website to collect data, while two exclusively used an app available for both Apple and Android devices (COVID Symptom Tracker and HowWeFeel), and only one utilized a survey on a social media platform (Facebook). While most of the programs had no form of follow-up with participants, COVIDNearYou and HelpBeatCOVID19 sent text message reminders, and COVID Symptom Tracker sent phone notifications every third day.

**Table 1.** Overview of self-reported symptom tracker programs.

Characteristic	BeatCOVID19Now	COVIDcast	COVIDNearYou	COVID Symptom Tracker	HelpBeatCOVID19	HowWeFeel
Host institution and partners	Swinburne University of Technology	Carnegie Mellon University Delphi Research Group; Facebook	Harvard Medical School; Boston Children's Hospital; Ending Pandemics; Google; Centers for Disease Control and Prevention (CDC)	Harvard TH Chan School of Public Health; Massachusetts General Hospital; King's College London; Stanford University School of Medicine; Zoe Global Limited	University of Alabama; Alabama Department of Public Health	Harvard TH Chan School of Public Health; Massachusetts Institute of Technology; Institute for Quantitative Social Science; McGovern Institute; Howard Hughes Medical Institute; Weizmann Institute of Science; Pinterest; Feeding America; Alex's Lemonade Stand; Chartio; Bill & Melinda Gates Foundation
Location	Melbourne, Australia	Pittsburgh, Pennsylvania, USA	Boston, Massachusetts, USA	Boston, Massachusetts, USA	Birmingham, Alabama, USA	Boston, Massachusetts, USA
Funding sources	Swinburne University of Technology	None	Ending Pandemics Crowdsourcing	Mass General Wellcome Trust (UK)	University of Alabama	Bill & Melinda Gates Foundation Crowdsourcing
Intended participants	Worldwide, 18+ years	United States residents, 18+ years	United States, Canada, and Mexico residents, 18+ years	United States residents, 18+ years; participants from other internal studies including RCTs	United States residents, 18+ years; particular focus on Alabama and neighboring states	United States residents, 18+ years
Date symptom tracker was initiated	March 26, 2020	April 6, 2020	March 22, 2020	April 4, 2020	Not available	April 3, 2020
Number of responses to date <sup>a</sup>	27,000+	2,573,240	54,000+	98,000+	57,000+	1,000,000+
Mechanism of recruiting participants/platform	Website; app in development <sup>a</sup>	Survey via Facebook	Website	Apple App Store, Google Play Store	Website	Apple App Store, Google Play Store
Follow-up	None	None	None	Daily phone notifications	Text messages every 3 days	None
Frequency of reporting	Daily	Daily	Weekly	Live data	Live data	Daily
Availability of summary tables for external synthesis/utilization	Yes	Yes	Yes	Yes	No	No <sup>a</sup>
Intended audience for the product	Public at large; state and local public health officials; international health organizations	Public at large; state and local public health officials; US policymakers; health care providers; health care systems	Public at large; CDC and national public health organizations; state and local public health officials; researchers; health care providers; health care systems	Public at large; participants of internal studies	Public at large; neighboring states; state and local health officials; local policymakers	Public at large; state and local public health officials



Characteristic	BeatCOVID19Now	COVIDcast	COVIDNearYou	COVID Symptom Tracker	HelpBeatCOVID19	HowWeFeel
Publicly available data privacy statement	Yes	Yes	Yes	Yes	Yes	Yes

<sup>a</sup>This data was collected at the time that the synthesis was performed and is subject to change.

The programs collected a variety of data elements, but several were common among them (Table 2). All of the symptom trackers collected demographic data on the participant's age, gender, and zip code. They also all collected information on symptoms experienced by the participant, although the time frame considered varied from the present to 7 days prior. Additionally, every program asked if the participant had been tested for COVID-19 at the time of the survey. Five of the six trackers also asked for information on any chronic conditions that the participant is experiencing, and if they are or are not a smoker.

Some of the programs had special interest in certain topics that were not explored by others. Only four of the programs asked the participant if they had been exposed to anyone who had COVID-19, while two asked if the participant came into direct contact with the public. Two programs asked if participants had received an annual flu shot this past year. Two programs asked questions related to the impact of the pandemic on participant's mental health. Two programs asked the participant to answer questions about others in their household in addition to themselves.

**Table 2.** Data elements across programs.

Data elements	Beat- COVID19Now	COVIDcast	COVIDNearYou	COVID Symptom Tracker	HelpBeatCOVID19	HowWeFeel
Is the survey being completed on behalf of another person?					✓	
Age	✓	✓	✓	✓	✓	✓
Gender/sex	✓	✓	✓	✓	✓	✓
Race/ethnicity	✓				✓	
Zip code	✓	✓	✓	✓	✓	✓
Number of people in the household	✓	✓			✓	
Employment status	✓					
Languages spoken in the household	✓					
Is the participant an essential worker?	✓					
Is the participant a health care worker?	✓			✓		
International travel within the past 2 months?	✓					
Has the participant traveled out of state within the past 5 days?		✓				
Travel within the past 2 weeks?			✓			
Does the participant come in direct contact with the public?				✓	✓	
How many people has the participant had direct contact with outside of their household?		✓				
Has the participant gone outside for work within the past 5 days?		✓				
What activities has the participant engaged in outside of their household?	✓					
Has the participant been in contact with health care professionals?			✓			
Has the participant visited a long-term care facility or nursing home within 5 days?		✓				
To what extent is the participant complying with social distancing guidelines?		✓				
How many days has the participant spent in quarantine or social isolation?			✓			
Has the participant been quarantined over the past 2 weeks?				✓		
Has the participant been quarantined over the past 24 hours?						✓
How is the participant feeling today? (good/not good)			✓	✓	✓	✓
Has the participant been exposed to anyone with COVID-19?		✓	✓	✓	✓	✓
Has the participant been tested for COVID-19?	✓	✓	✓	✓	✓	✓
Does the participant suspect they have COVID-19 despite not being tested?				✓		✓
Symptoms of the participant over the last 24 hours	✓					

Data elements	Beat-COVID19Now	COVIDcast	COVIDNearYou	COVID Symptom Tracker	HelpBeatCOVID19	HowWeFeel
Symptoms among the participant or household member(s) within 24 hours		✓				
Symptoms currently being experienced by the participant			✓	✓		✓
Symptoms over the last 7 days					✓	
How many days has the participant been experiencing symptoms?		✓				
What date did the participant begin experiencing symptoms?			✓			
Has the participant had difficulty completing normal activities over the past 24 hours?	✓					
Is anyone within the participant's household experiencing symptoms?					✓	✓
Number of people in the household who are sick		✓				
Number of people the participant knows in the community who are sick		✓				
Has the participant been to the hospital within the past 24 hours?		✓				
Is the participant at home or hospitalized?				✓		
Is the participant able to move freely?				✓		
Does the participant require outside help on a regular basis?				✓		
If the participant needs help, can they get it from someone close to them?				✓		
Highest temperature		✓				
Does the participant have a non-COVID-19 respiratory illness?	✓					
Impact on immediate mental health or changes in mood or behavior	✓	✓				
Is the participant worried about their ability to engage in daily activities or about the security of their future?	✓	✓				
Does the participant have health problems that require staying indoors regularly?				✓		
Chronic conditions	✓	✓		✓	✓	✓
Smoking status				✓	✓	✓
Height				✓		
Weight				✓	✓	
Pregnancy status					✓	✓
Has the participant had the flu vaccination?	✓	✓	✓			
Is the participant currently taking aspirin?				✓		
Is the participant currently taking nonsteroidal anti-inflammatory drugs (NSAIDs)?				✓		

Data elements	Beat-COVID19Now	COVIDcast	COVIDNearYou	COVID Symptom Tracker	HelpBeatCOVID19	HowWeFeel
Is the participant currently taking blood pressure medication?				✓		
Is the participant currently taking immunosuppressants?				✓		
Does the participant have access to transportation?					✓	
Does the participant have health insurance?					✓	
Type of domicile					✓	
Can the participant afford a medical co-pay if needed?					✓	
Has the participant completed the survey before?					✓	

## Discussion

### Principal Results

Self-reported symptom trackers have been shown to be beneficial in tracking and monitoring the spread and progression of influenza each year and may prove to be vital as the United States continues to loosen shelter-in-place guidelines across the country. Due to the nature of the rapidly changing pandemic, this resource will be updated at both 6- and 12-month intervals to better reflect the evolving pandemic response.

Two of the programs were created by groups who already have existing infrastructure for tracking influenza outbreaks each year, BeatCOVID19Now, which is a derivative of Flu-iiQ, and COVIDNearYou, the sister tracker to FluNearYou. Flu-iiQ, in particular, was developed to solicit patient-reported outcome measures during large-scale clinical trials to measure the presence or absence of disease within a small subset of a population, allowing for extremely sensitive measurements without requiring thousands of responses [19]. The flexibility of these programs to track symptoms associated with diverse flu-like illness is imperative in identifying outbreaks of disease both for the purposes of this current pandemic as well as future flu and other respiratory disease outbreaks [20].

The data elements collected varied between programs, but all asked for zip code data, which means that even groups that do not currently have their data geolocated on maps have the potential to do so in the future in order to make data accessible to state and local health officials. They also all collect data regarding testing status, which enables local, state, or national program managers or planners to see the impact of current testing expansion efforts. Almost all of the programs asked about race and/or ethnicity, which may highlight racial disparities in testing, symptoms, unemployment status, and other chronic health conditions. The similarities in the data elements being collected by the different programs indicates that collaboration to build a larger, single picture is a possibility; standardization could be beneficial to the programs and to the local leaders and planners, health care providers, and researchers who would receive the outputs. The differences in collected

data highlight areas of focus between the programs that other programs may want to consider incorporating as well.

Notable differences between the programs include unique data elements as well as the manner of recruitment. Two of the programs, BeatCOVID19Now and COVIDcast, are collecting information related to the mental health impact of the pandemic. This topic is currently being discussed in the scientific community since individuals with current mental health conditions can be at higher risk for infections [21,22]. Additionally, mental health conditions can be made worse by the anxiety and fear brought on by the pandemic [21]. Individuals without existing mental health conditions may develop emotional responses to the pandemic similar to disaster scenarios, particularly those who are working in response to the pandemic or those who are more susceptible to infection. Quarantine in general can spur a number of emotional responses that can remain after stay-at-home orders are lifted [22]. These programs could help to track the effect of mental health during the COVID-19 pandemic and help to inform prevention efforts for future pandemics requiring social isolation and quarantine. Another key difference was the reach of each program. Programs that partnered with or heavily relied on social media platforms (COVIDcast and HowWeFeel) had significantly more responses than those that did not utilize social media, suggesting that social media is a powerful recruitment tool for these efforts, even more so now since people depend on these platforms to stay connected due to social distancing measures. Therefore, its use should be considered by other groups going forward.

One of the notable results of this synthesis is the demonstrated overlap or duplication of effort between the programs. Each program is competing for the same group of potential respondents, who are more than likely going to be completing only one group's survey. Without ongoing coordination between groups, the beneficiaries of their work—the public, lawmakers, state and local health care officials, etc—will not obtain information reflective of the full potential of symptom tracking. Although many of the groups recognize this, active collaboration between the groups has been a difficult process, even among

the groups located in the same city (eg, Boston, Massachusetts) and based in the same institution.

A key challenge facing these programs is a lack of recognition at the national level. Only one of the trackers, COVIDNearYou, had a partnership with the Centers for Disease Control and Prevention, an extension of their ongoing partnership for FluNearYou. Despite this long-term collaboration, there is no outward support from the agency urging people to engage with this new program. The lack of local, state, or national promotion or outward partnership further exacerbates the potential for gaps between programs. Additionally, there is the potential that endorsement by local authorities or agencies could increase the number of responses, reaching people who were previously unaware of these programs and influencing them to contribute their data, which would in turn would allow for more complete data. This has been found to be the case in the United Kingdom, where the National Health System has endorsed the sister application to COVID Symptom Tracker, based at King's College in London. Because of this, at the time of interview, they had received ten times as many responses as their US counterparts [23].

### Limitations

Several limitations must be acknowledged for this study. First, our analysis was limited to English language programs, and therefore may have missed nuances of data collection which are more important to non-English speaking residents. Second, although the speed of framework analysis enables rapid evaluation of commonalities, it does not provide the in-depth rigor of a full systematic review. Third, our collected data did

evaluate differences in the number of responses to each program but not analyze the effectiveness, market penetration, or user demographics of evaluated programs. Fourth, we recognize that program participation is limited to only those who have access to the internet or cellular phone service, creating an unintended disparity among respondents based on their access to and utilization of technology. Therefore, the underlying reasons for the difference in response rate remain beyond the scope of this study. Last, this synthesis does not provide critical appraisal of programs or evaluate programs for effectiveness.

### Conclusion

Self-reported symptom tracking programs offer potential benefits as states and counties continue to reopen after the large-scale stay-at-home orders. Frequently reported data with high participation in geographic areas would allow officials to better monitor potential emerging hotspots and institute public health policy and reallocate resources more quickly to combat the spread of disease. However, there are unique challenges to address with self-reported symptom tracking programs to ensure successful implementation. Recognition or endorsement at the national, state, or local levels; increased funding to expand social media advertisements and partnerships; and collaboration between existing programs to generate a more comprehensive data picture would be essential steps in bolstering the utility of symptom tracking programs to achieve optimal effectiveness. If these challenges are addressed and symptom tracking programs become more widely used, the reopening process could be safer in the short term with the potential to monitor communities more closely for long-term management of the COVID-19 pandemic or future outbreaks.

### Acknowledgments

The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views, assertions, opinions, or policies of the Uniformed Services University of the Health Sciences, the Henry M Jackson Foundation for the Advancement of Military Medicine, or the Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the US Government.

This study was funded through a grant from the US Department of Defense, Defense Health Agency (award #HU0001-17-2-0001). The funding agency played no part in the design, analysis, and interpretation of data or writing of the manuscript.

### Conflicts of Interest

None declared.

### References

1. Hui D, I Azhar E, Madani T, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* 2020 Feb;91:264-266 [FREE Full text] [doi: [10.1016/j.ijid.2020.01.009](https://doi.org/10.1016/j.ijid.2020.01.009)] [Medline: [31953166](https://pubmed.ncbi.nlm.nih.gov/31953166/)]
2. COVID19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins Corona Virus Resource Center. URL: <https://coronavirus.jhu.edu/map.html> [accessed 2020-08-06]
3. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html> [accessed 2020-08-06]
4. Cohen J, Kupferschmidt K. Countries test tactics in 'war' against COVID-19. *Science* 2020 Mar 20;367(6484):1287-1288. [doi: [10.1126/science.367.6484.1287](https://doi.org/10.1126/science.367.6484.1287)] [Medline: [32193299](https://pubmed.ncbi.nlm.nih.gov/32193299/)]
5. Nicola M, O'Neill N, Sohrabi C, Khan M, Agha M, Agha R. Evidence based management guideline for the COVID-19 pandemic - Review article. *Int J Surg* 2020 May;77:206-216 [FREE Full text] [doi: [10.1016/j.ijsu.2020.04.001](https://doi.org/10.1016/j.ijsu.2020.04.001)] [Medline: [32289472](https://pubmed.ncbi.nlm.nih.gov/32289472/)]



6. Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Apr 17;69(15):458-464 [FREE Full text] [doi: [10.15585/mmwr.mm6915e3](https://doi.org/10.15585/mmwr.mm6915e3)] [Medline: [32298251](https://pubmed.ncbi.nlm.nih.gov/32298251/)]
7. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *N Engl J Med* 2020 May 21;382(21):2049-2055. [doi: [10.1056/nejmsb2005114](https://doi.org/10.1056/nejmsb2005114)]
8. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 2020 May 22;368(6493):860-868 [FREE Full text] [doi: [10.1126/science.abb5793](https://doi.org/10.1126/science.abb5793)] [Medline: [32291278](https://pubmed.ncbi.nlm.nih.gov/32291278/)]
9. Buehler JW, Centers for Disease Control and Prevention. CDC's vision for public health surveillance in the 21st century. *MMWR Suppl* 2012 Jul 27;61(3):1-2. [Medline: [22832989](https://pubmed.ncbi.nlm.nih.gov/22832989/)]
10. Burrell C, Howard C, Murphy F. Fenner and White's Medical Virology, Fifth Edition. London, UK: Academic Press; Nov 30, 2016.
11. Brownstein JS, Chu S, Marathe A, Marathe MV, Nguyen AT, Paolotti D, et al. Combining Participatory Influenza Surveillance with Modeling and Forecasting: Three Alternative Approaches. *JMIR Public Health Surveill* 2017 Nov 01;3(4):e83 [FREE Full text] [doi: [10.2196/publichealth.7344](https://doi.org/10.2196/publichealth.7344)] [Medline: [29092812](https://pubmed.ncbi.nlm.nih.gov/29092812/)]
12. Lu FS, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, et al. Accurate Influenza Monitoring and Forecasting Using Novel Internet Data Streams: A Case Study in the Boston Metropolis. *JMIR Public Health Surveill* 2018 Jan 09;4(1):e4 [FREE Full text] [doi: [10.2196/publichealth.8950](https://doi.org/10.2196/publichealth.8950)] [Medline: [29317382](https://pubmed.ncbi.nlm.nih.gov/29317382/)]
13. Baltrusaitis K, Brownstein JS, Scarpino SV, Bakota E, Crawley AW, Conidi G, et al. Comparison of crowd-sourced, electronic health records based, and traditional health-care based influenza-tracking systems at multiple spatial resolutions in the United States of America. *BMC Infect Dis* 2018 Aug 15;18(1):403 [FREE Full text] [doi: [10.1186/s12879-018-3322-3](https://doi.org/10.1186/s12879-018-3322-3)] [Medline: [30111305](https://pubmed.ncbi.nlm.nih.gov/30111305/)]
14. Abir M, Cutter C, Nelson C. COVID-19: A Stress Test for A US Health Care System Already Under Stress. *Health Affairs* 2020 Mar 11 [FREE Full text] [doi: [10.1377/hblog20200310.530622](https://doi.org/10.1377/hblog20200310.530622)]
15. Brunton G, Oliver S, Thomas J. Innovations in framework synthesis as a systematic review method. *Res Synth Methods* 2020 May 03;11(3):316-330. [doi: [10.1002/jrsm.1399](https://doi.org/10.1002/jrsm.1399)] [Medline: [32090479](https://pubmed.ncbi.nlm.nih.gov/32090479/)]
16. Koehlmoos T, Gazi R, Rashid M. Social franchising evaluations: A scoping review. LondonPPI-Centre, Social Science Research Unit, Institute of Education, University of London. 2011 Jun. URL: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3085> [accessed 2020-04-14]
17. Booth A, Carroll C. How to build up the actionable knowledge base: the role of 'best fit' framework synthesis for studies of improvement in healthcare. *BMJ Qual Saf* 2015 Nov 25;24(11):700-708. [doi: [10.1136/bmjqs-2014-003642](https://doi.org/10.1136/bmjqs-2014-003642)] [Medline: [26306609](https://pubmed.ncbi.nlm.nih.gov/26306609/)]
18. Carroll C, Booth A, Cooper K. A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents. *BMC Med Res Methodol* 2011 Mar 16;11(1):29 [FREE Full text] [doi: [10.1186/1471-2288-11-29](https://doi.org/10.1186/1471-2288-11-29)] [Medline: [21410933](https://pubmed.ncbi.nlm.nih.gov/21410933/)]
19. Osborne R, Norquist J, Elsworth G, Busija L, Mehta V, Herring T, et al. Development and Validation of the Influenza Intensity and Impact Questionnaire (FluIIQ™). *Value in Health* 2011 Jul;14(5):687-699. [doi: [10.1016/j.jval.2010.12.005](https://doi.org/10.1016/j.jval.2010.12.005)] [Medline: [21839407](https://pubmed.ncbi.nlm.nih.gov/21839407/)]
20. Chunara R, Aman S, Smolinski M, Brownstein J. Flu Near You: An Online Self-Reported Influenza Surveillance System in the USA. *Online J of Public Health Inform* 2013;5(1):e [FREE Full text] [doi: [10.5210/ajph.v5i1.4456](https://doi.org/10.5210/ajph.v5i1.4456)]
21. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020 Jul 11;26(7):1037-1040. [doi: [10.1038/s41591-020-0916-2](https://doi.org/10.1038/s41591-020-0916-2)] [Medline: [32393804](https://pubmed.ncbi.nlm.nih.gov/32393804/)]
22. Lochlainn M, Lee K, Sudre C. Key predictors of attending hospital with COVID19: An association study from the COVID Symptom Tracker App in 2,618,948 individuals. Preprint posted on April 29, 2020. medRxiv [FREE Full text] [doi: [10.1101/2020.04.25.20079251](https://doi.org/10.1101/2020.04.25.20079251)]
23. Drew D, Nguyen L, Steves C, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: [10.1126/science.abc0473](https://doi.org/10.1126/science.abc0473)] [Medline: [32371477](https://pubmed.ncbi.nlm.nih.gov/32371477/)]

## Abbreviations

**CDC:** Centers for Disease Control and Prevention

**ICU:** intensive care unit

*Edited by G Eysenbach; submitted 06.08.20; peer-reviewed by A Cravioto, T Burke; comments to author 01.09.20; revised version received 02.09.20; accepted 14.09.20; published 22.10.20.*

*Please cite as:*

*Koehlmoos TP, Janvrin ML, Korona-Bailey J, Madsen C, Sturdivant R*

*COVID-19 Self-Reported Symptom Tracking Programs in the United States: Framework Synthesis*

*J Med Internet Res 2020;22(10):e23297*

URL: <http://www.jmir.org/2020/10/e23297/>

doi: [10.2196/23297](https://doi.org/10.2196/23297)

PMID: [33006943](https://pubmed.ncbi.nlm.nih.gov/33006943/)

©Tracey Pérez Koehlmoos, Miranda Lynn Janvrin, Jessica Korona-Bailey, Cathaleen Madsen, Rodney Sturdivant. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study

Ranganathan Chandrasekaran<sup>1</sup>, PhD; Vikalp Mehta<sup>1</sup>, MSc; Tejali Valkunde<sup>1</sup>, MSc; Evangelos Moustakas<sup>2</sup>, PhD

<sup>1</sup>Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL, United States

<sup>2</sup>Middlesex University Dubai, Dubai, United Arab Emirates

**Corresponding Author:**

Ranganathan Chandrasekaran, PhD

Department of Information and Decision Sciences

University of Illinois at Chicago

601 S Morgan St

Chicago, IL, 60607

United States

Phone: 1 3129962847

Email: [ranga@uic.edu](mailto:ranga@uic.edu)

## Abstract

**Background:** With restrictions on movement and stay-at-home orders in place due to the COVID-19 pandemic, social media platforms such as Twitter have become an outlet for users to express their concerns, opinions, and feelings about the pandemic. Individuals, health agencies, and governments are using Twitter to communicate about COVID-19.

**Objective:** The aims of this study were to examine key themes and topics of English-language COVID-19–related tweets posted by individuals and to explore the trends and variations in how the COVID-19–related tweets, key topics, and associated sentiments changed over a period of time from before to after the disease was declared a pandemic.

**Methods:** Building on the emergent stream of studies examining COVID-19–related tweets in English, we performed a temporal assessment covering the time period from January 1 to May 9, 2020, and examined variations in tweet topics and sentiment scores to uncover key trends. Combining data from two publicly available COVID-19 tweet data sets with those obtained in our own search, we compiled a data set of 13.9 million English-language COVID-19–related tweets posted by individuals. We use guided latent Dirichlet allocation (LDA) to infer themes and topics underlying the tweets, and we used VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis to compute sentiment scores and examine weekly trends for 17 weeks.

**Results:** Topic modeling yielded 26 topics, which were grouped into 10 broader themes underlying the COVID-19–related tweets. Of the 13,937,906 examined tweets, 2,858,316 (20.51%) were about the impact of COVID-19 on the economy and markets, followed by spread and growth in cases (2,154,065, 15.45%), treatment and recovery (1,831,339, 13.14%), impact on the health care sector (1,588,499, 11.40%), and governments response (1,559,591, 11.19%). Average compound sentiment scores were found to be negative throughout the examined time period for the topics of spread and growth of cases, symptoms, racism, source of the outbreak, and political impact of COVID-19. In contrast, we saw a reversal of sentiments from negative to positive for prevention, impact on the economy and markets, government response, impact on the health care industry, and treatment and recovery.

**Conclusions:** Identification of dominant themes, topics, sentiments, and changing trends in tweets about the COVID-19 pandemic can help governments, health care agencies, and policy makers frame appropriate responses to prevent and control the spread of the pandemic.

(*J Med Internet Res* 2020;22(10):e22624) doi:[10.2196/22624](https://doi.org/10.2196/22624)

**KEYWORDS**

coronavirus; infodemiology; infoveillance; infodemic; twitter; COVID-19; social media; sentiment analysis; trends; topic modeling; disease surveillance

## Introduction

As the effects of the COVID-19 pandemic are felt worldwide, social media platforms are becoming inundated with content associated with the disease. Since its initial identification and reporting in Wuhan, China, the novel disease COVID-19 has spread to multiple countries across all continents and has become a global pandemic. The World Health Organization (WHO) declared the outbreak to be a pandemic on March 11, 2020, and the US government declared it to be a national emergency on March 13, 2020. As of June 30, 2020, the virus has infected over 10 million individuals and has caused approximately 503,000 deaths worldwide [1]. To contain the spread of the virus, several countries have implemented lockdown and quarantine measures and imposed travel bans, restricting people's movement. Schools have been closed, many workers have become unemployed, and numerous individuals are locked down in their homes. With millions of lives affected by the COVID-19 pandemic, social media platforms such as Twitter have become an outlet for users to express their concerns, opinions, and feelings about the pandemic.

Social media has emerged as a significant conduit for health-related information; the majority of people across multiple countries use some form of social media [1,2]. Pew Research surveys examining multiple countries have identified social media as an important source of health information [3]. In recent years, sharing and consuming health information via social media has become prevalent. It is unsurprising that social media has become a prominent platform for people to share information and feelings about COVID-19.

The science of understanding health-related information that is distributed via a digital medium such as the internet or social media with the aim to inform public health and public policy is known as infodemiology. A related term, infoveillance, refers to syndromic surveillance of public health-related concerns that is expressed and diffused on the internet through digital channels. Infoveillance has been particularly useful to identify

outbreak patterns and to study public perceptions of several diseases, including H1N1 influenza ("swine flu") [4], Ebola virus [5,6], and Zika virus [7-9]. Analysis of health event data posted on social media platforms not only provides firsthand evidence of health event occurrences but also enables faster access to real-time information that can help health professionals and policy makers frame appropriate responses to health-related events.

The COVID-19 outbreak has propelled an emergent set of studies that have examined public perceptions, thoughts, and concerns about this pandemic using social media data (Table 1). Most of these studies relied on data from the Twitter or Weibo platforms and analyzed data from early periods of the pandemic. The amount of data used in these studies varies from a few hundred tweets to a few million. These studies have collectively provided a rich body of knowledge on how Twitter users have reacted to the pandemic and their concerns in the early stages of the outbreak. Many of these studies did not differentiate between sources of tweets, such as whether the tweet originated from an individual or an organization such as a news channel or health agency. From an infoveillance perspective, it is important to understand the social media discourses pertaining to COVID-19 among the common public rather than by news agencies or other organizations. Further, there is limited understanding of the changes in public sentiments and discourse about COVID-19 over time. To address these gaps, we examined COVID-19-related tweets using a much larger data set covering a time period from January 1 to May 9, 2020. We performed a temporal assessment and examined variations in the topics and sentiment scores over a period of time from before to after the disease was declared a pandemic to uncover key trends.

Our research goals were to examine key themes and topics in COVID-19-related English-language tweets posted by individuals and to explore the trends and variations in how COVID-19-related tweets, key topics, and associated sentiments changed over a period of time from before to after the disease was declared a pandemic.

**Table 1.** Summary of key studies on the COVID-19 pandemic using social media data.

Source	Social media platform	Data set	Time period	Key findings
Abd-Alrazaq et al, 2020 [10]	Twitter	167,073 tweets	Tweets from February 2 to March 15, 2020	Identified 12 topics that were grouped into four themes, viz the origin of the virus; its sources; its impact on people, countries, and the economy; and ways of mitigating infection.
Li et al, 2020 [11]	Weibo	115,299 posts	Posts from December 23, 2019, to January 30, 2020	Positive correlation between the number of Weibo posts and number of reported cases in Wuhan. Qualitative analysis of 11,893 posts revealed main themes of disease causes, changing epidemiological characteristics, and public reaction to outbreak control and response measures.
Shen et al, 2020 [12]	Weibo	15 million posts	Posts from November 1, 2019, to March 31, 2020	Developed a classifier to identify “sick posts” pertaining to COVID-19. The number of sick posts positively predicted the officially reported COVID-19 cases up to 14 days ahead of official statistics.
Sarker et al, 2020 [13]	Twitter	499,601 tweets from 305 users who self-disclosed their COVID-19 test results	N/A <sup>a</sup>	203 users who tested positive for COVID-19 reported their symptoms: fever/pyrexia, cough, body ache/pain, fatigue, headache, dyspnea, anosmia and ageusia.
Tao et al, 2020 [14]	Weibo	15,900 posts	December 31, 2019, to March 16, 2020	Analysis of oral health–related information posted on Weibo revealed home oral care and dental services to be the most common tweet topics.
Wahbeh et al, 2020 [15]	Twitter	10,096 tweets from 119 medical professionals	December 1, 2019, to April 1, 2020	Identified eight themes: actions and recommendations, fighting misinformation, information and knowledge, the health care system, symptoms and illness, immunity, testing, and infection and transmission.
Budhwani et al, 2020 [16]	Twitter	193,862 tweets by US-based users	March 9 to March 25, 2020	Identified a large increase in the number of tweets referencing “Chinese virus” or “China virus.”
Rufai and Bunce, 2020 [17]	Twitter	203 viral tweets by 8 G7 <sup>b</sup> world leaders	November 17, 2019, to March 17, 2020	Identified three categories of themes: informative, morale-boosting, and political.
Park et al, 2020 [18]	Twitter	43,832 users and 78,233 relationships	Few weeks before February 29, 2020	Assessed speed of information transmission in networks and found that news containing the word “coronavirus” spread faster.
Lwin et al, 2020 [19]	Twitter	20,325,929 tweets from 7,033,158 users	January 28 to April 9, 2020	An examination of four emotions (fear, anger, sadness, and joy) revealed that emotions shifted from fear to anger, while sadness and joy also surfaced.
Pobiruchin et al, 2020 [20]	Twitter	21,755,802 tweets from 4,809,842 users	February 9 to April 11, 2020	Examined temporal and geographical variations of COVID-19–related tweets, focusing on Europe, and the categories and origins of shared external resources.

<sup>a</sup>N/A: not applicable.<sup>b</sup>G7: Group of Seven.

## Methods

### Data Collection

We collected all COVID-19–related tweets from January 1 to May 9, 2020. The Python programming language was used for our data collection and analyses, and Tableau was used as a supplementary tool for visualization purposes. We used three

sources to assemble the tweets required for our analysis. First, we relied on the COVID-19 Twitter data set at IEEE Dataport [21], which contained COVID-19–related tweets from March 20, 2020. Second, we used a Twitter data set posted in GitHub [22] that contained COVID-19–related tweets posted since January 21, 2020. Both these data sets are publicly available and provide a list of tweet IDs for all tweets related to COVID-19. Third, we collected COVID-19–related tweets for



the remaining period, including texts and metadata, from Twitter using GetOldTweets3, a Python 3 library that enables scraping of historical Twitter data [23]. Because we were combining tweets from multiple sources, we used a common set of keywords and phrases that other sources had used: *corona*, *coronavirus*, *covid-19*, *covid19*, and their variants, including their hashtag equivalents. The language-tag setting “EN” and the retweet tag “RT” were used to filter English-language tweets and retweets. We also used the retweets feature in GetOldTweets3 to filter out retweets. Due to restrictions of the Twitter platform, the public data sets contained only the tweet IDs. The process of extracting complete details of a tweet, including metadata, from Twitter using the tweet ID is referred to as hydration, and a number of tools have been developed for this purpose [11]. We used the Hydrator software [24] listed in IEEE Dataport to gather the complete text and metadata of the tweets.

### Data Preprocessing

Our next step was to classify all the tweets posted by individuals versus those that originated from organizations. We first gathered the unique Twitter user IDs of all the Twitter users in our data set. Following the approach outlined in [25], we used a naïve Bayes machine learning model to classify the tweeters into individuals versus organizations. We used a published data set that contained 8945 Twitter users and their profile descriptions, which human coders used to annotate users as individual or institutional [26]. To these data, we added 2000 Twitter user IDs pulled from our data set along with their associated profiles, and we manually annotated them (interrater reliability  $\kappa=0.84$ ). Using the combined data set of 10,945 users, we divided our data set into training versus validation sets using an 80:20 split, and we used these sets to train and test our classifier model, respectively. The naïve Bayes classifier yielded an accuracy of 83.2% with a precision of 0.82, a recall of 0.83, and an F1 score of 0.81; these values were considered satisfactory and are comparable to those in other studies [27–29]. [Multimedia Appendix 1](#) presents the confusion matrix. Our classifier performance was also robust across multiple split strategies for dividing the data set for training and validation. This classifier was then used to identify all the individual users in our full data set, and only tweets posted by individuals were retained for further assessment. We also eliminated duplicate tweets and retweets (filtered using the “RT” tag), resulting in a data set that contained only original tweets posted by individual users. We preprocessed and cleaned the tweets using the Natural Language Toolkit (NLTK), regular expression (RegEx), and the gensim Python library [30]. We removed stop words, user mentions, and links, and we also lemmatized the text of the tweets.

### Topic Modeling and Sentiment Analysis

Topic modeling is an unsupervised machine learning approach that is useful for discovering abstract topics that occur in a collection of textual documents. It helps uncover hidden semantic structures in a body of documents. Most topic modeling algorithms are based on probabilistic generative models that specify mechanisms for how documents are written in order to infer abstract topics. One popular topic modeling

algorithm, latent Dirichlet allocation (LDA), is an unsupervised generative probabilistic method for modeling a corpus of words [31]. A key advantage of LDA is that no prior knowledge of topics is needed. By tuning the LDA parameters, one can explore the formation of different topics and the resultant document clusters. Despite the usefulness of LDA, its outcomes can be difficult to interpret and can drastically vary based on the choice of parameters. With a large corpus of texts, the unsupervised nature of LDA can result in the generation of topics that are neither meaningful nor effective, requiring human intervention and multiple iterations [10]. An improvised variant to traditional LDA, the guided LDA algorithm [32], enables the provision of a set of seed words that are representative of the underlying topics so that the topic models are guided to learn topics that are of specific interest.

We used two broad approaches to prepare the initial set of topics and the seed words for guided LDA. First, we used the extant literature on COVID-19 infoveillance using Twitter to identify a broad set of topics and potential keywords. Second, we performed traditional LDA with multiple numbers of topics as inputs ( $n=10, 20, 30$ , and  $40$ ) iteratively and examined the word lists that were generated. We used both steps to generate a list of topics and anchor words for the guided LDA (see [Multimedia Appendix 2](#)). The GuidedLDA package in Python was used for the topic modeling. Through discussions, the authors then grouped the topics and identified dominant themes. Further, we computed a sentiment score for each tweet using the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool in Python. VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments in social media texts such as tweets [33].

To assess the sentiments of tweets, VADER provides a compound score metric that calculates the sum of all Lexicon ratings that have been normalized between  $-1$  (most extreme negative) and  $+1$  (most extreme positive); this method takes into account both the polarity (positive/negative) and the intensity of the emotion expressed. For each tweet, we classified the sentiment as positive, negative, or neutral based on the compound score. A tweet with a compound score greater than  $0.05$  was classified as positive, a tweet with a score between  $-0.05$  and  $0.05$  was classified as neutral, and a tweet with a score less than  $-0.05$  was classified as negative. To further understand the changes in the sentiment scores over time, we qualitatively analyzed the content of tweets to explore the rationale behind the changes in the compound sentiment scores. The authors manually examined the tweets pertaining to specific topics in weeks in which variations were observed to infer possible reasons for the variations in sentiment.

## Results

We obtained a total of 13,937,906 tweets from 10,868,921 unique users after eliminating 4,085,264 tweets posted by organizations and institutions. Our primary goal was to understand public perceptions and sentiments pertaining to COVID-19; hence, only tweets posted by individuals were retained for analysis.

## Themes and Topics From Text Mining

Our analysis of tweets yielded 26 subtopics, which we framed into 10 broad themes (Table 2). Of the 13,937,906 tweets we examined, 2,858,316 (20.51%) pertained to the theme of the impact of COVID-19 on the economy and markets, followed by spread and growth in cases (2,154,065, 15.45%), treatment and recovery (1,831,339, 13.14%), impact on the health care

sector (1,588,499, 11.40%), and government response to the pandemic (1,559,591, 11.19%). Although tweets related to the theme of racism formed only 4.14% (577,066/13,937,906) of the data, over 500,000 tweets were found to contain racist content. It should be noted that all the tweets we assessed were public discourses pertaining to broader themes, as our data set consisted of tweets posted by individuals about various issues pertaining to the COVID-19 pandemic.

**Table 2.** Themes and topics of COVID-19–related tweets (N=13,937,906), n (%).

Theme and topics	Value
<b>1. Source (origin)</b>	<b>966,372 (6.93)</b>
1.1 Outbreak	489,768 (3.51)
1.2 Alternative causes	476,606 (3.42)
<b>2. Prevention</b>	<b>1,076,840 (7.73)</b>
2.1 Social distancing	575,786 (4.13)
2.2 Disinfecting and cleanliness	501,054 (3.59)
3. Symptoms	558,332 (4.01)
<b>4. Spread and growth</b>	<b>2,154,065 (15.45)</b>
4.1 Modes of transmission	472,749 (3.39)
4.2 Spread of cases	617,946 (4.43)
4.3 Hotspots and locations	459,039 (3.29)
4.4 Death reports	604,331 (4.34)
<b>5. Treatment and recovery</b>	<b>1,831,339 (13.14)</b>
5.1 Drugs and vaccines	442,413 (3.17)
5.2 Therapies	483,109 (3.47)
5.3 Alternative methods	416,530 (2.99)
5.4 Testing	489,287 (3.51)
<b>6. Impact on the economy and markets</b>	<b>2,858,316 (20.51)</b>
6.1 Shortage of products	513,703 (3.69)
6.2 Panic buying	667,320 (4.79)
6.3 Stock markets	535,262 (3.84)
6.4 Employment	505,510 (3.63)
6.5 Impact on business	636,521 (4.57)
<b>7. Impact on health care sector</b>	<b>1,588,499 (11.4)</b>
7.1 Impact on hospitals and clinics	441,895 (3.17)
7.2 Policy changes	615,027 (4.41)
7.3 Frontline workers	531,577 (3.81)
<b>8. Government response</b>	<b>1,559,591 (11.19)</b>
8.1 Travel restrictions	519,406 (3.73)
8.2 Financial measures	485,277 (3.48)
8.3 Lockdown regulations	554,908 (3.98)
9. Political impact	767,486 (5.51)
10. Racism	577,066 (4.14)

## Trends in the Proportions of Positive, Negative, and Neutral COVID-19 Tweets

For each theme pertaining to COVID-19, we examined the trends in the proportions of positive, negative, and neutral tweets over time (Figure S1 in [Multimedia Appendix 3](#)). Of the total tweets concerning the source of the COVID-19 outbreak, the proportions of neutral and negative tweets remained fairly high (approximately 35% to 45%) in the weeks before the WHO announced that COVID-19 was a pandemic. The proportion of positive tweets exceeded those of negative and neutral tweets in the week of the WHO declaration. In the subsequent weeks, the proportion of positive tweets dropped to approximately 25%, whereas the proportions of neutral and negative tweets were approximately 30% to 45%. When we examined tweets pertaining to the prevention of COVID-19, the proportion of positive tweets exceeded those of neutral and negative tweets in almost all the weeks from February 2020, reaching approximately 40% in the beginning of May 2020.

The proportion of negative tweets was considerably higher than those of the positive and neutral tweets for the themes of symptoms (approximately 60%) and of spread and growth in cases (approximately 45%). This pattern was observed for almost all the weeks we examined. In February 2020, over 90% of tweets on the theme of symptoms were negative. Although this trend gradually declined over the next few weeks, it still formed over 50% in the last week of our examination. Similarly, negative tweets about the spread and increase in COVID-19 cases constituted between 40% and 50% from February 2020 until the beginning of May 2020. For the theme of treatment and recovery, the proportion of positive tweets (20%) gradually increased to over 40% over the 17-week period. The negative tweets in the initial weeks (30% to 35%) declined to 25% in April and early May 2020.

We noted a gradual increase in the proportion of positive tweets pertaining to the impact of COVID-19 on the economy and markets over time. Proportions of negative tweets were higher in the months of February and March 2020 but gradually declined to approximately 30% toward the beginning of May 2020.

An increase in the proportion of positive tweets over time was seen for the themes of government response and impact on the health care industry. The theme pertaining to government response captured the Twitter discourse by users concerning various measures taken by different governments to address COVID-19. The proportion of negative tweets about government response was approximately 45% up to mid-March 2020 and then declined to approximately 30% by the first week of May. The proportion of negative tweets on the theme of the political impacts of COVID-19 was considerably higher (>50%) from March 2020. We also noted a substantial proportion of negative tweets on the theme of racism.

## Trends in Sentiments of Themes of COVID-19 Tweets

We examined the trends pertaining to the changes in the sentiment scores of each of our themes and topics over the time period of examination. To plot the trends, we used the average

compound scores by topic and week. Our results are presented in Figure S2 ([Multimedia Appendix 3](#)).

Average compound sentiment scores were found to be negative throughout the time period of our examination for the themes of spread and growth of cases, symptoms, racism, source of the outbreak, and political impacts of COVID-19. In contrast, we saw a reversal of sentiments from negative to positive for the themes of prevention, impact on the economy and markets, government response, impact on the health care industry, and treatment and recovery; the negative sentiment scores in the initial weeks of the COVID-19 outbreak for the aforementioned themes changed to positive scores in the final few weeks of our examination. This reversal of sentiments is noteworthy, as it reflects a collective opinion of a fairly larger set of Twitter users on how the pandemic is being managed by key stakeholders.

## Trends in Sentiments of Topics of COVID-19 Tweets

We further examined the trends in the sentiment scores for topics underlying the broader themes. Compound scores from VADER were averaged over each topic for every week (Figure S3, [Multimedia Appendix 3](#)). This assessment helped us to understand the progression of sentiments for specific topics over the period we examined. To understand the variations in the sentiments, we also qualitatively examined the tweets for weeks in which changes were observed. Sample tweets for each of the themes and topics are shown in [Multimedia Appendix 4](#).

Our analysis revealed a consistently negative average compound score for the topic of the outbreak in Wuhan, China, for all the weeks that we examined. We found that Twitter users frequently referred to the geographical origin of the disease even in the later weeks of our examination. When we examined the topic of alternative causes of the outbreak, we found several tweets about hypothetical causes and conspiracy theories pertaining to COVID-19 (eg, use of SARS-CoV-2 as a bioweapon and origin of the virus in a lab in Wuhan). The average sentiment scores remained negative for weeks until the week of March 22 to 28, then showed a spike to positive values and continued to remain positive until the beginning of May 2020. This positive trend in later weeks is due to tweets dismissing the conspiracy theories that were circulated during the early weeks. Further, the spike in the positive score in the week of March 22 to 28 is partly due to a large number of tweets that contained references to “coronavirus as an act of God” and prayers to end the pandemic, as well as tweets that viewed COVID-19 as “nature’s way to heal the planet.” These types of tweets provide qualitative evidence for the positive sentiment scores we observed in the analysis, such as:

*This virus is certainly God’s call to humanity to wake up and recognise him before it is too late.*

*Wow. Earth is recovering, Air pollution is slowing down, Water pollution is clearing up, Natural wildlife is returning home, Coronavirus is earth’s vaccine. We’re the virus.*

*This planet will surely heal, in the most magical ways. I can feel the vibrations coming on.*

The average compound score for social distancing remained negative until the first week of March. During this time period,

COVID-19 had not yet spread worldwide. However, from the second week of March 2020, the average sentiment score was positive for all the weeks we examined, reflecting that the general public supported and had a favorable disposition toward social distancing as a mechanism to combat the spread of the virus. We observed that several Twitter appealed to others and advocated social distancing measures, such as:

*Kindly stay at home. Wash your hands. Practice social distancing.*

*Ran two miles even when I didn't want to! Made excuses all day! Get out there and do it! But practice social distancing, let's flatten this curve!*

The topic of disinfecting and cleanliness showed average positive sentiment scores for all weeks from the third week of January 2020. We found that Twitter users used gaming strategies such as challenges (eg, #SafeHandsChallenge) involving a chain of users to advocate cleanliness and create broader awareness about the importance of disinfecting and cleanliness. We found that many Twitter users shared tips about disinfecting groceries and products after shopping. We also found that some Twitter users condemned people who did not wear face masks or follow recommended safety protocols in public.

Three of the four topics under the theme of spread and growth exhibited negative average compound sentiment scores. The average compound scores for the topic of death reports were negative for all the weeks, with values ranging between  $-0.2$  and  $-0.5$ . The topic pertaining to spread of cases exhibited negative trends throughout, with average compound scores ranging between  $-0.1$  and  $-0.3$ . The topic of modes of transmission of COVID-19 also showed negative scores across all the weeks, with values between  $0$  and  $-0.2$ . The topic of hotspots and locations for COVID-19 transmission exhibited negative scores until February 2020 but showed positive scores thereafter. Tweets mentioning hotspots of COVID-19 transmission often included mentions of places such as churches, places of religious worship, beaches, events, and festive occasions with mass gatherings; these mentions primarily contributed to the positive values.

All four topics under the theme of treatment and recovery showed negative scores in the initial weeks, which changed to positive average compound scores from April 2020. For the topic of testing, Twitter users reacted negatively to the lack of availability of test kits and testing methods in the initial weeks of the pandemic (eg, tweets containing phrases such as “not all in the hospitals can be tested as they are often short with test kits” or “Many places are not testing people for coronavirus due to test shortage. Its annoying”); however, with the improvement in availability of COVID-19 test kits and test centers worldwide, the sentiment became positive:

*We got tested today. Easy as could be, no waiting, felt really safe, cheek swabs.*

*In lots of states, it's very easy to get tested now, even if you're asymptomatic.*

As more information on the efficacies of drugs such as remdesivir became available, Twitter users' sentiments regarding

drugs and medicines for COVID-19 became positive by the end of March 2020. Twitter users reacted to news about drugs, as can be seen in these example tweets:

*HCQ and Remdesivir, are effective at limiting duration of illness, hospitalization and viral spread if given early.*

*Remdesivir is effective in mitigating COVID-19 symptoms if taken early, ideally pre-hospitalization.*

We also noted a small increase in the average positive compound score for the topic of drugs and medicines in the last week of April; this can be attributed to the US Food and Drug Administration's authorization of emergency use of the antimalarial drug chloroquine to treat COVID-19 on April 27, 2020. Tweets such as these contributed to this positive compound score:

*Hydroxychloroquine protocol: effective, cheap and can be produced in many laboratories. HCQ functions as both a cure and a vaccine.*

However, it should be noted that this authorization was later revoked on June 15, 2020, a date outside the time period covered by our study. The negative sentiments about various therapies in the initial weeks of the pandemic also started to become positive in the third week of March. Twitter users positively reacted and shared information on plasma therapy and associated trials that were being conducted on patients with COVID-19. For instance, some tweets provided examples of positive sentiments from Twitter users:

*We desperately need a treatment for those severely suffering with Coronavirus. Blood plasma could be the answer.*

*The effects of coronavirus are scary for many families, but this treatment of using antibodies from recovered patients could save lives.*

The topic of alternative methods of treatment for COVID-19 (traditional Chinese medicine, Indian Ayurveda, etc.) had mildly positive sentiment scores for most of the time period in our examination.

Among the topics that comprise the theme of impact on the economy and markets, Twitter users' average compound scores for sentiments about the topic of employment remained negative for all the weeks we examined. Many Twitter users posted information about their job loss and unemployment:

*Lost my job a couple weeks ago due to Coronavirus and now it's impossible to find a new job.*

*Well....I just got the call. Lost my job due to Covid.*

Moreover, other users organized crowdfunding campaigns to help people who lost their jobs. Similar negative scores were seen across all the weeks for the topic pertaining to stock markets. Tweets pertaining to the topic of panic buying by consumers showed negative sentiment scores for all the weeks until mid-April, after which the scores began to be positive. Many users shared tweets about long queues and panic buying, as can be seen in these tweets:



*2020 and panic buying has reduced us to this. Waiting in line for at least 2 hours to get pull ups and baby wipes because no one else has them.*

*No panic buying y'all hear that? So leave some damn bread and milk for me please.*

Tweets about the topic of shortages (of food and essential items) swung between positive and negative scores in the first few weeks of the pandemic but became positive from mid-March until early May. Twitter users reacted positively to measures adopted by supermarkets and grocery stores to practice safety measures, as can be seen in this illustrative tweet:

*Yes! Longos Markets requires all customers to wear masks. Went there today, it was a good, safe shopping experience, better than any other store. Will definitely be shopping there again.*

Twitter users' sentiments about the topic of businesses exhibited negative scores in the initial weeks of the pandemic, primarily fueled by news about closures and losses. However, this sentiment score changed to positive in the first week of March 2020. The positive score seen here reflects the adaptation of businesses to the new pandemic and their reopening and revival across different countries. Tweets such as the following provide indicators of positive sentiments of users about reopening of businesses:

*Most of HK open for business now. Emphasis on testing and tracing.*

*Open for business. Trusting the people to take care of themselves. Freedom smells sweet.*

Sentiment scores pertaining to the topic of hospitals and clinics largely remained negative throughout the time period of our examination. Tweets about lack of beds, facilities, NS ventilators, overcrowding of patients, and the struggles of health care institutions to cater to the influx of COVID-19 patients contributed to the negative sentiment scores. Illustrative tweets about these negative sentiments include:

*Madrid hospitals now have double the number of intensive care patients than beds. Means you can no longer get intensive care in a Madrid hospital.*

*Lack of safety gear for healthcare workers, shortage of beds and doctors, inadequate labs to conduct tests - our healthcare system is very fragile!!*

However, we noted a reversal in the trends of sentiment scores for the topic of frontline health care workers. Twitter users' negative sentiments until the end of March 2020, reflecting the lack of personal protective equipment and gear for health care professionals, health worker burnout, and increased workload for health care workers, became positive in April and early May 2020 as the situation improved. We saw tweets in the initial period of the pandemic with negative sentiments, such as "the knighted geniuses at the top of the NHS can't even organise protective equipment for our doctors and nurses". In the later weeks of our examination, hashtags such as #coronawarriors and tweets hailing the services of frontline workers (eg, "Deepest gratitude to the #CoronaWarriors who are working tirelessly in these difficult times") contributed to positive sentiment scores. The topic of health policy, which reflects the

newer safety guidelines, protocols, and policies pertaining to patients with COVID-19 implemented by health care organizations, exhibited a negative trend in the early period, with tweets such as:

*The ventilator situation is even more dire than we know. Not every hospital had an allocation policy in place.*

*Spain has begun a no ventilator policy for anyone over 65.*

However, this topic showed a positive trend after the end of March 2020 as many agencies, governments, and health care institutions began to establish clear policies for treating patients with COVID-19. In the later weeks of our examination, many hospitals had framed clearer guidelines for use of masks, visitations, and restrictions pertaining to COVID-19. These illustrative tweets point to a possible rationale behind the positive sentiments pertaining to the topic of health policy in the later weeks of our examination:

*The hospital has an understandable policy during this crisis of limiting visitation for the safety of all & to reduce use of critical PPEs.*

*Spouse can't visit under hospital's no visitation policy. Psychologically excruciating but family all recognize it's the right thing, and hard to limit exceptions once you start"*

Twitter users' sentiments about the topic of travel restrictions imposed by governments worldwide were largely negative for most of the weeks we examined, except for the week of March 22 to 28, 2020. In this week, governments in populous countries such as India and Canada announced travel restrictions such as flight suspensions and isolation and quarantine measures for individuals entering these countries. Tweets indicated positive sentiments about travel restrictions:

*I believe it was a good move from India to have a complete travel restriction to all countries. When we don't have the health systems to treat huge populations, the best thing to do is to shut doors.*

*The fine for breaking self-quarantine / self-isolation in BC, Canada is \$25,000 AND jail time. Canada is taking travel and quarantine very seriously. Great job.*

Many Twitter users welcomed travel bans and restrictions and expressed positive sentiments about them.

Except for the first two weeks of April, the average compound sentiment scores about the topic of lockdown regulations remained negative in most of the weeks we examined. Twitter users' sentiments about stay-at-home orders, shutdowns, and lockdowns of complete cities were negative. This can be seen from these sample tweets:

*This lockdown really does do bad things to good peoples mental health, trying to stay positive is a task in itself when this shite feels never ending.*

*Lockdown extended for another 3 weeks I hate it here.*

However, the sentiment scores were positive in the weeks in which different governments announced financial measures



such as stimulus payments to people affected by closures and lockdowns. In some tweets, users expressed positive sentiments about the financial relief measures to help individuals suffering due to the impact of COVID-19:

*I got my stimulus check today! Woohoo!*

*Zoe and I finally received our stimulus/relief check from the federal government.*

*India is preparing a stimulus package that would put money directly into the accounts of more than 100 million poor people and support businesses hit the hardest by the 21-day lockdown.*

## Discussion

### Principal Findings

This study joins the growing body of inveillance studies on COVID-19 that examine social media data to uncover public opinions about the pandemic. We used a corpus of over 13 million tweets from January until the first week of May 2020 to uncover the trends in sentiments regarding various themes and topics. Our study is comprehensive, covering 26 different topics underlying COVID-19–related tweets under 10 broader themes. In response to a call made by Liu et al [34], we combined the topic modeling approach with sentiment analysis to observe the trends in sentiments for various themes and topics over time. By assimilating the collective opinions of several million users, we found interesting patterns in the trends pertaining to sentiments of themes and topics of COVID-19–related tweets.

We combined two publicly available sources with our own search to assemble a unique data set that contained English-language tweets about various topics associated with COVID-19. We further used a naïve Bayes classifier to segregate tweets made by individuals. We employed guided LDA to identify the underlying topics and associated themes, and we also examined the sentiments associated with the tweets and their changes over time.

Our key finding is that the impact of COVID-19 on the economy and markets was the most discussed issue by Twitter users. The number and proportions of tweets on this theme were remarkably higher than those of tweets on the other themes we uncovered. Further, users' sentiment was negative until the third week of March but gradually became positive in the final weeks we studied. Users' initial negative sentiments about shortages, panic buying, and businesses gradually turned positive from April 2020. Users started feeling positive about government responses to contain the pandemic, including financial measures to support and assist them in dealing with the disease outbreak.

Twitter users felt negative about continued spread and growth of the number of cases and the symptoms of COVID-19. However, we also found that Twitter users were more positive about treatment of and recovery from COVID-19 in later weeks than they were during the earlier stages of the pandemic. They expressed positive feelings by sharing information on testing, drugs, and newer therapies that show promise to contain the outbreak. Another notable finding is the Twitter users' gradual change in sentiment from negative to positive regarding

COVID-19 prevention measures such as social distancing and cleanliness. Twitter users who initially expressed negative sentiments regarding COVID-19's impacts on the health care sector, comprising hospitals, clinics, and frontline workers, gradually became positive in the later weeks.

Another key finding of our study is the continued negative sentiments about political fallouts due to COVID-19. Although leaders worldwide are struggling to contain the pandemic, we noted that Twitter users felt negative about how COVID-19 was used for political purposes. Similarly, we noted strong negative sentiments about racist content in users' tweets.

Our study offers several insights for health policy makers, administrators, and officials who are managing the impact of the COVID-19 pandemic. Identification of topics and associated sentiment changes provide pointers to how the general public are reacting to specific measures taken to tackle the pandemic. Variations in sentiment scores serve as a feedback mechanism for assessing public perceptions of various measures taken with respect to social distancing, cleanliness and disinfecting, lockdowns, travel restrictions, and efforts to revive the economy. Public sentiments have also started to become positive about COVID-19 testing, treatment, and vaccines as well as health policies. Our study shows that observing aggregate sentiments and changes in trends via social media posts can offer a cost-effective, timely, and valuable mechanism to gauge public perceptions regarding policy decisions being made to address the pandemic.

### Limitations

This study has a few limitations that should be kept in mind while interpreting the results. We relied on a large data set that was partly compiled by us and included two other publicly available data sets. These data sources contained tweets from varying dates and used different search terms and search strategies to gather the tweets. Our analysis may have inadvertently omitted certain COVID-19–related tweets that were not captured by the data sources. In addition, COVID-19–related tweets from users who chose to make their accounts private were not included in our study. Further, we did not consider any geographical boundaries when examining the tweets. Studies focusing on tweets from specific countries can find different topics and sentiments that reflect country-specific opinions and concerns. We also restricted our study to tweets in the English language and to those posted by individuals. It should be noted that our naïve Bayes classifier with over 80% accuracy helped us identify tweets posted by individuals. It is possible that some individual users posted tweets on behalf of organizations, and these tweets may have been included in our data set. A more refined approach with deep learning techniques to identify individual tweets can aid in classifying and assembling a tweet data set with increased accuracy. As a future research extension, tweets posted by organizations could be another frame of reference to understand their concerns and sentiments. Another important limitation is that our findings are reflective of Twitter users, who are fairly familiar with social media and use of technology. The results may not generalize to the larger worldwide population of people who do not use Twitter.

## Conclusion

As COVID-19 continues to affect millions of people worldwide, our study throws light on dominant themes, topics, sentiments and changing trends regarding this pandemic among Twitter users. By examining the changing sentiments and trends surrounding various themes and topics, government agencies,

health care organizations, businesses, and leaders who are working to address the COVID-19 pandemic can be informed about the larger public opinion regarding the disease and the measures they have taken so that adaptations and corrective courses of action can be applied to prevent and control the spread of COVID-19.

## Acknowledgments

The authors would like to thank Navya Shiva, Karansinh Raj, Paul Davis and Ronald Omega Pukadiyil for their assistance in the data collection for the study

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

The confusion matrix.

[DOCX File, 12 KB - [jmir\\_v22i10e22624\\_app1.docx](#)]

### Multimedia Appendix 2

Themes, topics, and associated keywords.

[DOCX File, 15 KB - [jmir\\_v22i10e22624\\_app2.docx](#)]

### Multimedia Appendix 3

Supplementary figures showing the trends in the proportions of positive, neutral, and negative tweets, sentiment score trends by theme, and trends in sentiment scores by topic.

[DOCX File, 1502 KB - [jmir\\_v22i10e22624\\_app3.docx](#)]

### Multimedia Appendix 4

Illustrative tweets for the topics and themes.

[DOCX File, 42 KB - [jmir\\_v22i10e22624\\_app4.docx](#)]

## References

1. Coronavirus disease (COVID-19) Situation Report – 162. World Health Organization. 2020 Jun 30. URL: <https://www.who.int/docs/default-source/coronaviruse/20200630-covid-19-sitrep-162.pdf> [accessed 2020-07-05]
2. Shannon S, Kent N. 8 charts on internet use around the world as countries grapple with COVID-19 Internet. Pew Research Center. 2020 Apr 02. URL: <https://www.pewresearch.org/fact-tank/2020/04/02/8-charts-on-internet-use-around-the-world-as-countries-grapple-with-covid-19/> [accessed 2020-10-20]
3. Silver L, Huang C, Taylor K. In Emerging Economies, Smartphone and Social Media Users Have Broader Social Networks. Pew Research Center. 2019 Aug 22. URL: [https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/08/PI-PG\\_2019-08-22\\_social-networks-emerging-economies\\_FINAL.pdf](https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/08/PI-PG_2019-08-22_social-networks-emerging-economies_FINAL.pdf) [accessed 2020-10-20]
4. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: [10.1371/journal.pone.0014118](https://doi.org/10.1371/journal.pone.0014118)] [Medline: [21124761](#)]
5. Odum M, Yoon S. What can we learn about the Ebola outbreak from tweets? Am J Infect Control 2015 Jun;43(6):563-571 [FREE Full text] [doi: [10.1016/j.ajic.2015.02.023](https://doi.org/10.1016/j.ajic.2015.02.023)] [Medline: [26042846](#)]
6. van Lent LG, Sungur H, Kunneman FA, van de Velde B, Das E. Too Far to Care? Measuring Public Attention and Fear for Ebola Using Twitter. J Med Internet Res 2017 Jun 13;19(6):e193 [FREE Full text] [doi: [10.2196/jmir.7219](https://doi.org/10.2196/jmir.7219)] [Medline: [28611015](#)]
7. Stefanidis A, Vraga E, Lamprianidis G, Radzikowski J, Delamater PL, Jacobsen KH, et al. Zika in Twitter: Temporal Variations of Locations, Actors, and Concepts. JMIR Public Health Surveill 2017 Apr 20;3(2):e22 [FREE Full text] [doi: [10.2196/publichealth.6925](https://doi.org/10.2196/publichealth.6925)] [Medline: [28428164](#)]
8. Daughton AR, Paul MJ. Identifying Protective Health Behaviors on Twitter: Observational Study of Travel Advisories and Zika Virus. J Med Internet Res 2019 May 13;21(5):e13090 [FREE Full text] [doi: [10.2196/13090](https://doi.org/10.2196/13090)] [Medline: [31094347](#)]
9. Pruss D, Fujinuma Y, Daughton AR, Paul MJ, Arnot B, Albers Szafir D, et al. Zika discourse in the Americas: A multilingual topic analysis of Twitter. PLoS One 2019;14(5):e0216922 [FREE Full text] [doi: [10.1371/journal.pone.0216922](https://doi.org/10.1371/journal.pone.0216922)] [Medline: [31120935](#)]

10. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *J Med Internet Res* 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: [10.2196/19016](https://doi.org/10.2196/19016)] [Medline: [32287039](https://pubmed.ncbi.nlm.nih.gov/32287039/)]
11. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. *JMIR Public Health Surveill* 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: [10.2196/18700](https://doi.org/10.2196/18700)] [Medline: [32293582](https://pubmed.ncbi.nlm.nih.gov/32293582/)]
12. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study. *J Med Internet Res* 2020 May 28;22(5):e19421 [FREE Full text] [doi: [10.2196/19421](https://doi.org/10.2196/19421)] [Medline: [32452804](https://pubmed.ncbi.nlm.nih.gov/32452804/)]
13. Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi M, Yang YC. Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource. *J Am Med Inform Assoc* 2020 Aug 01;27(8):1310-1315 [FREE Full text] [doi: [10.1093/jamia/ocaa116](https://doi.org/10.1093/jamia/ocaa116)] [Medline: [32620975](https://pubmed.ncbi.nlm.nih.gov/32620975/)]
14. Tao Z, Chu G, McGrath C, Hua F, Leung YY, Yang W, et al. Nature and Diffusion of COVID-19-related Oral Health Information on Chinese Social Media: Analysis of Tweets on Weibo. *J Med Internet Res* 2020 Jun 15;22(6):e19981 [FREE Full text] [doi: [10.2196/19981](https://doi.org/10.2196/19981)] [Medline: [32501808](https://pubmed.ncbi.nlm.nih.gov/32501808/)]
15. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining Physicians' Opinions on Social Media to Obtain Insights Into COVID-19: Mixed Methods Analysis. *JMIR Public Health Surveill* 2020 Jun 18;6(2):e19276 [FREE Full text] [doi: [10.2196/19276](https://doi.org/10.2196/19276)] [Medline: [32421686](https://pubmed.ncbi.nlm.nih.gov/32421686/)]
16. Budhwani H, Sun R. Creating COVID-19 Stigma by Referencing the Novel Coronavirus as the "Chinese virus" on Twitter: Quantitative Analysis of Social Media Data. *J Med Internet Res* 2020 May 06;22(5):e19301 [FREE Full text] [doi: [10.2196/19301](https://doi.org/10.2196/19301)] [Medline: [32343669](https://pubmed.ncbi.nlm.nih.gov/32343669/)]
17. Rufai SR, Bunce C. World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. *J Public Health (Oxf)* 2020 Aug 18;42(3):510-516 [FREE Full text] [doi: [10.1093/pubmed/fdaa049](https://doi.org/10.1093/pubmed/fdaa049)] [Medline: [32309854](https://pubmed.ncbi.nlm.nih.gov/32309854/)]
18. Park HW, Park S, Chong M. Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea. *J Med Internet Res* 2020 May 05;22(5):e18897 [FREE Full text] [doi: [10.2196/18897](https://doi.org/10.2196/18897)] [Medline: [32325426](https://pubmed.ncbi.nlm.nih.gov/32325426/)]
19. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. *JMIR Public Health Surveill* 2020 May 22;6(2):e19447 [FREE Full text] [doi: [10.2196/19447](https://doi.org/10.2196/19447)] [Medline: [32412418](https://pubmed.ncbi.nlm.nih.gov/32412418/)]
20. Pobiruchin M, Zowalla R, Wiesner M. Temporal and Location Variations, and Link Categories for the Dissemination of COVID-19-Related Information on Twitter During the SARS-CoV-2 Outbreak in Europe: Infoveillance Study. *J Med Internet Res* 2020 Aug 28;22(8):e19629 [FREE Full text] [doi: [10.2196/19629](https://doi.org/10.2196/19629)] [Medline: [32790641](https://pubmed.ncbi.nlm.nih.gov/32790641/)]
21. Lamsal R. Coronavirus (COVID-19) Tweets Dataset. IEEE Dataport. URL: <https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset> [accessed 2020-10-20]
22. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health Surveill* 2020 May 29;6(2):e19273 [FREE Full text] [doi: [10.2196/19273](https://doi.org/10.2196/19273)] [Medline: [32427106](https://pubmed.ncbi.nlm.nih.gov/32427106/)]
23. GetOldTweets3. Python Package Index. URL: <https://pypi.org/project/GetOldTweets3/> [accessed 2020-10-20]
24. Hydrator. GitHub. URL: <https://github.com/DocNow/hydrator> [accessed 2020-10-20]
25. Wood-Doughty Z, Mahajan P, Dredze M. Johns Hopkins or johnny-hopkins: Classifying Individuals versus Organizations on Twitter. 2018 Presented at: Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media; June 6, 2018; New Orleans, LA. [doi: [10.18653/v1/w18-1108](https://doi.org/10.18653/v1/w18-1108)]
26. Kwon KH, Priniski JH, Chadha M. Disentangling User Samples: A Supervised Machine Learning Approach to Proxy-population Mismatch in Twitter Research. *Commun Methods Meas* 2018 Feb 15;12(2-3):216-237. [doi: [10.1080/19312458.2018.1430755](https://doi.org/10.1080/19312458.2018.1430755)]
27. Zhu JM, Sarker A, Gollust S, Merchant R, Grande D. Characteristics of Twitter Use by State Medicaid Programs in the United States: Machine Learning Approach. *J Med Internet Res* 2020 Aug 17;22(8):e18401 [FREE Full text] [doi: [10.2196/18401](https://doi.org/10.2196/18401)] [Medline: [32804085](https://pubmed.ncbi.nlm.nih.gov/32804085/)]
28. Miller M, Banerjee T, Muppalla R, Romine W, Sheth A. What Are People Tweeting About Zika? An Exploratory Study Concerning Its Symptoms, Treatment, Transmission, and Prevention. *JMIR Public Health Surveill* 2017 Jun 19;3(2):e38 [FREE Full text] [doi: [10.2196/publichealth.7157](https://doi.org/10.2196/publichealth.7157)] [Medline: [28630032](https://pubmed.ncbi.nlm.nih.gov/28630032/)]
29. Hwang Y, Kim HJ, Choi HJ, Lee J. Exploring Abnormal Behavior Patterns of Online Users With Emotional Eating Behavior: Topic Modeling Study. *J Med Internet Res* 2020 Mar 31;22(3):e15700 [FREE Full text] [doi: [10.2196/15700](https://doi.org/10.2196/15700)] [Medline: [32229461](https://pubmed.ncbi.nlm.nih.gov/32229461/)]
30. Srinivasa-Desikan B. Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras. Birmingham, UK: Packt Publishing Ltd; 2018.
31. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. *J Mach Learn Res* 2003 Jan;3(Jan):993-1022 [FREE Full text]
32. Jagarlamudi J, Iii H, Udupa R. Incorporating Lexical Priors into Topic Models. 2012 Presented at: 13th Conference of the European Chapter of the Association for Computational Linguistics; April 23-27, 2012; Avignon, France URL: <https://www.aclweb.org/anthology/E12-1021.pdf>

33. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. 2015 Jan Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI URL: <http://eegilbert.org/papers/icwsm14.vader.hutto.pdf>
34. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: [10.2196/19118](https://doi.org/10.2196/19118)] [Medline: [32302966](https://pubmed.ncbi.nlm.nih.gov/32302966/)]

## Abbreviations

**LDA:** latent Dirichlet allocation

**NLTK:** Natural Language Toolkit

**RegEx:** regular expression

**VADER:** Valence Aware Dictionary and sEntiment Reasoner

**WHO:** World Health Organization

*Edited by G Eysenbach, G Fagherazzi; submitted 18.07.20; peer-reviewed by R Zowalla, L Xiong, Anonymous; comments to author 08.08.20; revised version received 26.08.20; accepted 26.09.20; published 23.10.20.*

*Please cite as:*

*Chandrasekaran R, Mehta V, Valkunde T, Moustakas E*

*Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Inveillance Study*

*J Med Internet Res 2020;22(10):e22624*

*URL: <http://www.jmir.org/2020/10/e22624/>*

*doi: [10.2196/22624](https://doi.org/10.2196/22624)*

*PMID: [33006937](https://pubmed.ncbi.nlm.nih.gov/33006937/)*

©Ranganathan Chandrasekaran, Vikalp Mehta, Tejali Valkunde, Evangelos Moustakas. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 23.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Physical Distancing Measures and Walking Activity in Middle-aged and Older Residents in Changsha, China, During the COVID-19 Epidemic Period: Longitudinal Observational Study

Yilun Wang<sup>1</sup>, MD, PhD; Yuqing Zhang<sup>2,3</sup>, DSC; Kim Bennell<sup>4</sup>, PhD; Daniel Kenta White<sup>5</sup>, PT, SCD; Jie Wei<sup>2,3,6</sup>, PhD; Ziyang Wu<sup>1</sup>, MD; Hongyi He<sup>1</sup>, MD; Shaohui Liu<sup>6</sup>, MS; Xianghang Luo<sup>7,8</sup>, MD, PhD; Shuo Hu<sup>8,9</sup>, MD, PhD; Chao Zeng<sup>1,2,3\*</sup>, MD, PhD; Guanghua Lei<sup>1,8,10\*</sup>, MD, PhD

<sup>1</sup>Department of Orthopaedics, Xiangya Hospital, Central South University, Changsha, China

<sup>2</sup>Division of Rheumatology, Allergy, and Immunology, Department of Medicine, Massachusetts General Hospital, Boston, MA, United States

<sup>3</sup>The Mongan Institute, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

<sup>4</sup>Centre for Health, Exercise and Sports Medicine, The University of Melbourne, Melbourne, Australia

<sup>5</sup>Department of Physical Therapy, University of Delaware, Newark, DE, United States

<sup>6</sup>Health Management Center, Xiangya Hospital, Central South University, Changsha, China

<sup>7</sup>Department of Endocrinology, Endocrinology Research Center, Xiangya Hospital, Central South University, Changsha, China

<sup>8</sup>National Clinical Research Center of Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

<sup>9</sup>Department of Nuclear Medicine, Xiangya Hospital, Central South University, Changsha, China

<sup>10</sup>Hunan Key Laboratory of Joint Degeneration and Injury, Changsha, China

\*these authors contributed equally

**Corresponding Author:**

Guanghua Lei, MD, PhD

Department of Orthopaedics

Xiangya Hospital

Central South University

Xiangya Road 87

Changsha, 410008

China

Phone: 86 0731 89753006

Email: [lei\\_guanghua@csu.edu.cn](mailto:lei_guanghua@csu.edu.cn)

## Abstract

**Background:** Physical distancing measures taken to contain COVID-19 transmission may substantially reduce physical activity levels and cause individuals to adopt a more sedentary lifestyle.

**Objective:** The objective of this study is to determine if there was any change in daily steps, an important component of daily physical activity, and examine risk factors for frequent low daily steps during the COVID-19 epidemic.

**Methods:** We used data collected from the Step Study, a population-based longitudinal study of walking activity among residents aged  $\geq 40$  years in Changsha, China. Daily steps were collected via a smartphone linked to WeChat, a social networking platform. We plotted mean daily steps and the prevalence of low daily steps ( $\leq 1500$  steps/day) 30 days before (reference period) and 30 days after (epidemic period) January 21, 2020 (date of the first COVID-19 case diagnosed in Changsha), and compared it with the same corresponding period from 2019. We examined the association of risk factors with the prevalence of frequent low daily steps ( $\leq 1500$  steps/day for  $\geq 14$  days) using logistic regression.

**Results:** Among 3544 participants (mean age 51.6 years;  $n=1226$  females, 34.6%), mean daily steps dropped from 8097 to 5440 and the prevalence of low daily steps increased from 3% (2287/76,136 person-day) to 18.5% (12,951/70,183 person-day) during the reference and epidemic periods, respectively. No such phenomenon was observed during the corresponding period in 2019. Older age ( $P$  for interaction=.001) and female sex ( $P$  for interaction<.001) were both associated with a higher prevalence of frequent low daily steps and were more pronounced during the epidemic period. More education was associated with a lower prevalence of frequent low daily steps during the reference period but not the epidemic period ( $P$  for interaction=.34). Body mass index or comorbidity were not associated with frequent low daily steps during either period.



**Conclusions:** Daily steps of Changsha residents aged  $\geq 40$  years dropped significantly during the COVID-19 period, especially among older adults and females. Although successful physical distancing, measured by the rapid downward trend in daily step counts of residents, played a critical role in the containment of the COVID-19 epidemic, our findings of an increase in the prevalence of frequent low daily steps raise concerns about unintended effects on physical activity.

(*J Med Internet Res* 2020;22(10):e21632) doi:[10.2196/21632](https://doi.org/10.2196/21632)

## KEYWORDS

COVID-19; pandemic; physical distancing; steps; walking activity

## Introduction

COVID-19 has caused morbidity and mortality worldwide [1]. To control this highly contagious infectious disease, many countries have implemented “physical distancing” and “shelter-in-place” measures to contain COVID-19 transmission [2,3]. However, such measures may substantially reduce physical activity levels and cause individuals to adopt a more sedentary lifestyle [4].

There is ample evidence that regular physical activity is crucial for health, with physical inactivity characterized as the fourth leading cause of global mortality [5]. Results from the National Health and Nutrition Examination Survey (NHANES) and the Women’s Health Study both showed that a lower number of daily steps, which is an important component of daily physical activity, was significantly associated with higher all-cause mortality [6,7]. Studies also demonstrated that individuals who walked less than 1500 daily steps for 14 days were at a higher risk of muscle mass loss and low insulin sensitivity [8,9]. To date, several studies have reported changes in daily steps of residents in areas affected by COVID-19 [10–14]; however, none of these studies specifically described the prevalence of frequent low daily steps ( $\leq 1500$  steps/day for  $\geq 14$  days over one month) [8,9], a strong predictor of poor health outcomes, and examined its risk factors.

Using data collected from the Step Study, a longitudinal study conducted among the urban residents of Changsha, China, we therefore described the trends of daily steps around the COVID-19 epidemic period in 2020 and compared them with that of a similar period one year earlier. We examined the relationships between several sociodemographic factors, anthropometric factors, and comorbidity and the prevalence of frequent low daily steps.

## Methods

### Study Population

The Step Study (registration number: ChiCTR1800017977) is a population-based longitudinal study initiated in September 2018 in Changsha, China. The aims of the Step Study were to describe patterns of walking activity, factors related to walking activity, and its sequelae among the community-living population. Participants in the Step Study comprised individuals who had their annual physical checkup at Xiangya Hospital of Central South University in Changsha, China [15,16]. To be eligible for the Step Study, participants had to meet the following criteria: (1) be a resident of Changsha; (2) be aged  $\geq 40$  years; (3) own a personal smartphone and have a WeChat

account; and (4) be willing and able to give consent. Individuals with severe mental illness were ineligible. During each annual physical checkup, participants were queried about their sociodemographic and lifestyle factors, comorbidities, health-related symptoms and signs (eg, joint pain), and ability to perform daily activities. They also received clinical examinations and laboratory tests, including physical function tests (eg, lower limb muscle strength measurements).

For each participant, walking activity measured as daily step counts was collected through a smartphone linked to WeChat. WeChat is a multipurpose social network platform (Tencent Inc) with approximately 1.1 billion monthly active users in China [17]. One of its apps can extract daily step count information from the accelerometer sensor in a smartphone. Thus, if a participant is a WeChat user and wears a smartphone, his/her step counts can be captured by WeChat’s app.

### Study Outcome

The outcome variable was daily step counts recorded by the accelerometer sensor in the smartphone and extracted by WeChat. To be eligible for the current analysis of daily step counts, we required participants to wear their smartphone for  $\geq 10$  hours on a given day, a standard convention for measuring daily walking [18,19]. The smartphone wear time was calculated as the difference between the times of the first recorded step count and the last recorded step count each day. This algorithm was used in the Activity Inequality Project to calculate daily step counts for more than 700,000 individuals across 111 countries [19]. Participants with no valid daily step count were excluded in the current analysis. We defined a low daily step count as  $\leq 1500$  steps/day [8,9]. If a participant had  $\geq 14$  days of low daily step counts over a 30-day period, we considered that person as having experienced frequent low daily steps [8,9].

We conducted two validation studies to assess the accuracy of daily step counts collected from WeChat. To determine the accuracy of steps measured at various walking speeds by iOS and Android devices, we visually counted steps from 14 subjects walking on a treadmill at 4.8, 6.4, 8.0, and 9.6 km/h while subjects held/wore smartphones in different positions (ie, pants pocket, hand, and arm). These methods are consistent with previous studies [20–22]. We found step count accuracy to be high, with intraclass correlation coefficients (ICCs) ranging from 0.64 to 0.99. Second, we assessed the accuracy of step counts extracted from WeChat in free-living conditions. Specifically, 36 participants from the Step Study were instructed to wear both a Fitbit Charge 3 (Fitbit Inc), as the criterion measure [23], and their personal smartphone for 7 consecutive days [24]. The results also demonstrated a moderate to high

agreement on step counts, with ICCs ranging from 0.67 to 0.81. Detailed information from these validation studies is shown in [Multimedia Appendix 1](#).

### Study Exposures

Information on age, sex, educational level, height, weight, and comorbidity were obtained from the annual physical checkup visit. Body mass index was calculated. The modified Charlson Comorbidity Index (CCI) was computed based on self-reported comorbidities [25].

### Statistical Analyses

On January 22, 2020, one day after the first case of COVID-19 was diagnosed in Changsha (January 21, 2020, ie, the index day) the municipal government issued an emergency notice to implement physical distancing measures (ie, staying at home, closing schools, working from home if possible, travelling only when necessary, and cancelling mass gatherings) [26,27]. In this analysis, we defined the time interval from January 22 to February 20, 2020 (30 days after the index date), as the COVID-19 epidemic period, and the time interval from December 22, 2019, to January 20, 2020 (30 days before the index date), as the reference period. Since the COVID-19 epidemic occurred around the holiday season of Chinese New Year, for the purpose of comparison we also used data collected between January 2 and March 3, 2019, which corresponded to the same Chinese lunar calendar period. Details of the selection of study periods are shown in [Multimedia Appendix 2](#).

First, we plotted the mean daily steps from December 22, 2019, to February 20, 2020 (around COVID-19 epidemic period), and mean daily steps from January 2 to March 3, 2019 (historic comparison period), respectively, using a 3-day moving average smooth method [28]. Second, we used the same approach to plot the prevalence of low daily steps ( $\leq 1500$  steps/day) for the corresponding periods. We calculated the mean difference (MD) and its 95% confidence interval for daily steps using Generalized Estimating Equations (GEE) between the epidemic period and reference period in 2020, and the corresponding periods in 2019, respectively. Specifically, we included each qualified daily step count into the GEE model using the PROC GENMOD procedure

in SAS (Version 9.4; SAS Institute) with identity links to calculate the MD and its 95% CI between the epidemic and reference periods. We added the REPEATED statement to account for correlation of the repeated measurements of individuals' daily step counts [29].

Similarly, prevalence ratios (PR) were calculated for the prevalence of low daily steps ( $\leq 1500$  steps/day) between the two comparative time periods in 2019 and 2020, respectively [30]. Finally, we estimated the prevalence of frequent low daily steps ( $\leq 1500$  steps/day for  $\geq 14$  days over 30 days) during epidemic and reference periods and examined whether age, sex, BMI, educational level, and comorbidity were associated with the prevalence of frequent low daily steps using logistic regression. We tested the additive effect measure of modification of physical distancing with each of the risk factors mentioned above by adding an interaction term in the regression model [31].

All  $P$  values for interaction were two-sided and  $P$  for interaction  $< .05$  was considered statistically significant for all tests. All statistical analyses were conducted using SAS (Version 9.4).

### Ethical Considerations

This study was approved by the Ethics Committee of Xiangya Hospital, Central South University (#201806910), and written informed consent was obtained from study participants.

## Results

A total of 7262 Changsha residents aged  $\geq 40$  years had a physical checkup at the study center between September 2018 and January 2020, and 4145 of them (57.1%) had a WeChat account and agreed to participate in the Step Study. Of these, 3544 with at least one valid daily step count during the study period were included in the analysis ([Table 1](#)). The mean age of participants was 51.6 (SD 8.9) years, 1226 (34.6%) participants were females, and the mean BMI of participants was 24.0 (SD 4.3)  $\text{kg/m}^2$ . Overall, 2616 (73.8%) participants were Android users, 765 (21.6%) were iOS users, and 163 (4.6%) participants' phone types were unknown.

**Table 1.** Baseline characteristics of included participants.

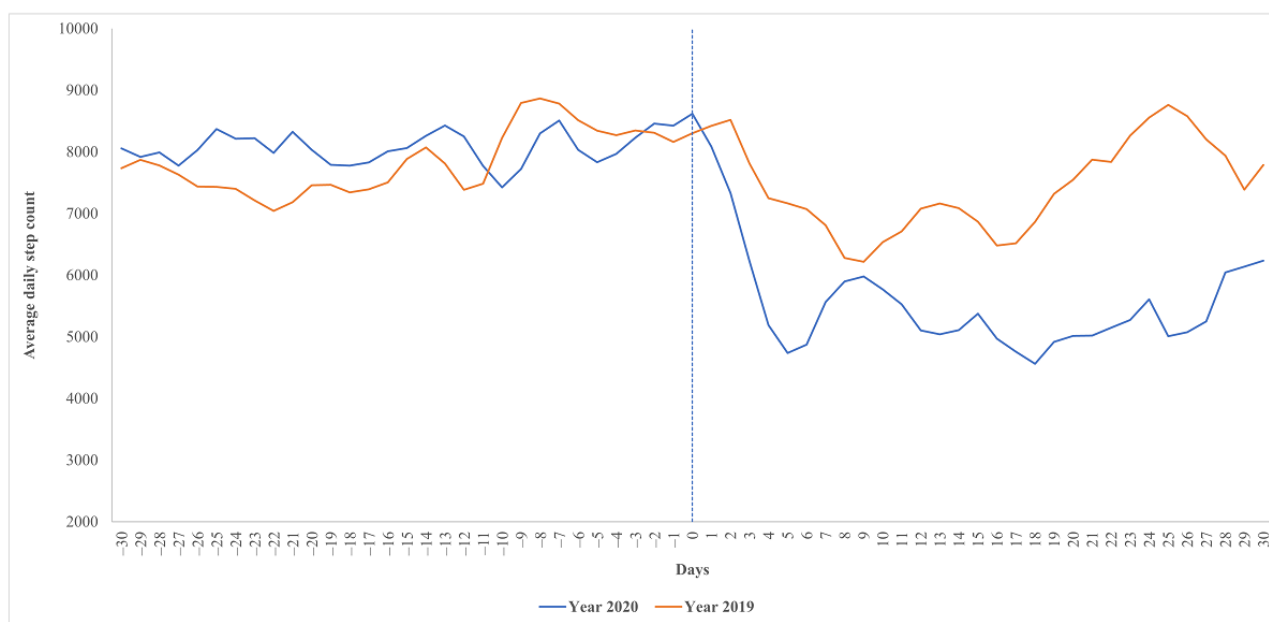
Characteristics	Total sample (N=3544)	Males (n=2318)	Females (n=1226)
Age (years), mean (SD)	51.6 (8.9)	51.6 (8.8)	51.5 (9.0)
<b>Age (years), n (%)</b>			
40-49	1733 (48.9)	1117 (48.2)	616 (50.2)
50-59	1190 (33.6)	818 (35.3)	372 (30.3)
60-70	452 (12.7)	271 (11.7)	181 (14.8)
≥70	169 (4.8)	112 (4.8)	57 (4.7)
BMI (kg/m <sup>2</sup> ), mean (SD) <sup>a</sup>	24.0 (4.3)	24.8 (4.2)	22.5 (4.3)
<b>BMI (kg/m<sup>2</sup>), n (%)<sup>b</sup></b>			
<25	1908 (59.6)	1043 (49.2)	865 (80.0)
≥25	1294 (40.4)	1075 (50.8)	219 (20.0)
<b>Education, n (%)<sup>b</sup></b>			
High school or below	564 (22.1)	328 (19.7)	236 (26.5)
Junior college	621 (24.3)	372 (22.4)	249 (27.9)
University or above	1369 (53.6)	962 (57.9)	407 (45.6)
<b>Charlson Comorbidity Index, n (%)</b>			
0	2759 (77.8)	1789 (77.2)	970 (79.1)
≥1	785 (22.2)	529 (22.8)	256 (20.9)

<sup>a</sup>N=3202.<sup>b</sup>N=2554.

As shown in [Figure 1](#) and [Table 2](#), daily steps (mean 8097 steps/day, range 6942-9153 steps/day) during the reference period (30 days prior to the COVID-19 epidemic in 2020) were similar to that during the corresponding period in 2019 (mean 7872 steps/day, range 6649-8912 steps/day). However, the daily steps decreased substantially after implementing physical

distancing measures, from 8624 steps/day on the day before the index day to 4121 steps/day on Day 4 after the index day; this trend continued during the rest of the epidemic period. Compared with the reference period, the mean daily steps dropped by 2678 steps (95% CI 2582-2763). However, this trend was not observed during the corresponding period in 2019.

**Figure 1.** Average daily step count around the Chinese Lunar New Year period among participants in 2019 and 2020. Day 0 represents the index date; in 2020, this represents the date of the first COVID-19 case diagnosed in Changsha.



**Table 2.** Associations of time period with daily step count.

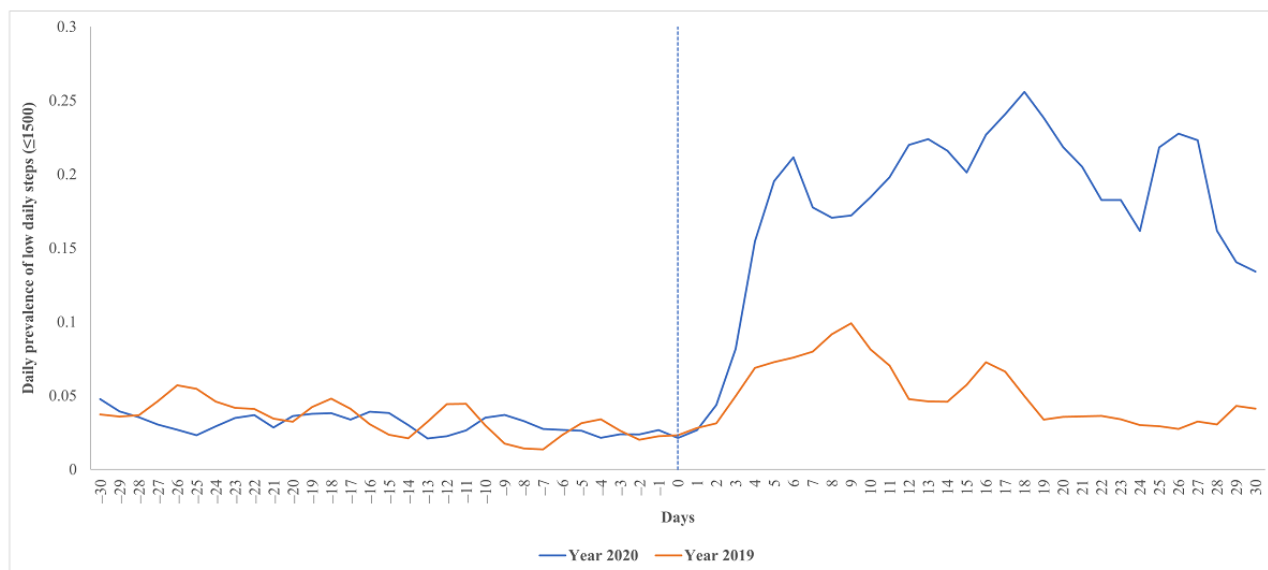
Time period	2019 daily step count		2020 daily step count	
	Mean (SD)	Mean difference (95% CI) <sup>a</sup>	Mean (SD)	Mean difference (95% CI) <sup>a</sup>
Reference	7872 (4842)	0 (reference)	8097 (4793)	0 (reference)
Epidemic	7472 (4979)	-413 (-501 to -325)	5440 (4571)	-2672 (-2763 to -2582)

<sup>a</sup>Mean differences were adjusted for age and sex.

Figure 2 and Table 3 present the prevalence of low daily steps ( $\leq 1500$  steps/day) in 2019 and 2020. The prevalence of low daily steps was similar during the reference period in 2020 (3.0%, 2287/76,136 person-day) and the corresponding period in 2019 (3.3%, 1006/30,647 person-day). In contrast, the prevalence of low daily steps increased substantially after implementing physical distancing measures, from 2.0%

(53/2693) on the day prior to the index date to 25.5% (639/2505) on Day 4 after the index date; this trend continued during the rest of the follow-up period. Compared with the reference period, the PR of low daily steps was 6.2 (95% CI 5.8-6.7). However, no such trend was observed during the entire historic comparison period in 2019.

**Figure 2.** Prevalence of low daily step count ( $\leq 1500$  steps/day) around the Chinese Lunar New Year period among participants in 2019 and 2020. Day 0 represents the index date; in 2020, this represents the date of the first COVID-19 case diagnosed in Changsha.



**Table 3.** Associations of time period with prevalence of low daily steps ( $\leq 1500$  steps/day).

Time period	2019 low daily steps		2020 low daily steps	
	Person-day, n (%)	Prevalence ratio (95% CI) <sup>a</sup>	Person-day, n (%)	Prevalence ratio (95% CI) <sup>a</sup>
Reference	1006 (3.3)	1.0 (reference)	2287 (3.0)	1.0 (reference)
Epidemic	1488 (5.1)	1.6 (1.4-1.7)	12,951 (18.5)	6.2 (5.8-6.7)

<sup>a</sup>Prevalence ratios were adjusted for age and sex.

As shown in Table 4, only 12 (0.4%) of 2879 participants had walked less than  $\leq 1500$  steps/day for 14 days or more (frequent low daily steps) during the 2020 reference period; however, the prevalence of frequent low daily steps increased to 7.4% (196/2655) during the COVID-19 epidemic period, after physical distancing measures were implemented. Older age ( $P$  for interaction=.001) and female sex ( $P$  for interaction<.001) were both associated with a higher prevalence of frequent low daily steps than their counterparts during the reference and COVID-19 epidemic periods, and such associations were more

pronounced during the epidemic period. Participants with a university education or more had a lower prevalence of frequent low daily steps than those with only high school or less education during the reference period but this was not observed during the COVID-19 epidemic period. During the physical distancing period, no significant interaction between education and prevalence of frequent low daily steps was observed ( $P$  for interaction=.34). There was no apparent association of either BMI or CCI with the prevalence of frequent low daily steps during either period.



**Table 4.** Association of prevalence of frequent low daily steps cases ( $\leq 1500$  steps/day for  $\geq 14$  days) and basic characteristics.

Characteristics	Reference period			Epidemic period			<i>P</i> values for interaction
	Participants, n	Cases, n (%)	PR <sup>a</sup> (95% CI) <sup>b</sup>	Participants, n	Cases, n (%)	PR (95% CI) <sup>b</sup>	
Total	2879	12 (0.4)	N/A <sup>c</sup>	2655	196 (7.4)	N/A	N/A
<b>Age (years)</b>							.001
40-49	1543	1 (0.1)	1.0 (ref) <sup>d</sup>	1470	84 (5.7)	1.0 (ref)	N/A
50-59	957	8 (0.8)	9.9 (1.2-0.0)	871	69 (7.9)	1.7 (1.2-2.3)	N/A
60-70	308	2 (0.7)	9.2 (0.8-102.6)	257	32 (12.5)	2.2 (1.5-3.4)	N/A
$\geq 70$	71	1 (1.4)	21.0 (1.3-331.8)	57	11 (19.3)	3.0 (1.6-5.7)	N/A
<b>Sex</b>							<.001
Males	1904	7 (0.4)	1.0 (ref)	1793	75 (4.2)	1.0 (ref)	N/A
Females	975	5 (0.5)	2.2 (0.6-7.9)	862	121 (14.0)	3.4 (2.5-4.6)	N/A
<b>Body mass index</b>							.250
<25	1557	7 (0.5)	1.0 (ref)	1415	109 (7.7)	1.0 (ref)	N/A
$\geq 25$	1006	3 (0.3)	0.8 (0.2-3.2)	269	61 (6.2)	1.2 (0.9-1.7)	N/A
<b>Education</b>							.335
High school or below	441	6 (1.4)	1.0 (ref)	389	45 (11.6)	1.0 (ref)	N/A
Junior college	522	2 (0.4)	0.3 (0.1-1.7)	460	31 (6.7)	0.7 (0.4-1.0)	N/A
University or above	1228	2 (0.2)	0.2 (0.0-1.1)	1134	90 (7.9)	1.0 (0.7-1.4)	N/A
<b>Charlson Comorbidity Index</b>							.345
0	2294	9 (0.4)	1.0 (ref)	2115	150 (7.1)	1.0 (ref)	N/A
$\geq 1$	585	3 (0.5)	0.9 (0.2-3.8)	540	46 (8.5)	1.1 (0.7-1.5)	N/A

<sup>a</sup>PR: prevalence ratio.<sup>b</sup>Prevalence ratios were adjusted for age and sex.<sup>c</sup>N/A: not applicable.<sup>d</sup>Ref: reference.

## Discussion

### Principal Findings

Using objective data collected from the longitudinal Step Study, we found that daily steps among middle-aged and older residents in Changsha dropped rapidly and substantially (2672 fewer daily steps on average) after implementing physical distancing measures during the COVID-19 epidemic period. In addition, more than 7% (196/2655) of residents had walked  $\leq 1500$  steps/day for  $\geq 14$  days over the one-month epidemic period compared with 0.4% (12/2879) of residents in the month prior to the epidemic. The reduction of steps/day during the COVID-19 epidemic was more pronounced among older adults and females.

### Comparison With Previous Studies

To date, several studies have reported changes in daily steps during the COVID-19 epidemic [10-14]. One worldwide study based on a smartphone app (Argus) showed that mean daily steps in different regions decreased by 5.5% and 27.3% (287 and 1432 steps/day, respectively) within 10 and 30 days after the COVID-19 pandemic was declared [10]. Another study that

used a wristwatch with an embedded accelerometer (Withings) demonstrated a marked decrease in daily steps (from 25% to 54%) following the official dates of home confinement in countries adopting a total lockdown [11]. Similar findings were also reported in other countries [12-14]. Our results demonstrated that such a change also occurred in the Chinese population. Furthermore, we examined daily steps within the same period in the previous year and observed no such change during this period, which enabled us to minimize the potential impact of the holiday season of Chinese New Year on daily steps.

In addition, we found that the effect of implementation of physical distancing measures on frequent low daily steps was more pronounced among older adults and females. Previous studies have examined associations between walking activities and sociodemographic factors, anthropometric factors, and comorbidity. The results from the NHANES report indicated that those of advancing age (OR 1.95, per 16.7-year increments) and female sex (OR 1.86) both had higher odds of walking less than 5000 steps/day [32]. Another study found that education was associated with increased walking activity, with one additional year of education associated with a 560 daily steps

increase [33]. Our results corroborate these findings. However, after the implementation of physical distancing measures, the prevalence of frequent low daily steps among residents with university or above education was similar to those with high school or below education, suggesting both groups followed physical distancing measures and reduced their outdoor walking activities. Nevertheless, the magnitude of relative increase in the prevalence of frequent low daily steps during the COVID-19 epidemic appeared to be greater among residents with university or above education than among those with high school or below education, indicating the former are more likely to follow instructions and communicate effectively with health providers [33,34]. Previous studies also showed that both BMI and comorbidities (eg, hypertension and diabetes) were associated with fewer daily steps [6,7,35,36], but this was not the case in the present study, nor was any association modified by the implementation of physical distancing measures.

### Public Health Implications

Physical distancing measures play a critical role in containing COVID-19 transmission and monitoring population mobility data can provide evidence as to whether people are complying with these measures [37]; however, the impact of physical distancing on other aspects of daily life should not be overlooked. Recently, Hall and colleagues [4] commented that “The world is experiencing an extraordinary, life-altering challenge due to the COVID-19 pandemic. Many countries have become accustomed to a new normal – ‘social distancing’ and ‘shelter-in-place’ are now a part of everyday vernacular and life.” The authors warned that this health crisis has the potential to further impact and accelerate the pandemic of physical inactivity and sedentary behavior.

Our data showed that average daily steps among residents of Changsha dropped by more than 30% (from 8097 to 5440) during the COVID-19 epidemic period after implementation of physical distancing measures. These data raise concerns about the potential adverse effects of such measures on health and well-being. Previous studies have showed that higher daily steps are associated with better cardiometabolic profiles and lower all-cause mortality [6,7,38]. Other studies have also found that a decrease in daily steps of 2000 steps, irrespective of previous habitual step counts, increases the risk of insulin sensitivity and higher cardiovascular events [39,40]. Thus, our findings have important implications for public health recommendations and the prevention of other health crises during the COVID-19 pandemic. Furthermore, our data showed that the reduction in daily steps is much more common among older adults and females. In general, older adults and females are more likely to develop sedentary behaviors, placing them at greater risk of various diseases related to inactivity [32,41]; thus, the worsening trend of physical inactivity during the COVID-19 epidemic period compounds the risk of sequelae related to a sedentary lifestyle.

It is uncertain as to how long it will take to completely control the COVID-19 pandemic worldwide. If similar trends toward a sedentary lifestyle are seen in other countries, the avoidance of further sedentary lifestyle behaviors and promotion of regular physical activity during this time are an urgent global public

health issue. It is also unknown whether the observed decline in daily walking is a temporary phenomenon that may revert back to baseline levels after the disease is under control; thus, further longitudinal studies are needed so that evidence-based strategies can be developed and implemented to encourage greater participation in regular physical activity.

### Strengths and Limitations

Several strengths of our study are noteworthy. First, we used data from a large population-based longitudinal study (Step Study), which allowed not only investigation of COVID-19–related changes in daily step counts from over 3500 residents, but also inclusion of a historic comparison period from the prior year to account for a secular trend in daily step counts. Second, the study took advantage of a social network platform (WeChat) to capture daily step count data via smartphone in real time among the community-living population. As a smartphone has become a daily necessity for most adults, this approach allowed long-term monitoring of daily step count trends. This contrasts to previous studies using wearable devices (accelerometers) that generally only collect step count data for a relatively short time (eg, a week or less) [42]. It can be challenging to extrapolate such short-term data to describe step count patterns over a long period owing to various potential confounders [42].

Potential limitations of our study also deserve comment. First, the WeChat app may underestimate daily step counts because some individuals may not always carry their smartphone with them [43], especially when participants were housebound. Thus, some light walking activities at home may not be captured. However, one previous study reported that the average wear time of smartphones among Chinese citizens was more than 13 hours during the day time, indicating that most walking activities should be captured by a smartphone [19]. Second, although our validation study demonstrated that daily step counts collected from iOS and Android devices both showed close agreement with actual step counts under controlled laboratory settings, previous studies found that iOS and Android devices have different accuracies in capturing the daily step count under free-living conditions [44]. This could have led to misclassification of daily step counts in this study. Third, participants in our study were slightly younger (51.6 versus 54.8 years) and more likely to be male (65.4% versus 51.2%) than those who are aged  $\geq 40$  years and live in Changsha, according to the latest census data in 2010. However, our results showed that the prevalence of frequent low daily steps increased more among women and older people, suggesting that the overall prevalence of frequent low steps among all residents in Changsha aged  $\geq 40$  years may be even higher than what was reported in our study. Fourth, participants in our study were urban residents in Changsha who came to Xiangya Hospital for their annual health checkup. The percentage of WeChat users among these individuals (57.1%) was higher than that of the population with the same age range in China (41.5%) [45]. Thus, our findings may not be generalizable to residents living in other parts of China, especially those living in rural areas. Finally, we were unable to capture the intensity of steps (eg, slow versus fast steps); however, total steps/day has recently been shown

to be an important predictor of mortality independent of step intensity [6].

## Conclusions

Using data collected from a large population-based longitudinal study, we demonstrated that walking activity, indicated by daily step count, decreased rapidly and substantially during the COVID-19 epidemic period among middle-aged and older adult

Chinese residents living in urban areas. These results suggest that appropriate strategies need to be taken to encourage residents to actively engage in regular physical activity while maintaining personal hygiene and physical distancing. They also call for further studies to evaluate whether the low levels of walking activity observed following the implementation physical distancing measures will be maintained and whether they will have significant adverse impacts on health outcomes.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (81772413, 81702207, 81702206, 81930071), the National Health and Medical Research Council Principal Research Fellowship, the National Key Research and Development Project (2018YFB1105705), and the Key Research and Development Program of Hunan Province (2018SK2070, 2018SK2071). No funding bodies had any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors thank Xiaoxiao Li, Wei Li, Huizhong Long, Jiatian Li, Haochen Wang, Yuqing Wang, Ning Wang, Bei Xu, Zidan Yang, Jing Wu, Zhenglei Zhu, and Xin Huang (Xiangya Hospital) for their contribution to this study.

## Authors' Contributions

GL and CZ had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. CZ and GL are joint corresponding authors. All authors have read the manuscript, provided critical feedback on intellectual content, and approved the final manuscript. CZ, GL, YW, and YZ contributed to the concept and design. All authors contributed to the acquisition, analysis, or interpretation of data. YW, CZ, GL, and YZ drafted the manuscript. All authors critically revised the manuscript for important intellectual content. YZ and JW contributed to statistical analysis. CZ, GL, and JW obtained funding. YW, CZ, and GL provided administrative, technical, or material support. GL, YZ, and CZ provided supervision.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Validation study.

[DOCX File, 55 KB - [jmir\\_v22i10e21632\\_app1.docx](#)]

### Multimedia Appendix 2

Diagram of the study periods. Chinese New Year fell on January 25 in 2020 and February 5 in 2019.

[PNG File, 234 KB - [jmir\\_v22i10e21632\\_app2.png](#)]

## References

1. Coronavirus disease (COVID-19) Pandemic. World Health Organization. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> [accessed 2020-03-30]
2. Merriam-Webster. Social Distancing. 2020. URL: <https://www.merriam-webster.com/dictionary/social%20distancing> [accessed 2020-03-30]
3. Miller K. Here's what a shelter in place order means during the coronavirus pandemic. Prevention. URL: <https://www.prevention.com/health/amp31738348/shelter-in-place-ordercoronavirus/> [accessed 2020-03-30]
4. Hall G, Laddu DR, Phillips SA, Lavie CJ, Arena R. A tale of two pandemics: How will COVID-19 and global trends in physical inactivity and sedentary behavior affect one another? Prog Cardiovasc Dis 2020 Apr 08;1 [FREE Full text] [doi: [10.1016/j.pcad.2020.04.005](https://doi.org/10.1016/j.pcad.2020.04.005)] [Medline: [32277997](#)]
5. Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks. World Health Organization. URL: [http://apps.who.int/iris/bitstream/10665/44203/1/9789241563871\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/44203/1/9789241563871_eng.pdf) [accessed 2020-04-05]
6. Saint-Maurice PF, Troiano RP, Bassett DR, Graubard BI, Carlson SA, Shiroma EJ, et al. Association of Daily Step Count and Step Intensity With Mortality Among US Adults. JAMA 2020 Mar 24;323(12):1151-1160. [doi: [10.1001/jama.2020.1382](https://doi.org/10.1001/jama.2020.1382)] [Medline: [32207799](#)]

7. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of Step Volume and Intensity With All-Cause Mortality in Older Women. *JAMA Intern Med* 2019 May 29;1 [FREE Full text] [doi: [10.1001/jamainternmed.2019.0899](https://doi.org/10.1001/jamainternmed.2019.0899)] [Medline: [31141585](https://pubmed.ncbi.nlm.nih.gov/31141585/)]
8. Breen L, Stokes KA, Churchward-Venne TA, Moore DR, Baker SK, Smith K, et al. Two weeks of reduced activity decreases leg lean mass and induces "anabolic resistance" of myofibrillar protein synthesis in healthy elderly. *J Clin Endocrinol Metab* 2013 Jun;98(6):2604-2612. [doi: [10.1210/jc.2013-1502](https://doi.org/10.1210/jc.2013-1502)] [Medline: [23589526](https://pubmed.ncbi.nlm.nih.gov/23589526/)]
9. Knudsen SH, Hansen LS, Pedersen M, Dejgaard T, Hansen J, Hall GV, et al. Changes in insulin sensitivity precede changes in body composition during 14 days of step reduction combined with overfeeding in healthy young men. *J Appl Physiol* (1985) 2012 Jul;113(1):7-15. [doi: [10.1152/japplphysiol.00189.2011](https://doi.org/10.1152/japplphysiol.00189.2011)] [Medline: [22556394](https://pubmed.ncbi.nlm.nih.gov/22556394/)]
10. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, et al. Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study. *Annals of Internal Medicine* 2020 Jun 29;29. [doi: [10.7326/m20-2665](https://doi.org/10.7326/m20-2665)]
11. Pépin JL, Bruno RM, Yang R, Vercamer V, Jouhaud P, Escourrou P, et al. Wearable Activity Trackers for Monitoring Adherence to Home Confinement During the COVID-19 Pandemic Worldwide: Data Aggregation and Analysis. *J Med Internet Res* 2020 Jun 19;22(6):e19787 [FREE Full text] [doi: [10.2196/19787](https://doi.org/10.2196/19787)] [Medline: [32501803](https://pubmed.ncbi.nlm.nih.gov/32501803/)]
12. Di Sebastiano KM, Chulak-Bozzer T, Vanderloo LM, Faulkner G. Don't Walk So Close to Me: Physical Distancing and Adult Physical Activity in Canada. *Front Psychol* 2020;11:1895 [FREE Full text] [doi: [10.3389/fpsyg.2020.01895](https://doi.org/10.3389/fpsyg.2020.01895)] [Medline: [32849110](https://pubmed.ncbi.nlm.nih.gov/32849110/)]
13. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, RADAR-CNS Consortium. Using Smartphones and Wearable Devices to Monitor Behavioral Changes During COVID-19. *J Med Internet Res* 2020 Sep 25;22(9):e19992 [FREE Full text] [doi: [10.2196/19992](https://doi.org/10.2196/19992)] [Medline: [32877352](https://pubmed.ncbi.nlm.nih.gov/32877352/)]
14. Vetrovsky T, Frybova T, Gant I, Semerad M, Cimler R, Bunc V, et al. The detrimental effect of COVID-19 nationwide quarantine on accelerometer-assessed physical activity of heart failure patients. *ESC Heart Fail* 2020 Jul 22;1 [FREE Full text] [doi: [10.1002/ehf2.12916](https://doi.org/10.1002/ehf2.12916)] [Medline: [32696600](https://pubmed.ncbi.nlm.nih.gov/32696600/)]
15. Zeng C, Wei J, Terkeltaub R, Yang T, Choi HK, Wang Y, et al. Dose-response relationship between lower serum magnesium level and higher prevalence of knee chondrocalcinosis. *Arthritis Res Ther* 2017 Oct 24;19(1):236 [FREE Full text] [doi: [10.1186/s13075-017-1450-6](https://doi.org/10.1186/s13075-017-1450-6)] [Medline: [29065924](https://pubmed.ncbi.nlm.nih.gov/29065924/)]
16. Zeng C, Wei J, Li H, Yang T, Zhang F, Pan D, et al. Relationship between Serum Magnesium Concentration and Radiographic Knee Osteoarthritis. *J Rheumatol* 2015 Jul;42(7):1231-1236. [doi: [10.3899/jrheum.141414](https://doi.org/10.3899/jrheum.141414)] [Medline: [26034158](https://pubmed.ncbi.nlm.nih.gov/26034158/)]
17. The number of WeChat users reached 1.1 billion in 2019. SOHU. URL: [https://www.sohu.com/a/362122894\\_513553](https://www.sohu.com/a/362122894_513553) [accessed 2020-03-30]
18. Tudor-Locke C, Camhi SM, Troiano RP. A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003-2006. *Prev Chronic Dis* 2012;9:E113 [FREE Full text] [doi: [10.5888/pcd9.110332](https://doi.org/10.5888/pcd9.110332)] [Medline: [22698174](https://pubmed.ncbi.nlm.nih.gov/22698174/)]
19. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. *Nature* 2017 Dec 20;547(7663):336-339 [FREE Full text] [doi: [10.1038/nature23018](https://doi.org/10.1038/nature23018)] [Medline: [28693034](https://pubmed.ncbi.nlm.nih.gov/28693034/)]
20. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. *Scand J Med Sci Sports* 2018 Jul;28(7):1818-1827. [doi: [10.1111/sms.13074](https://doi.org/10.1111/sms.13074)] [Medline: [29460319](https://pubmed.ncbi.nlm.nih.gov/29460319/)]
21. Hekler EB, Buman MP, Grieco L, Rosenberger M, Winter SJ, Haskell W, et al. Validation of Physical Activity Tracking via Android Smartphones Compared to ActiGraph Accelerometer: Laboratory-Based and Free-Living Validation Studies. *JMIR mHealth uHealth* 2015 Apr 15;3(2):e36 [FREE Full text] [doi: [10.2196/mhealth.3505](https://doi.org/10.2196/mhealth.3505)] [Medline: [25881662](https://pubmed.ncbi.nlm.nih.gov/25881662/)]
22. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA* 2015 Feb 10;313(6):625-626. [doi: [10.1001/jama.2014.17841](https://doi.org/10.1001/jama.2014.17841)] [Medline: [25668268](https://pubmed.ncbi.nlm.nih.gov/25668268/)]
23. Straiton N, Alharbi M, Bauman A, Neubeck L, Gullick J, Bhindi R, et al. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review. *Maturitas* 2018 Jun;112:85-93. [doi: [10.1016/j.maturitas.2018.03.016](https://doi.org/10.1016/j.maturitas.2018.03.016)] [Medline: [29704922](https://pubmed.ncbi.nlm.nih.gov/29704922/)]
24. Spence JC, Burgess J, Rodgers W, Murray T. Effect of pretesting on intentions and behaviour: a pedometer and walking intervention. *Psychol Health* 2009 Sep;24(7):777-789. [doi: [10.1080/08870440801989938](https://doi.org/10.1080/08870440801989938)] [Medline: [20205026](https://pubmed.ncbi.nlm.nih.gov/20205026/)]
25. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care* 1996 Jan;34(1):73-84. [doi: [10.1097/00005650-199601000-00006](https://doi.org/10.1097/00005650-199601000-00006)] [Medline: [8551813](https://pubmed.ncbi.nlm.nih.gov/8551813/)]
26. Key points of prevention and control of COVID-2019 by Center for Disease Control and Prevention in Changsha city. Changsha Government. URL: [http://www.changsha.gov.cn/jdhy/hygg/hygg/202001/t20200122\\_6403516.html](http://www.changsha.gov.cn/jdhy/hygg/hygg/202001/t20200122_6403516.html) [accessed 2020-03-15]
27. Take epidemic prevention and control as the top priority during COVID-2019 in Changsha city. Changsha Gpvernment. URL: [http://www.changsha.gov.cn/szf/ywdt/zwdt/202001/t20200122\\_6403423.html](http://www.changsha.gov.cn/szf/ywdt/zwdt/202001/t20200122_6403423.html) [accessed 2020-03-15]
28. Harmsen SM, Chang YH, Hattrup SJ. Simple Moving Average: A Method of Reporting Evolving Complication Rates. *Orthopedics* 2016 Sep 01;39(5):e869-e876. [doi: [10.3928/01477447-20160517-02](https://doi.org/10.3928/01477447-20160517-02)] [Medline: [27220115](https://pubmed.ncbi.nlm.nih.gov/27220115/)]
29. Orelie J. Model Fitting in PROC GENMOD. SAS. URL: <https://support.sas.com/resources/papers/proceedings/proceedings/sugi26/p264-26.pdf> [accessed 2020-10-21]



30. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol* 2005 Aug 01;162(3):199-200. [doi: [10.1093/aje/kwi188](https://doi.org/10.1093/aje/kwi188)] [Medline: [15987728](https://pubmed.ncbi.nlm.nih.gov/15987728/)]
31. Rothman K. Chapter 11, Measuring Interaction. In: *Epidemiology, An Introduction*. New York: Oxford University Press; 2012.
32. Sisson SB, Camhi SM, Tudor-Locke C, Johnson WD, Katzmarzyk PT. Characteristics of step-defined physical activity categories in U.S. adults. *Am J Health Promot* 2012 Jan;26(3):152-159. [doi: [10.4278/ajhp.100326-QUAN-95](https://doi.org/10.4278/ajhp.100326-QUAN-95)] [Medline: [22208412](https://pubmed.ncbi.nlm.nih.gov/22208412/)]
33. Kari JT, Viinikainen J, Böckerman P, Tammelin TH, Pitkänen N, Lehtimäki T, et al. Education leads to a more physically active lifestyle: Evidence based on Mendelian randomization. *Scand J Med Sci Sports* 2020 Jul;30(7):1194-1204. [doi: [10.1111/sms.13653](https://doi.org/10.1111/sms.13653)] [Medline: [32176397](https://pubmed.ncbi.nlm.nih.gov/32176397/)]
34. Goldman DP, Smith JP. Can patient self-management help explain the SES health gradient? *Proc Natl Acad Sci U S A* 2002 Aug 06;99(16):10929-10934. [doi: [10.1073/pnas.162086599](https://doi.org/10.1073/pnas.162086599)] [Medline: [12140364](https://pubmed.ncbi.nlm.nih.gov/12140364/)]
35. Sumner J, Uijtdewilligen L, Yee ACH, Xian SNH, Barreira TV, Sloan RA, et al. Volume and Intensity of Stepping Activity and Cardiometabolic Risk Factors in a Multi-ethnic Asian Population. *Int J Environ Res Public Health* 2020 Jan 30;17(3):E863 [FREE Full text] [doi: [10.3390/ijerph17030863](https://doi.org/10.3390/ijerph17030863)] [Medline: [32019086](https://pubmed.ncbi.nlm.nih.gov/32019086/)]
36. Schuna JM, Brouillette RM, Foil HC, Fontenot SL, Keller JN, Tudor-Locke C. Steps per day, peak cadence, body mass index, and age in community-dwelling older adults. *Med Sci Sports Exerc* 2013 May;45(5):914-919. [doi: [10.1249/MSS.0b013e31827e47ac](https://doi.org/10.1249/MSS.0b013e31827e47ac)] [Medline: [23247705](https://pubmed.ncbi.nlm.nih.gov/23247705/)]
37. Buckee CO, Balsari S, Chan J, Crosas M, Dominici F, Gasser U, et al. Aggregated mobility data could help fight COVID-19. *Science* 2020 Apr 10;368(6487):145-146. [doi: [10.1126/science.abb8021](https://doi.org/10.1126/science.abb8021)] [Medline: [32205458](https://pubmed.ncbi.nlm.nih.gov/32205458/)]
38. Hajna S, Ross NA, Dasgupta K. Steps, moderate-to-vigorous physical activity, and cardiometabolic profiles. *Prev Med* 2018 Feb;107:69-74 [FREE Full text] [doi: [10.1016/j.ypmed.2017.11.007](https://doi.org/10.1016/j.ypmed.2017.11.007)] [Medline: [29126915](https://pubmed.ncbi.nlm.nih.gov/29126915/)]
39. Dwyer T, Ponsonby A, Ukoumunne OC, Pezic A, Venn A, Dunstan D, et al. Association of change in daily step count over five years with insulin sensitivity and adiposity: population based cohort study. *BMJ* 2011 Jan 13;342(jan13 1):c7249-c7249. [doi: [10.1136/bmj.c7249](https://doi.org/10.1136/bmj.c7249)] [Medline: [21233153](https://pubmed.ncbi.nlm.nih.gov/21233153/)]
40. Yates T, Haffner SM, Schulte PJ, Thomas L, Huffman KM, Bales CW, et al. Association between change in daily ambulatory activity and cardiovascular events in people with impaired glucose tolerance (NAVIGATOR trial): a cohort analysis. *Lancet* 2014 Mar 22;383(9922):1059-1066. [doi: [10.1016/S0140-6736\(13\)62061-9](https://doi.org/10.1016/S0140-6736(13)62061-9)] [Medline: [24361242](https://pubmed.ncbi.nlm.nih.gov/24361242/)]
41. Varona-Pérez P, Pérez-Jiménez D, Alfonso-Sagué K, García-Pérez RM, Bonet-Gorbea M, Fernández-González J. Patterns of Physical Activity and Associated Factors in Cubans Aged 15-69 Years. *MEDICC Rev* 2016 Oct;18(4):20-25. [Medline: [27829650](https://pubmed.ncbi.nlm.nih.gov/27829650/)]
42. Shephard RJ, Park H, Park S, Aoyagi Y. Objective Longitudinal Measures of Physical Activity and Bone Health in Older Japanese: the Nakanojo Study. *J Am Geriatr Soc* 2017 Apr;65(4):800-807. [doi: [10.1111/jgs.14553](https://doi.org/10.1111/jgs.14553)] [Medline: [27943243](https://pubmed.ncbi.nlm.nih.gov/27943243/)]
43. Duncan MJ, Wunderlich K, Zhao Y, Faulkner G. Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. *J Sports Sci* 2018 Aug;36(15):1695-1704. [doi: [10.1080/02640414.2017.1409855](https://doi.org/10.1080/02640414.2017.1409855)] [Medline: [29179653](https://pubmed.ncbi.nlm.nih.gov/29179653/)]
44. Fuller D, Colwell E, Low J, Orychok K, Tobin MA, Simango B, et al. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. *JMIR mHealth uHealth* 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: [10.2196/18694](https://doi.org/10.2196/18694)] [Medline: [32897239](https://pubmed.ncbi.nlm.nih.gov/32897239/)]
45. The 46th China Statistical Report on Internet Development. CNNIC. URL: [https://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwjbg/202009/t20200929\\_71257.htm](https://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwjbg/202009/t20200929_71257.htm) [accessed 2020-09-30]

## Abbreviations

**CCI:** Charlson Comorbidity Index  
**GEE:** Generalized Estimating Equations  
**ICC:** intraclass correlation coefficients  
**MD:** mean difference  
**NHANES:** National Health and Nutrition Examination Survey  
**PR:** prevalence ratio



*Edited by G Eysenbach; submitted 19.06.20; peer-reviewed by M Alharbi, H Sasai; comments to author 12.09.20; revised version received 01.10.20; accepted 02.10.20; published 26.10.20.*

*Please cite as:*

Wang Y, Zhang Y, Bennell K, White DK, Wei J, Wu Z, He H, Liu S, Luo X, Hu S, Zeng C, Lei G

*Physical Distancing Measures and Walking Activity in Middle-aged and Older Residents in Changsha, China, During the COVID-19 Epidemic Period: Longitudinal Observational Study*

*J Med Internet Res* 2020;22(10):e21632

URL: <http://www.jmir.org/2020/10/e21632/>

doi: [10.2196/21632](https://doi.org/10.2196/21632)

PMID: [33027035](https://pubmed.ncbi.nlm.nih.gov/33027035)

©Yilun Wang, Yuqing Zhang, Kim Bennell, Daniel Kenta White, Jie Wei, Ziyang Wu, Hongyi He, Shaohui Liu, Xianghang Luo, Shuo Hu, Chao Zeng, Guanghua Lei. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Investigating the Prevalence of Reactive Online Searching in the COVID-19 Pandemic: Infoveillance Study

Rafael A Badell-Grau<sup>1\*</sup>, BSc, MRes; Jordan Patrick Cuff<sup>1\*</sup>, BSc, MRes; Brendan P Kelly<sup>1</sup>, BS, MS; Helen Waller-Evans<sup>2</sup>, BSc, PGDip, DPhil; Emyr Lloyd-Evans<sup>1</sup>, MBiochem, DPhil

<sup>1</sup>School of Biosciences, Cardiff University, Cardiff, United Kingdom

<sup>2</sup>Medicine Discovery Institute, Cardiff University, Cardiff, United Kingdom

\*these authors contributed equally

**Corresponding Author:**

Emyr Lloyd-Evans, MBiochem, DPhil

School of Biosciences

Cardiff University

Sir Martin Evans Building

Museum Avenue

Cardiff, CF10 3AX

United Kingdom

Phone: 44 (0)29 2087 4304

Email: [lloyd-evanse@cardiff.ac.uk](mailto:lloyd-evanse@cardiff.ac.uk)

## Abstract

**Background:** The ongoing pandemic has placed an unprecedented strain on global society, health care, governments, and mass media. Public dissemination of government policies, medical interventions, and misinformation has been remarkably rapid and largely unregulated during the COVID-19 pandemic, resulting in increased misinterpretations, miscommunication, and public panic. Being the first full-scale global pandemic of the digital age, COVID-19 has presented novel challenges pertinent to government advice, the spread of news and misinformation, and the trade-off between the accessibility of science and the premature public use of unproven medical interventions.

**Objective:** This study aims to assess the use of internet search terms relating to COVID-19 information and misinformation during the global pandemic, identify which were most used in six affected countries, investigate any temporal trends and the likely propagators of key search terms, and determine any correlation between the *per capita* cases and deaths with the adoption of these search terms in each of the six countries.

**Methods:** This study uses relative search volume data extracted from Google Trends for search terms linked to the COVID-19 pandemic alongside *per capita* case and mortality data extracted from the European Open Data Portal to identify the temporal dynamics of the spread of news and misinformation during the global pandemic in six affected countries (Australia, Germany, Italy, Spain, the United Kingdom, and the United States). A correlation analysis was carried out to ascertain any correlation between the temporal trends of search term use and the rise of *per capita* mortality and disease cases.

**Results:** Of the selected search terms, most were searched immediately following promotion by governments, public figures, or viral circulation of information, but also in relation to the publication of scientific resources, which were sometimes misinterpreted before further dissemination. Strong correlations were identified between the volume of these COVID-19-related search terms (overall mean Spearman rho 0.753, SD 0.158), and *per capita* mortality (mean *per capita* deaths Spearman rho 0.690, SD 0.168) and cases (mean *per capita* cases Spearman rho 0.800, SD 0.112).

**Conclusions:** These findings illustrate the increased rate and volume of the public consumption of novel information during a global health care crisis. The positive correlation between mortality and online searching, particularly in countries with lower COVID-19 testing rates, may demonstrate the imperative to safeguard official communications and dispel misinformation in these countries. Online news, government briefings, and social media provide a powerful tool for the dissemination of important information to the public during pandemics, but their misuse and the presentation of misrepresented medical information should be monitored, minimized, and addressed to safeguard public safety. Ultimately, governments, public health authorities, and scientists have a moral imperative to safeguard the truth and maintain an accessible discourse with the public to limit fear.

**KEYWORDS**

chloroquine; coronavirus; COVID-19; fake news; Google Trends; ibuprofen; infodemiology; misinformation

**Introduction**

The COVID-19 pandemic has encouraged an unprecedented international panic. Since its emergence in late 2019 in the Hubei Province of China, COVID-19 has spread worldwide, and its associated infectivity and death rate have challenged world leaders, health care systems, and the public [1,2]. Unlike comparable previous pandemics, such as the Spanish flu in 1918, the internet has provided to the public a source of connectivity and a means to rapidly acquire emerging information about the virus [1]. The information available is, however, not always verifiable or scientifically supported.

The dissemination of government policy and cutting-edge medical research is unquestionably important in the remit of a global pandemic, but misinterpretation is commonplace. The desperation of the public encourages the opportunistic adoption of unverified medical interventions. The misuse and misrepresentation of such information presents a critical challenge to governments and to the public. Equally, the public may seek out and enable misinformation (eg, the virus being spread by 5G towers [3]), which is rapidly distributed via social media [4]. The increased dependence of the public on social media and other inherently biased sources of information may inflate the rate at which misinformation spreads, possibly fostering disenfranchisement with government and health care organizations [5-7]. This could ultimately provoke disregard toward restrictions enforced for public safety, lead to reduced supplies of medicines and personal protective equipment (PPE), or potentially even to reduced medical engagement and worsening of chronic conditions, increasing pressure on already strained health care providers.

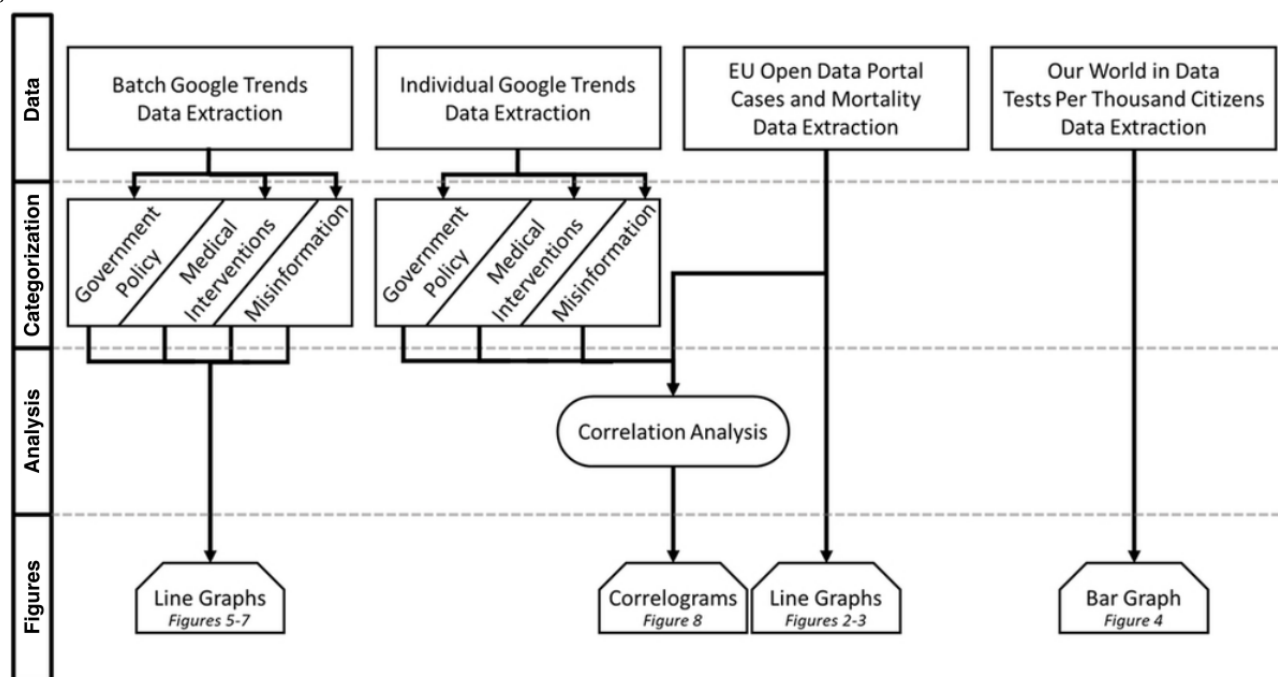
Given the rapid flow of digital information during the COVID-19 pandemic, real-time data collection and analysis provides an unparalleled opportunity to assess the public response to information as it emerges. Through internet-derived information from social media, news, and search engine use, public reactions and perceptions can be assessed in real time [4,8-11]. Google Trends (GT) has been used for not only the analysis of epidemiologically relevant data regarding influenza [9] and disease outbreaks [10] but also, more recently, COVID-19 [11]. By assessing the temporal dynamics of search terms related to the pandemic, particularly those relating to misinformation, it is possible to infer likely sources, propagators, and impacts. This study employs GT for the analysis of search terms used during the COVID-19 pandemic relating to

government policy, potential treatments, and misinformation, specifically in three English and three non-English speaking countries: Australia, the United Kingdom, the United States, Germany, Italy, and Spain. The aims of this study are to identify any correlation between the relative search volumes (RSVs) for information relating to the first wave of the pandemic, and to discuss these search volumes in the context of emerging news, alongside the prevalence of cases and deaths in each of the six focal countries.

**Methods****Mortality, Case, and Testing Data Extraction**

Worldwide mortality and case data, and country population sizes were extracted from the European Union Open Data Portal [12] on April 17, 2020 (Figure 1). Data were retained only for the six focal countries. Dates for which no data were available from November 1, 2019, to the first recorded numbers for that country were marked as zero. *Per capita* cases and deaths were also calculated using the included population sizes and retained for later analyses and figures. *Per capita* values, although not widely reported by the media at this time, were used in this study to correct for the large variation in population sizes of the focal countries, and to better represent the proportional pressure upon each country.

The objective reliability of these data is questionable given the internationally variable extent of testing and the resultant predicted inaccuracy of the case numbers in each country. International variations in the definition of COVID-19-related deaths and failures to report the full extent of case numbers also warrant skepticism. In the remit of this study, however, these data represent the immediate perceived threat and pressure elicited upon the societies of each focal country, thus providing a suitable comparison against the temporal dynamics of the search terms used. The numbers of COVID-19 tests per thousand citizens were downloaded from Our World in Data [13]; given the irregularity of testing and resultant unavailability of data for some countries, these data were not used for correlation analysis. The number of tests completed by April 17, 2020, was recorded, except for Germany and Spain for which values represented the tests per thousand completed by April 19 and 13, respectively, due to a lack of data for April 17. Testing data represent the number of tests performed, rather than the number of individuals tested, given the wider availability of these data; the nature of Australia's testing units is, however, unclear.

**Figure 1.** Data extraction and workflow.

### Search Volume Data Extraction

Data were extracted from GT on April 17, 2020, for the period of November 1, 2019, to April 17, 2020, which includes a brief period before the first confirmed case of COVID-19 for comparison. These data provide a proxy for public interest in government policy, emerging health care interventions, and misinformation, later contextualized as a response to the release of such information. The data extracted from GT are RSVs for predetermined search terms, allowing comparison of search rates for different terms via Google, the most widely used internet search engine, especially in the countries selected [14,15]. These RSVs are presented for each date of a given time period within a given country. Data are normalized relative to the highest RSV peak in that time period (this peak represented as 100).

Data were extracted for searches generated from Australia, Germany, Italy, Spain, the United Kingdom, and the United States. These countries were selected due to their widespread use of Google (precluding China and many other Asiatic countries), variation in the extent to which they were impacted by the pandemic, nuances in their responses to the pandemic, and the accessibility of their news and media in one predominant language. All search terms were preceded by “coronavirus” to ensure relevance to the pandemic; “coronavirus” was selected over “COVID19” and similar terms due to its greater prevalence of searches (eg, in the United States, “coronavirus vaccine” yielded four-fold the search volume of “corona vaccine” and “covid vaccine”, and twenty-fold that of “covid19 vaccine” and “covid-19 vaccine”).

All search terms were selected based on their widespread media coverage and their high Google search volumes. Their placement in the broad categories of *government policy*, *medical interventions*, and *misinformation* were based on the context of their wide reporting by media, government, research, and health

care organizations of those particular countries. The designation of search terms as *medical interventions* did not equate to their effectiveness in treating COVID-19 but scientific discussion around, or political endorsement of, their experimental or genuine use in treating the virus. Chloroquine, for example, was not empirically shown to benefit patients at the time of this study, and its early endorsement during the pandemic largely emanated from the United States, but international research nonetheless endeavored to ascertain any benefit it conferred to patients with COVID-19, this being the primary focus of its initial widespread news coverage. Misinformation search terms were labeled as such when there was no empirical evidence nor active published peer-reviewed research regarding their relevance to COVID-19, and their media coverage indicative of their potential for controversy; such search terms could often be traced back to an initial misinterpretation or false statement, some of which are highlighted in the discussion. All terms were identified as COVID-19 misinformation by Dhillon et al [16]. Other search terms relevant to COVID-19 were considered but for a contained and meaningful statistically significant comparison only those with relatively high and comparable RSVs within the three aforementioned categories were included. Search terms for which variations were possible (eg, chloroquine vs hydroxychloroquine) were included as the variation with the greatest GT search volume with the simpler terminology routinely having the greatest search volume.

Searches were carried out in the language native to each country unless the English terms provided a greater number of results (ie, where English phraseology was adopted). Searches were carried out in batches to identify relative differences in search volumes, with three batches coarsely defined as “government policies,” “medical interventions,” and “misinformation.” All search batches contained “coronavirus chloroquine” as a standard to facilitate some comparison between categories given its relatively central positioning in most batches. Chloroquine was selected for its relatively average search volume across

countries and categories, acting as an anchor to facilitate visual comparison between higher and lower RSVs. The search term RSVs were all also individually downloaded (independently normalized with the highest peak being 100) for subsequent correlation analysis to evenly represent the extent of searching and focus on the temporal dynamics. Given the representation of numbers less than one as “<1” by GT, all RSVs of “<1” were converted to 0.5 to facilitate quantitative comparison.

The government policy search terms comprised chloroquine (control standardization term), social distancing, sanitizer, mask, isolation, gloves, and testing (Textbox 1). Social distancing was implemented by many countries as an early and maintained means to prevent viral spread, as was isolation, although the latter may also have been searched in association with the well-being and mental health consequences of reduced social contact during lockdown. The use of sanitizer for cleansing of

hands was also encouraged by governments throughout the pandemic, although depleting public availability in most countries led many to attempt to create homemade sanitizer [17]. Masks and gloves were employed as a protective means to prevent spread, although predominantly by frontline health care workers; public purchase of this PPE was problematic in many countries, resulting in reduced availability for medical practitioners [18,19]. Testing and tracing was carried out for coronavirus, but the extent of testing and the national focus on its importance varied internationally [13]. The US spelling of “sanitizer” was maintained for the UK searches given a higher prevalence than the UK spelling “sanitiser.” Due to the GT search limit, the government policy search was split into two batches (batch 1: chloroquine, social distancing, sanitizer, mask, and isolation, and batch 2: chloroquine, gloves, and testing, with linguistic variations for Germany, Italy, and Spain).

**Textbox 1.** Google Trends search terms used in each of the three categories.

#### Government policy

- Australia, the United Kingdom, and the United States: social distancing, sanitizer, mask, isolation, gloves, testing
- Germany: social distancing, desinfektion-smittel, maske, isolation, handschuhe, testen
- Italy: distanziamento sociale, disinfettante, maschera, isolamento, guanti, analisi
- Spain: distanciamiento social, desinfectante, mascara, aislamiento, guantes, pruebas

#### Medical interventions

- Australia, the United Kingdom, and the United States: chloroquine, remdesivir, paracetamol, vaccine, ibuprofen
- Germany: chloroquin, remdesivir, paracetamol, impstoff, ibuprofen
- Italy: chlorochina, remdesivir, paracetamolo, vaccino, ibuprofene
- Spain: cloroquina, remdesivir, paracetamol, vacuna, ibuprofeno

#### Misinformation

- Australia, the United Kingdom, and the United States: 5G, man made, lab
- Germany: 5G, hergestellt, labor
- Italy: 5G, creato, laboratorio
- Spain: 5G, creado, laboratorio

The medical intervention search terms comprised chloroquine (control standardization term), remdesivir, paracetamol, vaccine, and ibuprofen (Textbox 1). All of these search terms pertain to treatments that were suggested to have potential effects against COVID-19 symptoms. The public focus on vaccines reflected the ongoing development of vaccines and the desire for relief from the pandemic [20]. Paracetamol and ibuprofen were used to subdue pain associated with COVID-19 symptoms, but public perception became antagonistic toward using ibuprofen for COVID-19 symptoms, which shifted focus toward paracetamol [21].

The misinformation search terms comprised chloroquine (control term), 5G, man made, and lab (Textbox 1). These search terms pertain to internationally prevalent misinformation related to COVID-19, often specifically suggesting a disingenuous cause or source of the viral spread. Specifically, these entail theories that the virus was being spread by the new 5G phone masts, that the virus was manufactured, and that the virus was released

from a laboratory [3,16,22-24], all of which have subsequently been debunked [25-27].

#### Statistical Analysis

Statistical analyses and plotting of data were carried out using R version v4.0.0 (R Foundation for Statistical Computing) [28]. Line graphs were created for *per capita* cases and deaths, and a bar chart for tests per thousand citizens using *ggplot* in the *ggplot2* package version 3.3.0 in R [29], with colors assigned via the *RColorBrewer* package v1.1-2 [30]. The data were identified as nonnormally distributed via Shapiro–Wilk tests, so nonparametric statistical analyses were selected. Correlations between RSVs and *per capita* deaths and cases were tested using Spearman rho rank correlation via the *rcor* function of the *Hmisc* package version 4.4-0 [31]. The output was then presented in a correlogram via the *corrplot* function of the *corrplot* package version 0.84 [32], with colors assigned via the *viridis* package v0.5.1 [33]. Line graphs were created for each of the three



categories of search terms for each country to aid comparison of both the extent and temporality of RSV trends in GraphPad Prism version 8 (GraphPad Software) [34]. All statistical data is included in [Multimedia Appendix 1](#).

### Information Sources and Reliability

The sources for the non-search term data (The European Union Open Data Portal and Our World in Data) are reputable sources that derive their data from official national reports, scientific publications, and other reliable sources. The data extracted from these sources align with those published internationally in response to the pandemic situation as it develops. The GT data are collected and presented by Google based on the input of users of their service, thus should be fully reliable. Although most sources cited in this report are from reputable scientific, government, or public health authority sources, others discussed throughout the manuscript are taken from mass media, social media, and other heavily biased sources or from scientific articles that discuss such sources; these sources are being referred to on the basis of these biases or simply to refer to the temporal development and emergence of global news, for which bias is an important factor. The paper discusses the reporting of this information in an objective manner, with no subscription to the reported ideals or beliefs represented in the text.

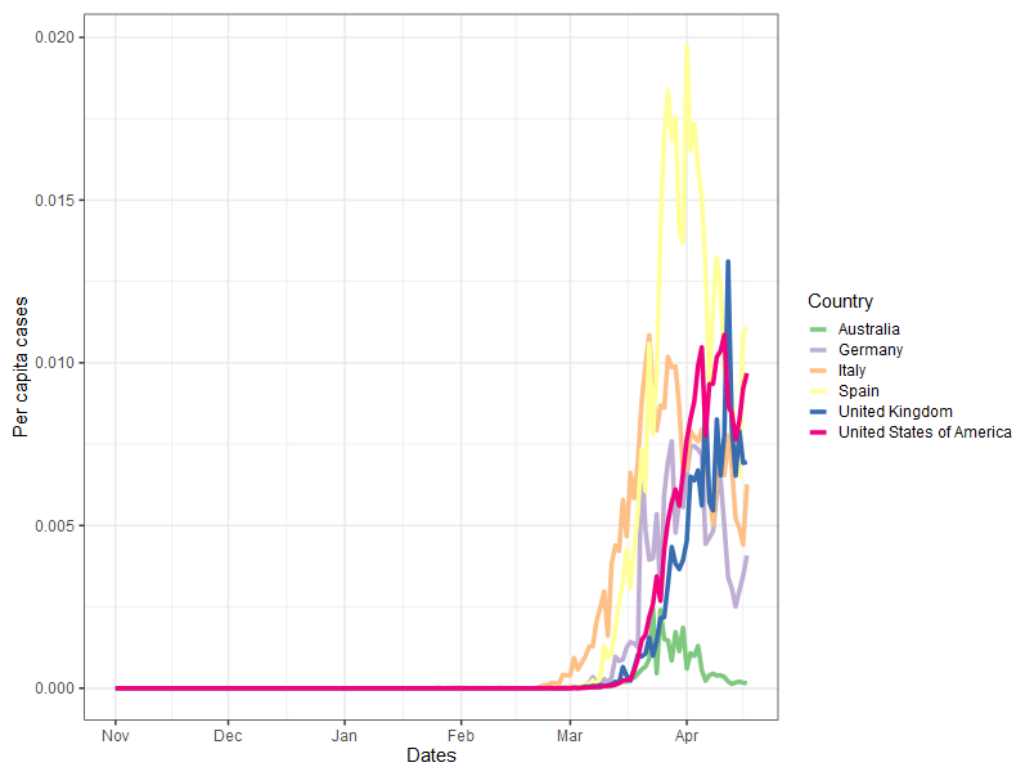
## Results

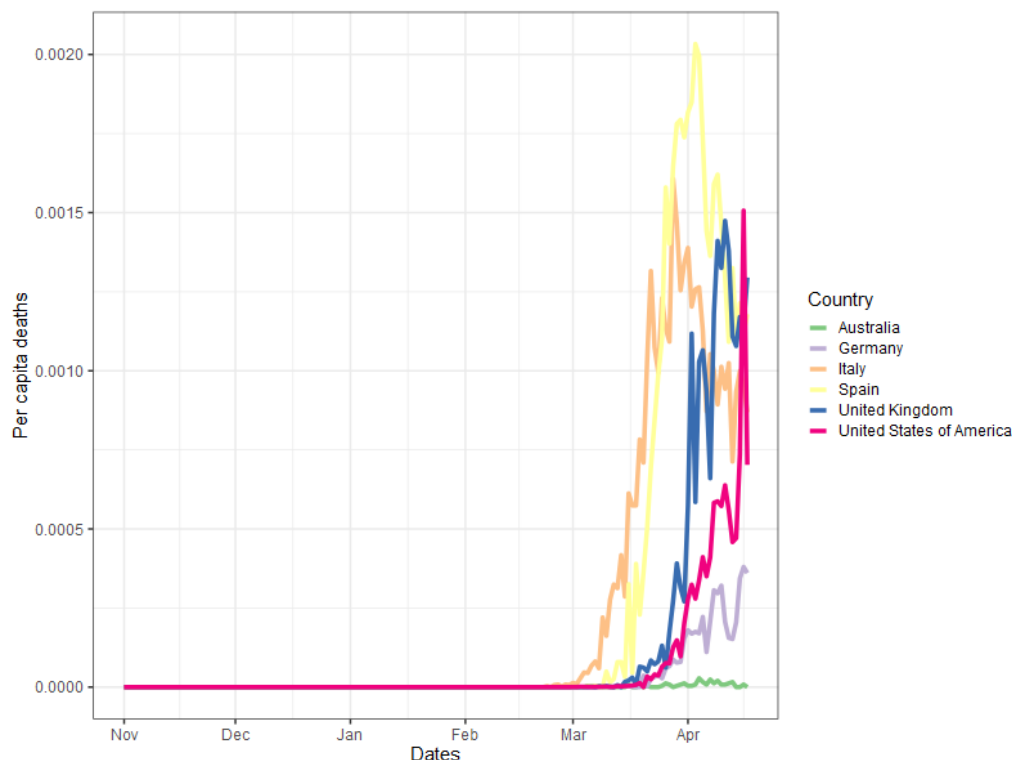
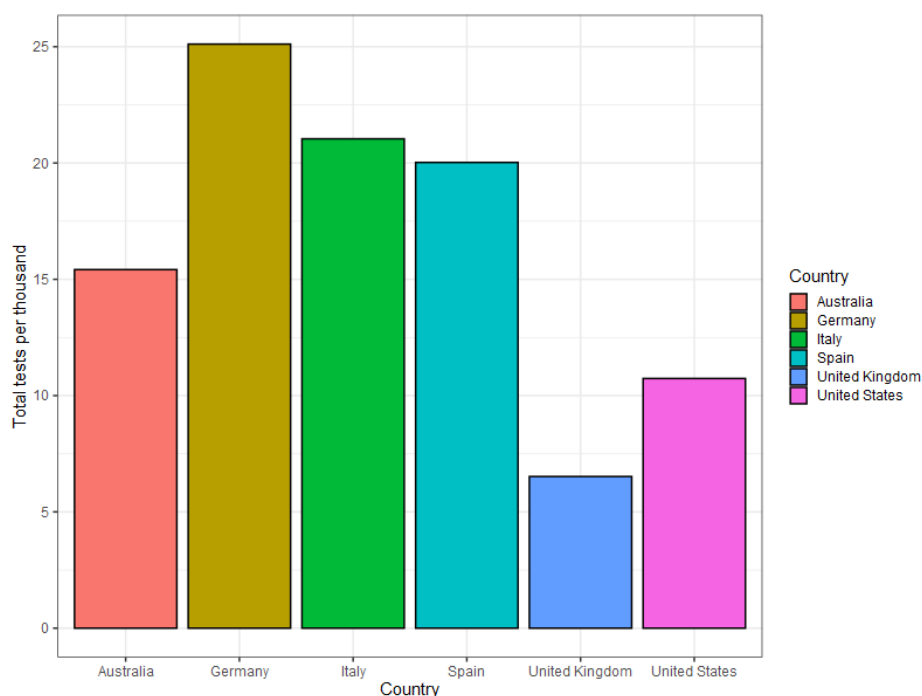
### Mortality, Case, and Test Results

All countries show similar *per capita* case (Figure 2) and death (Figure 3) trends temporally, with both beginning to

exponentially increase in most countries between late February and early March. Of the six countries, Italy was the first to present a substantial number of cases and deaths (mid-February 2020). Australia, the United Kingdom, and the United States were the last to experience rapidly increasing *per capita* case numbers (~March 10). The *per capita* case number trends are relatively similar in their extents for most countries, except for Spain, which exhibits approximately 50% more peak *per capita* cases than the second highest peak, the United Kingdom (Spain: 0.01937, UK: 0.013113), and Australia, which exhibits approximately a quarter of the peak *per capita* cases for the majority of the countries (Australia: 0.002445, average for other countries, excluding Spain: 0.010603, SD 0.0023). The *per capita* deaths similarly increase last for Australia, the United Kingdom, the United States, and Germany. Germany and the United States display a shallower trajectory of *per capita* death increases, and Australia shows a minor peak of *per capita* deaths. Spain again exhibits the greatest peak of *per capita* deaths, but only with an approximate 30% increase over the peaks of Italy, the United Kingdom, and the United States (Spain: 0.002033, Italy: 0.001607, the United Kingdom: 0.001474, the United States: 0.001506), compared to the ~50% increase over the second highest peak for *per capita* cases. Testing for COVID-19 varied massively between countries, with Germany showing the highest tests per thousand, with around 25 tests per thousand, and the United Kingdom showing the lowest with around 6.5 tests per thousand (Figure 4).

**Figure 2.** Per capita cases of COVID-19 during the study period.



**Figure 3.** Per capita COVID-19–related deaths during the study period.**Figure 4.** Total COVID-19 tests per thousand citizens in the six focal countries, as of April 17, 2020, except for Germany and Spain, which are represented by April 19 and 13, respectively, due to a lack of data for April 17.

## Search Volume Results

Of the government policy search terms (Figure 5), “testing” was prevalent in all countries, and “isolation” relatively high in all but the United States and Germany. “Sanitizer” was highly searched in Germany, Italy, and Spain. “Masks” was highly searched in Australia, Germany, Italy, and the United States, and to a lesser extent the United Kingdom. “Gloves” was

searched less in all but Italy and Spain. “Social distancing” was searched less except in Australia and the United Kingdom, where this term was the third most searched. Most search terms peaked at a similar time (mid-March) in most countries, although “mask” also peaked in late January to early February in Australia, the United Kingdom, and the United States, and to a lesser extent in Germany and Italy. In Germany and Italy, “sanitizer” and “mask” peaked in early March, 2-3 weeks earlier

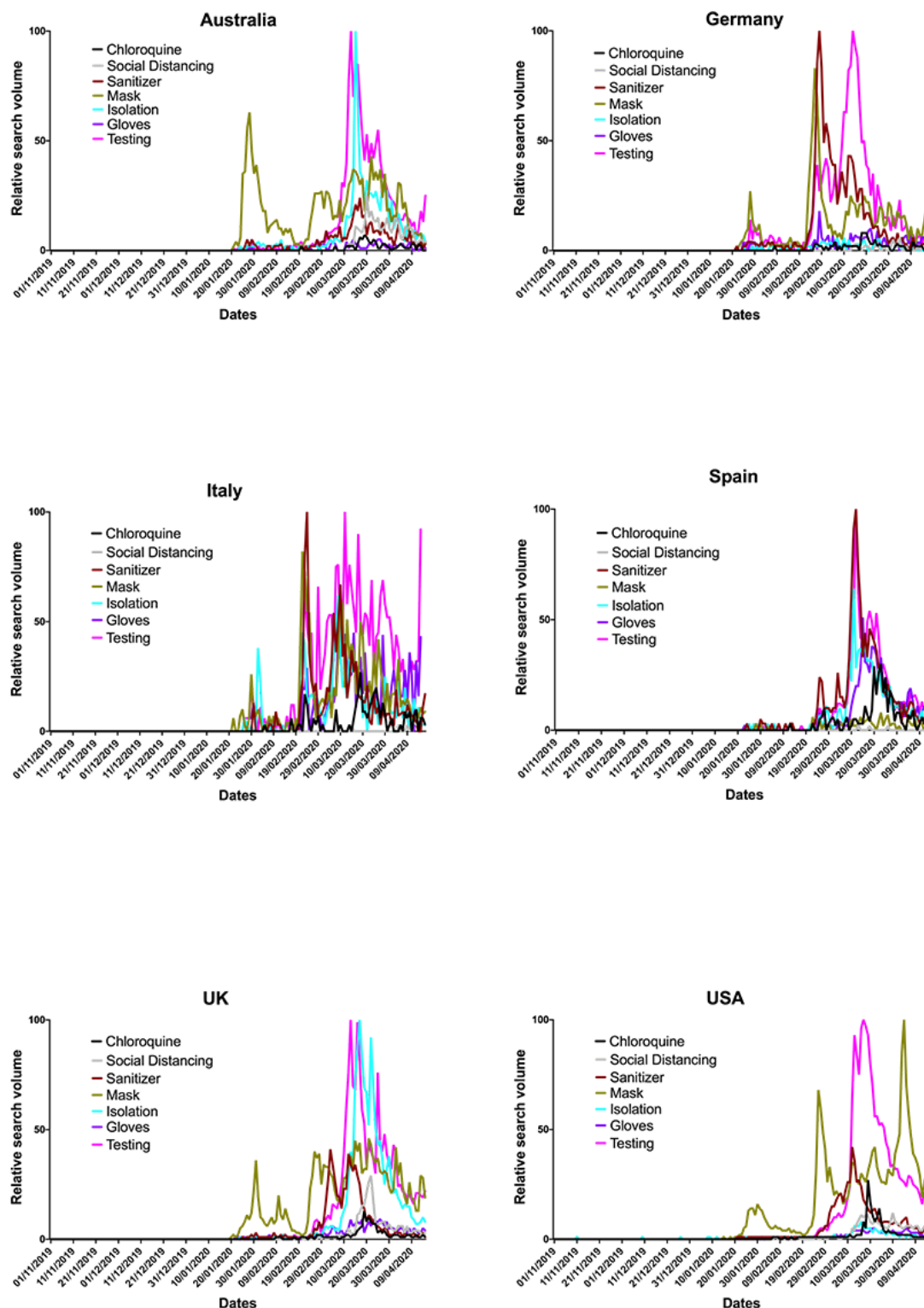
than a later peak coinciding with that of “testing” in other countries. In the United States, “gloves” was searched most at the end of February, but also with a second peak in early April, unlike the other countries. In Italy, searches of “testing” peaked sporadically from late February to mid-April (the end of the search period), with larger peaks spread further across the period.

Of the medical intervention search terms (Figure 6), “vaccine” was highly searched in all countries, peaking in late March, except in Germany, where it peaked in late February, and Italy, where it peaked sporadically from the end of January to mid-April (the end of the search period). In the United Kingdom, “vaccine” had a second peak in mid-April. The other medical interventions had relatively small peaks, often in mid- or late March. “Remdesivir” peaked higher in Italy relative to the other countries. “Chloroquine” peaked much higher in the United States relative to the other countries, also having a smaller peak in the United Kingdom. “Ibuprofen” had the highest peak in

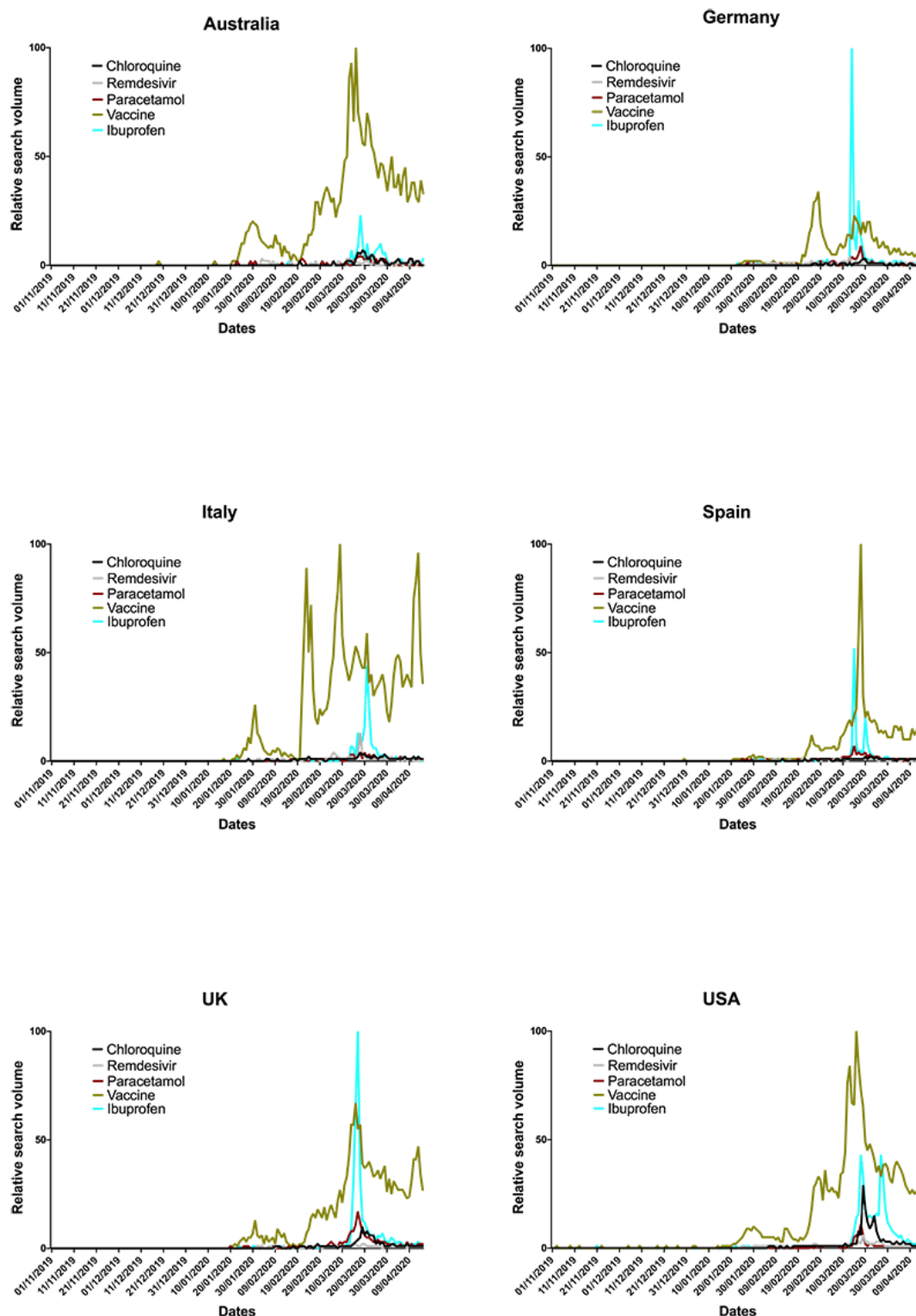
Germany and the United Kingdom, peaking in all countries in mid- or late March and having a second peak in early April in the United States.

Of the misinformation search terms (Figure 7), “5G” had erratic smaller peaks throughout mid- and late March but peaked in most countries in early April, with Germany and Spain displaying reduced peaks. “Man made” was mostly searched in mid-March, with some more widespread erratic peaks in all but the United Kingdom, and a substantial peak in late January and early February in Australia, Italy, and Spain. “Lab” was searched relatively little in Australia, the United Kingdom, and, to some extent, the United States. “Lab” was, however, highly searched in Italy and Spain in late March, with Italy also exhibiting large peaks in late January and late February, and was searched at similar intervals in Germany but never so proportionally high as Italy and Spain. In most cases, peaks of “lab” coincide with peaks of “man made.”

**Figure 5.** Government policy relative search volumes (RSVs) extracted from Google Trends (GT). Grouped RSV data, normalized to the highest RSV peak in the time period (represented as 100) were extracted from GT on April 17, 2020, for the period of November 1, 2019, to April 17, 2020. Search terms included “coronavirus chloroquine” (control term), “coronavirus social distancing,” “coronavirus sanitizer,” “coronavirus mask,” “coronavirus isolation,” “coronavirus gloves,” and “coronavirus testing,” with variations to reflect the language native to each country.

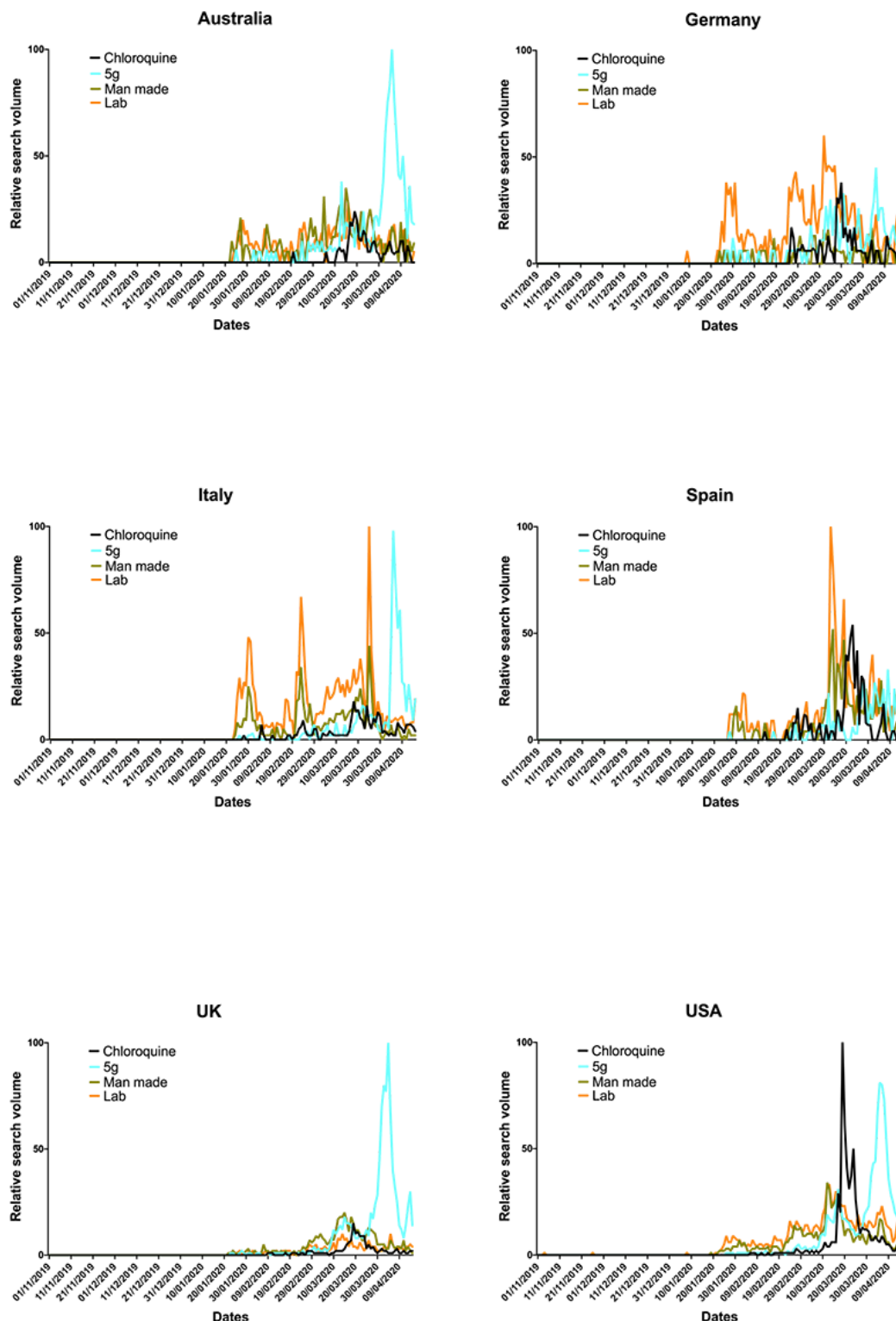


**Figure 6.** Medical intervention relative search volumes (RSVs) extracted from Google Trends (GT). Grouped RSV data, normalized to the highest RSV peak in the time period (represented as 100) were extracted from GT on April 17, 2020, for the period of November 1, 2019, to April 17, 2020. Search terms included “coronavirus chloroquine,” “coronavirus remdesivir,” “coronavirus paracetamol,” “coronavirus vaccine,” and “coronavirus ibuprofen,” with variations to reflect the language native to each country.





**Figure 7.** Misinformation relative search volumes (RSVs) extracted from Google Trends (GT). Grouped RSV data, normalized to the highest RSV peak in the time period (represented as 100) were extracted from GT on April 17, 2020, for the period of November 1, 2019, to April 17, 2020. Search terms included “coronavirus chloroquine,” “coronavirus remdesivir,” “coronavirus paracetamol,” “coronavirus vaccine,” and “coronavirus ibuprofen,” with variations to reflect the language native to each country.



## Correlation Analysis Results

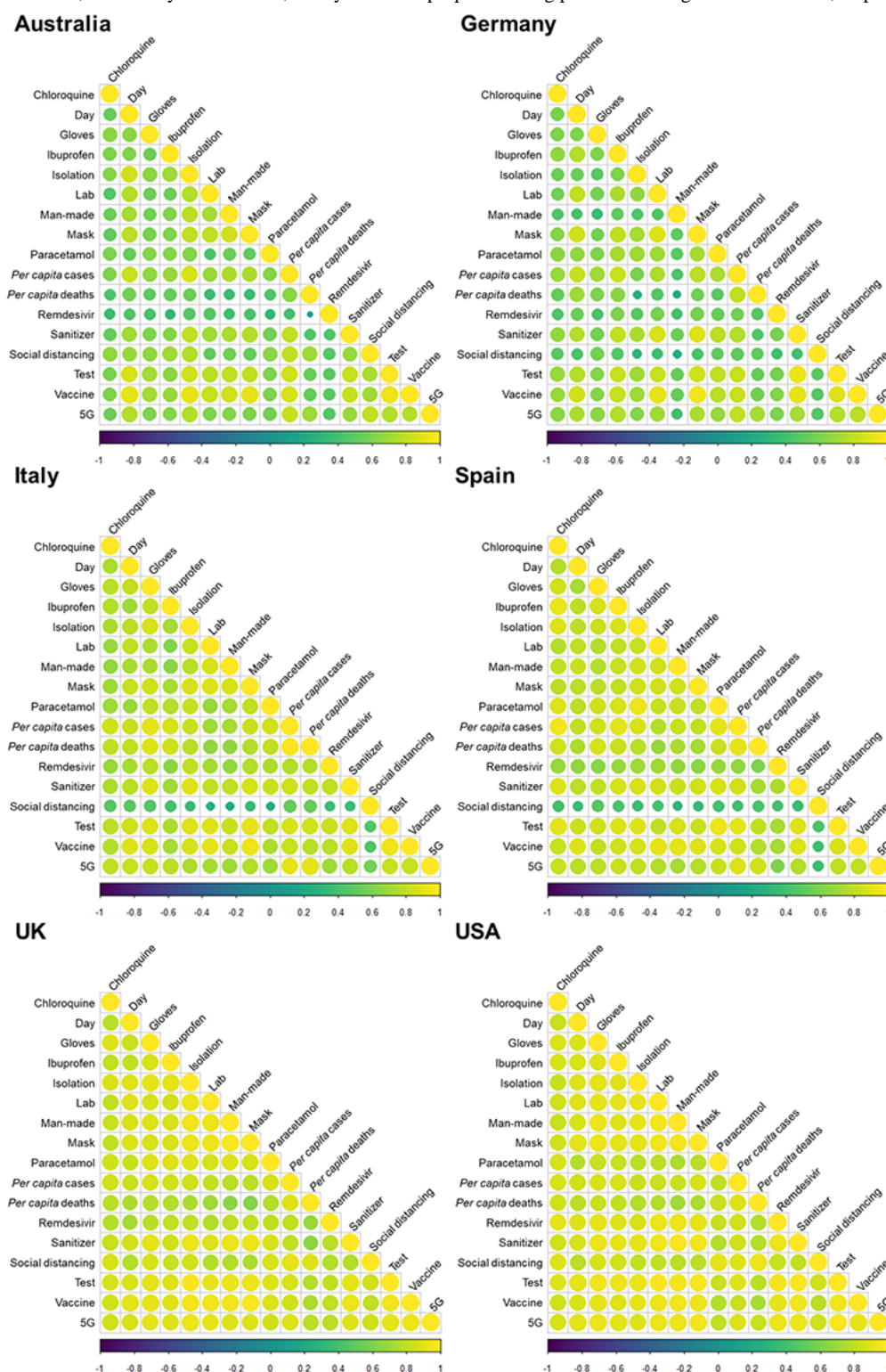
In all countries, almost all normalized search terms significantly positively correlated with one another (overall mean Spearman rho 0.753, SD 0.158) and *per capita* deaths (mean *per capita* deaths Spearman rho 0.690, SD 0.168) and cases (mean *per capita* cases Spearman rho 0.800, SD 0.112; [Figure 8](#), [Table 1](#),

and [Multimedia Appendix 1](#) Table S1); the only exception was the nonsignificant association between *per capita* deaths and remdesivir RSV in Australia (Spearman rho 0.134,  $P=.08$ ). Overall, stronger correlations were identified more universally for the United Kingdom (mean Spearman rho 0.851, SD 0.066) and the United States (mean Spearman rho 0.873, SD 0.058), while relatively weaker correlations were shown for Australia

(mean Spearman rho 0.641, SD 0.150) and Germany (mean Spearman rho 0.632, SD 0.157; Figure 8, Table 1, and Multimedia Appendix 1 Table S1). In Italy and Spain, the

weakest correlations were those between social distancing and all other variables.

**Figure 8.** Correlograms for the search factors and per capita case and death rates for each country. The size of each circle indicates the strength of the correlation as does the color, denoted by the scale bar, with yellow and purple denoting positive and negative correlations, respectively.



**Table 1.** Mean Spearman rho rank coefficients and their standard deviations are given for each country and overall results for all six countries.

Countries	Overall correlation, mean (SD)	Time correlation, mean (SD)	Cases correlation, mean (SD)	Deaths correlation, mean (SD)
All	0.753 (0.158)	0.769 (0.122)	0.800 (0.112)	0.690 (0.168)
Australia	0.641 (0.150)	0.701 (0.136)	0.732 (0.122)	0.495 (0.146)
Germany	0.632 (0.157)	0.681 (0.143)	0.719 (0.123)	0.535 (0.153)
Italy	0.753 (0.147)	0.772 (0.093)	0.819 (0.103)	0.796 (0.104)
Spain	0.766 (0.152)	0.764 (0.120)	0.826 (0.133)	0.762 (0.116)
UK	0.851 (0.066)	0.835 (0.052)	0.846 (0.037)	0.750 (0.075)
US	0.873 (0.058)	0.861 (0.047)	0.858 (0.024)	0.802 (0.056)

## Discussion

This study aims to identify any correlation between the internet searching of defined COVID-19–relevant search terms and the *per capita* cases and deaths in six countries. We identified a positive correlation between the cases and deaths relating to COVID-19 and online searches surrounding government policies, medical interventions, and scientific misinformation.

### Principal Results

Between November 1, 2019 and April 17, 2020, *per capita* deaths and cases showed a similar trend across the six countries, with all having reached or passed peak daily new cases during the first wave of the pandemic. However, Australia and Germany experienced fewer deaths during this time period, allowing for a direct comparison of the search trends across countries with high and low COVID-19 cases and deaths. Where the ratio of mortality to cases is higher, such as the United Kingdom, which had the highest excess deaths in Europe during this period [35], this could reflect strained health care provision, delayed or reduced effectiveness of preventative measures, a poorer testing effort, or a combination of all of these [36,37]. Disparity in testing across countries may also have exacerbated differences in mortality. The importance of testing is illustrated by its high RSV across all countries (Figure 5) and the finding that the greatest degree of testing (Germany) aligns with relatively low mortality and weak correlations between RSVs and caseload. Where testing and contact tracing have been employed (eg, Germany, South Korea), they have been undoubtedly effective in mitigating increases in cases and deaths [38,39], possibly leading to an increased media and public interest, predominantly, it seems, in countries where it is lacking.

Overall, stronger correlations were observed in the United Kingdom and the United States. The English-speaking majority of these countries could explain this, given the widespread use of English on social media and in international news. In direct contrast, Australia had some of the weakest overall correlations; the combined low *per capita* deaths and cases, and the earlier application of travel restrictions and a 2 week quarantine [40,41] may have fostered a greater sense of safety and, therefore, less need by individuals to focus on the pandemic, evidenced by reduced interest in medical interventions. The overall strength of correlations being weakest in Australia and Germany, where the case and death figures are lower, supports the association

between reduced public pressure and a less coordinated uptake of news and misinformation.

That, in almost all cases, *per capita* deaths and cases correlated with the search term RSVs further suggests a strong relationship between the pressure elicited upon the public and their receptibility to pandemic-related digital information. The virus was internationally recognized and regularly reported by January [2], with many of the proposed preventative measures and medical interventions being widely searched online before cases and deaths began to emerge (Figures 5 and 6). The peak of most RSVs in mid-March, aligning approximately with peak *per capita* deaths and cases (Figures 1 and 2), also coincides with the beginning of lockdown in many countries [42], suggesting that populations were well informed prelockdown and ready for substantial changes to living conditions. The more dramatic peaks of search term RSVs following the beginning of March may denote the public searching news-relevant topics in far greater volume due to their willingness to follow government guidance, increased anxiety, and free time. In Italy, however, RSV peaks arrived earlier, likely due to the earlier arrival of the virus. The later peaks, which are often larger, may be propagated by greater exposure to mainstream and social media while at home and increased levels of anxiety (Multimedia Appendix 1 Figure S1) [43,44] thus creating a “second wave.” This relatively erratic persistent search behavior, particularly surrounding misinformation, could indicate heightened public panic especially as *per capita* deaths increase.

The data in this study highlights the utility of infoveillance in assessing public readiness for and adoption of preventative measures. The early interest in masks observed in the United States and Australia could indicate a willingness for, or pre-emptive fear of, the use of PPE. Despite some antimask sentiment in politicians [45] and possible reluctance by governments to impose mask-wearing for fear of appearing dictatorial, the public may be more prepared for discourse surrounding PPE than expected given the high RSVs. Conversely, social distancing consistently correlated weakly with other search terms, specifically in Germany, Italy, and Spain, despite all three countries entering nationwide lockdowns and observing government-mandated social distancing rules. Given the use of translated search terms, where these received more searches than the English equivalent, this is unlikely to be due to linguistic differences, despite these comprising only the non-English-speaking countries. In some countries, strict enforcement of social distancing may not have been necessary



due to greater compliance with guidelines (Germany). Alternatively, social distancing may not have been so heavily emphasized or adhered to in some countries, resulting in government-enforced curfews with fines for noncompliance, as experienced in Italy [46,47]. Regardless, clear and repeated guidance should be provided by governments to ensure compliance by their citizens. Good government response has been credited with the rapid reduction of lockdown measures in some countries, but such responses need to be data driven [38,48], and GT can provide an effective proxy for the extent of public adherence to this guidance.

Public interest in medical interventions was similarly moderately consistent between countries, with ibuprofen and chloroquine being the most searched. Some of this search intensity likely arose from misinformation, for example, the high RSVs for ibuprofen coincided with a scientific correspondence to *The Lancet* hypothesizing a heightened risk to a subset of patients with hypertension and diabetes should they take ibuprofen to combat COVID-19 [21]. This correspondence became misrepresented on messaging platforms and in media [16,49] as “evidence” that ibuprofen worsened COVID-19 symptoms. Furthermore, a second ibuprofen RSV peak in April in the United States coincided with a viral social media message claiming that patients with COVID-19 using ibuprofen did not recover [16]. The European Medicines Agency and the US Food and Drug Administration quickly discredited this as misinformation, possibly explaining the ephemerality of the RSV peak [25,26]. Paracetamol was highly searched simultaneously with ibuprofen, suggesting that people were seeking alternatives [50]. That the ibuprofen RSV comprises the highest medical intervention search peak in the United Kingdom and Germany, and a relatively high peak in other countries, compared to lower RSVs for experimental COVID-19 disease-modifying drugs such as remdesivir [51], confirms the capacity of misinformation to penetrate the public consciousness. Although this may also reflect the less familiar names and scientific background of the experimental drugs. This is further evidenced by the much larger RSVs for *vaccine* across all countries, a term familiar with most people, yet a therapeutic option that is clearly much further from public availability than therapies such as remdesivir [52]. The second peak of interest in vaccines in the United Kingdom was likely propagated by UK media reporting the initiation of clinical trials at the University of Oxford [53]. It is worth noting, however, that one experimental drug, namely, chloroquine, was searched with far greater intensity in the United States. This is likely due to US government briefings that supported chloroquine as a potential treatment for COVID-19 [54] based on a small clinical study [55], which led to multiple larger studies that ultimately did not support the outcomes [55,56] with most trials now suspended as reviewed in [57] and following some reports of accidental self-poisoning [58]. The important role of clear guidance from government is further exemplified from the suggestion during US government briefings that consideration should be given to the internal use of disinfectant and UV light in combating COVID-19. This is a clear example of misinformation arising from misinterpreted scientific literature that led to widescale panic and increased calls to poison centers [59-61].

Similarly, mass media and elected representatives have also propagated theories that SARS-CoV-2 is either man-made or was leaked from a laboratory in Wuhan. A quickly retracted scientific preprint appeared to propagate this theory by providing it an undue sense of credibility [24]. Although the man-made theory was scientifically discredited [27], public discussion moved toward a “leak” of the virus [16,23], highlighting the evolution and adaptability of misinformation, especially when supported by public figures [62]. Editors, reviewers, and authors should maintain stringent safeguards to ensure appropriate publishing, even of preprints, especially regarding such sensitive topics [63]. Similar conspiracy theories with large RSVs (Figure 7) arose via mainstream and social media outlets suggesting the spread of COVID-19 by 5G towers. The theory itself was in early circulation and despite being discredited as misinformation in January, long before the search intensity peaked [22], it led to vigilante attacks on phone masts and engineers in uninformed attempts to arrest viral spread [3,16,22]. The danger of misguided intervention led by misinformation outlines a clear requirement for mechanisms to reduce the spread of, while rationally and widely discrediting, these theories via perceivably credible sources such as national governments or professional medical bodies [1]. That the search volume surrounding 5G and ibuprofen dissipated so rapidly after documented attempts made by public health authorities such as the World Health Organization to curb the spread of this misinformation [25,26] best illustrates this point. It is, therefore, clear that during this pandemic the consumption of mass media, social media, government announcements, and health organization releases has influenced the public perception around both the causes and treatments of COVID-19 and, as perhaps best evidenced by the high RSVs for ibuprofen and 5G, has contributed to both public panic and health care issues such as reduced stocks of essential medicines caused by stockpiling [50].

### Limitations of GT Data

This study used GT data for six countries in which it is the most popular, but not the only, internet search engine. However, as the most widely used, it provides the best snapshot of user searches so that appropriate statistical studies can be conducted. As with any searches, the data presented in this study do not confirm subscription of those searching the terms to the ideals, interventions, or policies that they represent; many of the queries that contribute to these data may have been submitted by critics and sceptics. Even such searches, however, ratify the increased public awareness, discussion, and spread of the information denoted by the search terms. A greater volume of people reached by the information will undoubtedly suggest a greater number subscribing to the theories and ideas. The progression of a global pandemic is incredibly complicated and unpredictable, and the findings of this study focus on GT data from just one time period in a currently ongoing situation. Although this study bears relevance primarily to the beginning of the pandemic, this is arguably the most critical point at which to limit spread; however, the findings may not prove as relevant to periods when the public have adjusted to the situation.

## Conclusions

Infoveillance has already proven to be a valuable tool during the COVID-19 pandemic through detection of novel symptoms [11], assessment of behaviors such as self-medication [64], and identification of outbreaks [65]. This study focuses on the public response during the early developing pandemic, particularly surrounding misinformation, government policy, and medical interventions.

A study exploring the use of GT for digital epidemiology found that search term RSVs were influenced far more by media clamor than by epidemiological burden [66]. This pandemic is unique in that rapidly emerging medical research deposited in preprint archives has been accessible and consumed by the media and public pre-peer review, leading to potentially dangerous misinterpretation, as has occurred with chloroquine [58]. Although our findings ratify this, we also identified a

positive correlation between internet searching and COVID-19 deaths and cases, indicating a more synergistic combined effect of epidemiological burden and media attention. The prevalence and online spread of misinformation has been reported previously for COVID-19 [1] with regard to social media platforms, and, as in this study, the findings ultimately identified an important role for public health organizations and governments in providing accessible online information and refutation of misinformation. Medical misinformation has drastic health care consequences and pre-existing misinformation, particularly that surrounding vaccines, will be a significant future obstacle in overcoming COVID-19 [6]. The presentation of accurate information, including infodemiology data as illustrated in this study, to maintain societal ease is vital, and there is an imperative for scientists, public health authorities, and governments to collaborate to rigorously maintain this.

## Acknowledgments

RABG was funded by Cardiff University. JPC was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) through the South West Biosciences Doctoral Training Partnership (grant BB/M009122/1). BPK was funded by Idorsia Pharmaceuticals. HWE was funded by Cardiff University. Work in the ELE lab was supported by the BBSRC (BB/S002774/1).

## Authors' Contributions

RABG and JPC cowrote the manuscript, provided sources, extracted and prepared data, performed statistical analyses, and prepared figures. BPK provided sources and synthesis, and assisted with preparing the manuscript. HWE provided data, sources, and supervision, and assisted with preparing and writing the manuscript. ELE designed the study, provided early data searches, and contributed to writing the manuscript. All authors contributed to the editing of the manuscript and design of the study.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Table S1 and Figure S1.

[DOCX File, 194 KB - [jmir\\_v22i10e19791\\_app1.docx](#)]

## References

1. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *J Med Internet Res* 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: [10.2196/19016](#)] [Medline: [32287039](#)]
2. WHO timeline - COVID-19. World Health Organization. 2020. URL: <https://www.who.int/news-room/detail/08-04-2020-who-timeline---covid-19> [accessed 2020-04-28]
3. Jolley D, Paterson JL. Pylons ablaze: examining the role of 5G COVID-19 conspiracy beliefs and support for violence. *Br J Soc Psychol* 2020 Jul;59(3):628-640 [FREE Full text] [doi: [10.1111/bjso.12394](#)] [Medline: [32564418](#)]
4. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. *J Med Internet Res* 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: [10.2196/19118](#)] [Medline: [32302966](#)]
5. Jaiswal J, Halkitis PN. Towards a more inclusive and dynamic understanding of medical mistrust informed by science. *Behav Med* 2019;45(2):79-85. [doi: [10.1080/08964289.2019.1619511](#)] [Medline: [31343962](#)]
6. Carrieri V, Madio L, Principe F. Vaccine hesitancy and (fake) news: quasi-experimental evidence from Italy. *Health Econ* 2019 Nov;28(11):1377-1382 [FREE Full text] [doi: [10.1002/hec.3937](#)] [Medline: [31429153](#)]
7. Speed E, Mannion R. The rise of post-truth populism in pluralist liberal democracies: challenges for health policy. *Int J Health Policy Manag* 2017 May 01;6(5):249-251 [FREE Full text] [doi: [10.1517/ijhpm.2017.19](#)] [Medline: [28812811](#)]
8. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009 Mar 27;11(1):e11 [FREE Full text] [doi: [10.2196/jmir.1157](#)] [Medline: [19329408](#)]



9. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 2009 Nov 15;49(10):1557-1564. [doi: [10.1086/630200](https://doi.org/10.1086/630200)] [Medline: [19845471](https://pubmed.ncbi.nlm.nih.gov/19845471/)]
10. Verma M, Kishore K, Kumar M, Sondh AR, Aggarwal G, Kathirvel S. Google search trends predicting disease outbreaks: an analysis from India. *Healthc Inform Res* 2018 Oct;24(4):300-308 [FREE Full text] [doi: [10.4258/hir.2018.24.4.300](https://doi.org/10.4258/hir.2018.24.4.300)] [Medline: [30443418](https://pubmed.ncbi.nlm.nih.gov/30443418/)]
11. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. *Int Forum Allergy Rhinol* 2020 Jul;10(7):839-847 [FREE Full text] [doi: [10.1002/alr.22580](https://doi.org/10.1002/alr.22580)] [Medline: [32279437](https://pubmed.ncbi.nlm.nih.gov/32279437/)]
12. EU open data portal. European Union. 2020. URL: <https://data.europa.eu/euodp/en/data/> [accessed 2020-04-15]
13. Total COVID-19 test per 1,000 people. Our World in Data. 2020. URL: <https://ourworldindata.org/grapher/full-list-cumulative-total-tests-per-thousand> [accessed 2020-04-29]
14. Clement J. Worldwide desktop market share of leading search engines from January 2010 to July 2020. Statista. 2020. URL: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/> [accessed 2020-04-28]
15. Search engine market share worldwide. StatCounter. 2020. URL: <https://gs.statcounter.com/search-engine-market-share> [accessed 2020-04-28]
16. Dhillon P, Breuer M, Hirst N. COVID-19 breakthroughs: separating fact from fiction. *FEBS J* 2020 Jun 05:A [FREE Full text] [doi: [10.1111/febs.15442](https://doi.org/10.1111/febs.15442)] [Medline: [32500928](https://pubmed.ncbi.nlm.nih.gov/32500928/)]
17. 2020. URL: <https://www.fda.gov/drugs/information-drug-class/qa-consumers-hand-sanitizers-and-covid-19>
18. Dargaville T, Spann K, Celina M. Opinion to address the personal protective equipment shortage in the global community during the COVID-19 outbreak. *Polym Degrad Stab* 2020 Jun;176:109162 [FREE Full text] [doi: [10.1016/j.polymdegradstab.2020.109162](https://doi.org/10.1016/j.polymdegradstab.2020.109162)] [Medline: [32292217](https://pubmed.ncbi.nlm.nih.gov/32292217/)]
19. Ranney ML, Griffith V, Jha AK. Critical supply shortages - the need for ventilators and personal protective equipment during the Covid-19 pandemic. *N Engl J Med* 2020 Apr 30;382(18):e41. [doi: [10.1056/NEJMp2006141](https://doi.org/10.1056/NEJMp2006141)] [Medline: [32212516](https://pubmed.ncbi.nlm.nih.gov/32212516/)]
20. Koirala A, Joo YJ, Khatami A, Chiu C, Britton PN. Vaccines for COVID-19: the current state of play. *Paediatr Respir Rev* 2020 Sep;35:43-49 [FREE Full text] [doi: [10.1016/j.prrv.2020.06.010](https://doi.org/10.1016/j.prrv.2020.06.010)] [Medline: [32653463](https://pubmed.ncbi.nlm.nih.gov/32653463/)]
21. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir Med* 2020 Apr;8(4):e21. [doi: [10.1016/s2213-2600\(20\)30116-8](https://doi.org/10.1016/s2213-2600(20)30116-8)]
22. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *J Med Internet Res* 2020 May 06;22(5):e19458 [FREE Full text] [doi: [10.2196/19458](https://doi.org/10.2196/19458)] [Medline: [32352383](https://pubmed.ncbi.nlm.nih.gov/32352383/)]
23. Chaturvedi P, Ramalingam N, Singh A. Is COVID-19 man-made? *Cancer Res Stat Treat* 2020;3(2):284. [doi: [10.4103/crst.crst.166.20](https://doi.org/10.4103/crst.crst.166.20)]
24. Pradhan P, Pandey AK, Mishra A, Gupta P, Tripathi PM, Menon MB, et al. Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag. *BioRxiv* 2020 Feb 02:A. [doi: [10.1101/2020.01.30.927871](https://doi.org/10.1101/2020.01.30.927871)]
25. FDA advises patients on use of non-steroidal anti-inflammatory drugs (NSAIDs) for COVID-19. 2020. URL: <https://www.fda.gov/drugs/drug-safety-and-availability/fda-advises-patients-use-non-steroidal-anti-inflammatory-drugs-nsaids-covid-19> [accessed 2020-03-19]
26. EMA gives advice on the use of non-steroidal anti-inflammatories for COVID-19. European Medicines Agency. 2020 Mar 18. URL: <https://www.ema.europa.eu/en/news/ema-gives-advice-use-non-steroidal-anti-inflammatories-covid-19> [accessed 2020-03-18]
27. Calisher C, Carroll D, Colwell R, Corley RB, Daszak P, Drosten C, et al. Statement in support of the scientists, public health professionals, and medical professionals of China combatting COVID-19. *Lancet* 2020 Mar;395(10226):e42-e43. [doi: [10.1016/s0140-6736\(20\)30418-9](https://doi.org/10.1016/s0140-6736(20)30418-9)]
28. R: A Language and Environment for Statistical Computing. URL: <https://www.r-project.org/>
29. Wickham H. ggplot2: Elegant Graphics for Data Analysis 3.3.0. New York City: Springer; Jun 08, 2016.
30. Neuwirth E. RColorBrewer: ColorBrewer Palettes R package version 1.1-2. The R Foundation. URL: <https://cran.r-project.org/web/packages/RColorBrewer/index.html>
31. Harrell FE, Dupont C. Hmisc: Harrell miscellaneous R package version 4.4-0. The R Foundation. 2008. URL: <https://cran.r-project.org/web/packages/Hmisc/index.html>
32. Wei T, Simko V. R package "corrplot": visualization of a correlation matrix (version 0.84). The R Foundation. 2017. URL: <https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html>
33. Granier S. Viridis: default color maps from 'matplotlib'. R package version 0.5.1. The R Foundation. 2018. URL: <https://cran.r-project.org/web/packages/viridis/viridis.pdf>
34. GraphPad. URL: <http://www.graphpad.com/>
35. Campbell A, Morgan E. Comparisons of all-cause mortality between European countries and regions: January to June 2020. Office for National Statistics. 2020. URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/comparisonsofallcausemortalitybetweeneuropeancountriesandregions/januarytojune2020>
36. Panovska-Griffiths J. Opinion: coronavirus - five reasons why the UK death toll is so high. UCL News. 2020 Jun 11. URL: <https://www.ucl.ac.uk/news/2020/jun/opinion-coronavirus-five-reasons-why-uk-death-toll-so-high>
37. Propper C, Stoye G, Zaranko B. The wider impacts of the coronavirus pandemic on the NHS. Institute for Fiscal Studies. 2020 Apr 09. URL: <https://www.ifs.org.uk/publications/14798>

38. Choi H, Cho W, Kim M, Hur J. Public health emergency and crisis management: case study of SARS-CoV-2 outbreak. *Int J Environ Res Public Health* 2020 Jun 04;17(11):3984 [FREE Full text] [doi: [10.3390/ijerph17113984](https://doi.org/10.3390/ijerph17113984)] [Medline: [32512742](https://pubmed.ncbi.nlm.nih.gov/32512742/)]
39. Stafford N. Covid-19: why Germany's case fatality rate seems so low. *BMJ* 2020 Apr 07;369:m1395. [doi: [10.1136/bmj.m1395](https://doi.org/10.1136/bmj.m1395)] [Medline: [32265194](https://pubmed.ncbi.nlm.nih.gov/32265194/)]
40. Coronavirus (COVID-19) advice for international travellers. Australian Government. 2020. URL: <https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/coronavirus-covid-19-restrictions/coronavirus-covid-19-advice-for-international-travellers>
41. COVID-19 National Incident Room Surveillance Team. COVID-19, Australia: epidemiology report 11 (reporting week to 23:59 AEST 12 April 2020). *Commun Dis Intell* (2018) 2020 Apr 17:44. [doi: [10.33321/cdi.2020.44.34](https://doi.org/10.33321/cdi.2020.44.34)] [Medline: [32299330](https://pubmed.ncbi.nlm.nih.gov/32299330/)]
42. Dunford D, Dale B, Stylianou N, Lowther E, Ahmed M, de la Torre Arenas I. Coronavirus: the world in lockdown in maps and charts. *BBC News*. 2020 Apr 07. URL: <https://www.bbc.co.uk/news/world-52103747>
43. Rettie H, Daniels J. Coping and tolerance of uncertainty: predictors and mediators of mental health during the COVID-19 pandemic. *Am Psychol* 2020 Aug 03:A. [doi: [10.1037/amp0000710](https://doi.org/10.1037/amp0000710)] [Medline: [32744841](https://pubmed.ncbi.nlm.nih.gov/32744841/)]
44. Ozamiz-Etxebarria N, Idoiaga Mondragon N, Dosil Santamaría M, Picaza Gorrotxategi M. Psychological symptoms during the two stages of lockdown in response to the COVID-19 outbreak: an investigation in a sample of citizens in Northern Spain. *Front Psychol* 2020;11:1491. [doi: [10.3389/fpsyg.2020.01491](https://doi.org/10.3389/fpsyg.2020.01491)] [Medline: [32625157](https://pubmed.ncbi.nlm.nih.gov/32625157/)]
45. Pagliari C. What do world leaders really mean by refusing to wear face masks? The University of Edinburgh. 2020 Jun 17. URL: <https://www.ed.ac.uk/covid-19-response/expert-insights/what-messages-are-world-leaders-giving-when-they-r>
46. Giuffrida A, Cochrane L. Italy imposes draconian rules to stop spread of coronavirus. *The Guardian*. 2020 Feb 23. URL: <https://www.theguardian.com/world/2020/feb/23/italy-draconian-measures-effort-halt-coronavirus-outbreak-spread>
47. Vinceti M, Filippini T, Rothman K, Ferrari F, Goffi A, Maffei G, et al. Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking. *EClinicalMedicine* 2020 Aug;25:100457 [FREE Full text] [doi: [10.1016/j.eclinm.2020.100457](https://doi.org/10.1016/j.eclinm.2020.100457)] [Medline: [32838234](https://pubmed.ncbi.nlm.nih.gov/32838234/)]
48. Bandyopadhyay G, Meltzer A. Let us unite against COVID-19 - a New Zealand perspective. *Ir J Psychol Med* 2020 May 14:1-4 [FREE Full text] [doi: [10.1017/ipm.2020.44](https://doi.org/10.1017/ipm.2020.44)] [Medline: [32404225](https://pubmed.ncbi.nlm.nih.gov/32404225/)]
49. Carius BM, Schauer SG. Ibuprofen during the COVID-19 pandemic: social media precautions and implications. *West J Emerg Med* 2020 Apr 24;21(3):497-498 [FREE Full text] [doi: [10.5811/westjem.2020.4.47686](https://doi.org/10.5811/westjem.2020.4.47686)] [Medline: [32421496](https://pubmed.ncbi.nlm.nih.gov/32421496/)]
50. Romano S, Galante H, Figueira D, Mendes Z, Rodrigues AT. Time-trend analysis of medicine sales and shortages during COVID-19 outbreak: data from community pharmacies. *Res Social Adm Pharm* 2020 May 23:30611-30612 [FREE Full text] [doi: [10.1016/j.sapharm.2020.05.024](https://doi.org/10.1016/j.sapharm.2020.05.024)] [Medline: [32482587](https://pubmed.ncbi.nlm.nih.gov/32482587/)]
51. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res* 2020 Mar;30(3):269-271 [FREE Full text] [doi: [10.1038/s41422-020-0282-0](https://doi.org/10.1038/s41422-020-0282-0)] [Medline: [32020029](https://pubmed.ncbi.nlm.nih.gov/32020029/)]
52. Flanagan C, Griffin R, Langreth R. The race to develop coronavirus treatments and vaccines is heating up. *Bloomberg*. 2020. URL: <https://www.bloomberg.com/features/2020-coronavirus-drug-vaccine-status/>
53. Oxford COVID-19 vaccine programme opens for clinical trial recruitment. *COVID-19 Oxford Vaccine Trial*. 2020 Mar 27. URL: <https://covid19vaccinetrial.co.uk/press-release-trial-open>
54. Sattui SE, Liew JW, Graef ER, Coler-Reilly A, Berenbaum F, Duarte-García A, et al. Swinging the pendulum: lessons learned from public discourse concerning hydroxychloroquine and COVID-19. *Expert Rev Clin Immunol* 2020 Aug 11:1-8. [doi: [10.1080/1744666X.2020.1792778](https://doi.org/10.1080/1744666X.2020.1792778)] [Medline: [32620062](https://pubmed.ncbi.nlm.nih.gov/32620062/)]
55. Gautret P, Lagier J, Parola P, Hoang VT, Meddeb L, Mailhe M, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents* 2020 Jul;56(1):105949. [doi: [10.1016/j.ijantimicag.2020.105949](https://doi.org/10.1016/j.ijantimicag.2020.105949)]
56. Borba MGS, Val FFA, Sampaio VS, Alexandre MAA, Melo GC, Brito M, CloroCovid-19 Team. Effect of high vs low doses of chloroquine diphosphate as adjunctive therapy for patients hospitalized with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection: a randomized clinical trial. *JAMA Netw Open* 2020 Apr 24;3(4):e208857 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.8857](https://doi.org/10.1001/jamanetworkopen.2020.8857)] [Medline: [32330277](https://pubmed.ncbi.nlm.nih.gov/32330277/)]
57. Mahase E. Hydroxychloroquine for covid-19: the end of the line? *BMJ* 2020 Jun 15;369:m2378. [doi: [10.1136/bmj.m2378](https://doi.org/10.1136/bmj.m2378)] [Medline: [32540958](https://pubmed.ncbi.nlm.nih.gov/32540958/)]
58. Chai PR, Ferro EG, Kirshenbaum JM, Hayes BD, Culbreth SE, Boyer EW, et al. Intentional hydroxychloroquine overdose treated with high-dose diazepam: an increasing concern in the COVID-19 pandemic. *J Med Toxicol* 2020 Jul;16(3):314-320 [FREE Full text] [doi: [10.1007/s13181-020-00790-8](https://doi.org/10.1007/s13181-020-00790-8)] [Medline: [32514696](https://pubmed.ncbi.nlm.nih.gov/32514696/)]
59. Kampf G, Todt D, Pfaender S, Steinmann E. Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. *J Hosp Infect* 2020 Mar;104(3):246-251 [FREE Full text] [doi: [10.1016/j.jhin.2020.01.022](https://doi.org/10.1016/j.jhin.2020.01.022)] [Medline: [32035997](https://pubmed.ncbi.nlm.nih.gov/32035997/)]
60. Kuehn BM. Spike in poison control calls related to disinfectant exposures. *JAMA* 2020 Jun 09;323(22):2240. [doi: [10.1001/jama.2020.8307](https://doi.org/10.1001/jama.2020.8307)] [Medline: [32515822](https://pubmed.ncbi.nlm.nih.gov/32515822/)]

61. Inagaki H, Saito A, Sugiyama H, Okabayashi T, Fujimoto S. Rapid inactivation of SARS-CoV-2 with deep-UV LED irradiation. *Emerg Microbes Infect* 2020 Dec;9(1):1744-1747 [FREE Full text] [doi: [10.1080/22221751.2020.1796529](https://doi.org/10.1080/22221751.2020.1796529)] [Medline: [32673522](https://pubmed.ncbi.nlm.nih.gov/32673522/)]
62. Brennen JS, Simon F, Howard PN, Nielsen RK. Types, sources, and claims of COVID-19 misinformation. Reuters Institute. 2020 Apr 07. URL: <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>
63. Majumder MS, Mandl KD. Early in the epidemic: impact of preprints on global discourse about COVID-19 transmissibility. *Lancet Glob Health* 2020 May;8(5):e627-e630 [FREE Full text] [doi: [10.1016/S2214-109X\(20\)30113-3](https://doi.org/10.1016/S2214-109X(20)30113-3)] [Medline: [32220289](https://pubmed.ncbi.nlm.nih.gov/32220289/)]
64. Onchonga D. A Google Trends study on the interest in self-medication during the 2019 novel coronavirus (COVID-19) disease pandemic. *Saudi Pharm J* 2020 Jul;28(7):903-904 [FREE Full text] [doi: [10.1016/j.jsps.2020.06.007](https://doi.org/10.1016/j.jsps.2020.06.007)] [Medline: [32641903](https://pubmed.ncbi.nlm.nih.gov/32641903/)]
65. Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY, Snot Force Alliance. Correlations of online search engine trends with coronavirus disease (COVID-19) incidence: infodemiology study. *JMIR Public Health Surveill* 2020 May 21;6(2):e19702 [FREE Full text] [doi: [10.2196/19702](https://doi.org/10.2196/19702)] [Medline: [32401211](https://pubmed.ncbi.nlm.nih.gov/32401211/)]
66. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health* 2017 Sep;7(3):185-189 [FREE Full text] [doi: [10.1016/j.jegh.2017.06.001](https://doi.org/10.1016/j.jegh.2017.06.001)] [Medline: [28756828](https://pubmed.ncbi.nlm.nih.gov/28756828/)]

## Abbreviations

**BBSRC:** Biotechnology and Biological Sciences Research Council

**GT:** Google Trends

**PPE:** personal protective equipment

**RSV:** relative search volume

*Edited by G Eysenbach, G Fagherazzi; submitted 08.05.20; peer-reviewed by B Akinwunmi, M Verma; comments to author 13.07.20; revised version received 06.08.20; accepted 27.08.20; published 27.10.20.*

*Please cite as:*

Badell-Grau RA, Cuff JP, Kelly BP, Waller-Evans H, Lloyd-Evans E

*Investigating the Prevalence of Reactive Online Searching in the COVID-19 Pandemic: Infoveillance Study*

*J Med Internet Res* 2020;22(10):e19791

URL: <https://www.jmir.org/2020/10/e19791>

doi: [10.2196/19791](https://doi.org/10.2196/19791)

PMID: [32915763](https://pubmed.ncbi.nlm.nih.gov/32915763/)

©Rafael A Badell-Grau, Jordan Patrick Cuff, Brendan P Kelly, Helen Waller-Evans, Emyr Lloyd-Evans. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 27.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Review

## Artificial Intelligence for COVID-19: Rapid Review

Jiayang Chen<sup>1</sup>, MBBS; Kay Choong See<sup>1,2</sup>, MBBS, MRCP, MPH, MHPE, EDIC

<sup>1</sup>Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

<sup>2</sup>Division of Respiratory & Critical Care Medicine, Department of Medicine, National University Hospital, Singapore, Singapore

**Corresponding Author:**

Jiayang Chen, MBBS

Yong Loo Lin School of Medicine

National University of Singapore

10 Medical Dr

Singapore, 117597

Singapore

Phone: 65 84369198

Email: [jiayang\\_chen1997@live.com](mailto:jiayang_chen1997@live.com)

**Abstract**

**Background:** COVID-19 was first discovered in December 2019 and has since evolved into a pandemic.

**Objective:** To address this global health crisis, artificial intelligence (AI) has been deployed at various levels of the health care system. However, AI has both potential benefits and limitations. We therefore conducted a review of AI applications for COVID-19.

**Methods:** We performed an extensive search of the PubMed and EMBASE databases for COVID-19–related English-language studies published between December 1, 2019, and March 31, 2020. We supplemented the database search with reference list checks. A thematic analysis and narrative review of AI applications for COVID-19 was conducted.

**Results:** In total, 11 papers were included for review. AI was applied to COVID-19 in four areas: diagnosis, public health, clinical decision making, and therapeutics. We identified several limitations including insufficient data, omission of multimodal methods of AI-based assessment, delay in realization of benefits, poor internal/external validation, inability to be used by laypersons, inability to be used in resource-poor settings, presence of ethical pitfalls, and presence of legal barriers. AI could potentially be explored in four other areas: surveillance, combination with big data, operation of other core clinical services, and management of patients with COVID-19.

**Conclusions:** In view of the continuing increase in the number of cases, and given that multiple waves of infections may occur, there is a need for effective methods to help control the COVID-19 pandemic. Despite its shortcomings, AI holds the potential to greatly augment existing human efforts, which may otherwise be overwhelmed by high patient numbers.

(*J Med Internet Res* 2020;22(10):e21476) doi:[10.2196/21476](https://doi.org/10.2196/21476)

**KEYWORDS**

coronavirus; deep learning; machine learning; medical informatics; computing; SARS virus; COVID-19; artificial intelligence; review

**Introduction**

COVID-19, caused by SARS-CoV-2 [1], was first discovered in December 2019 and has since become a global pandemic [2]. An emerging viral pandemic like COVID-19 exerts significant pressure on limited health care resources [3]. To prevent human efforts of disease containment from being overwhelmed, we need tools that can streamline the diagnosis, surveillance, and treatment of COVID-19 [4]. This need is particularly pressing in relatively resource-scarce settings, such as low- or middle-income countries [5,6].

Digital methods such as artificial intelligence (AI) hold the potential to greatly enhance medical care [7]. AI implies the use of a computer to model intelligent behavior without human intervention [8]. It has been applied to many areas of medicine [9], especially to aid the detection and prevention of disease [10]. AI techniques being used in medicine are broad, ranging from computer vision to deep learning techniques [11]. Unlike the if-then rules used in traditional computer programming, AI methods emulate the decision-making process of humans via two major approaches. The first major approach is supervised machine learning, which aims to develop a predictive algorithm using regression (linear or multiple) or classification methods



(eg, decision trees, neural networks). The other major AI approach is unsupervised machine learning, which allows computers to explore large amounts of unclassified data and to discover novel disease or treatment patterns [12].

An example of how AI has been applied to COVID-19 is demonstrated by BlueDot, a Canadian company specializing in infectious disease forecasting [13]. Using an AI engine that continuously gathers data for a multitude of diseases from a range of different sources globally, BlueDot was able to predict the COVID-19 outbreak and alert its users even before the World Health Organization did [14]. Another example is an AI-powered chatbot named SGDormBot, which has been used for symptom-based mass screening of migrant workers for COVID-19 in Singapore [15].

Nonetheless, while AI has been promoted as a tool to help manage the COVID-19 pandemic, AI has both potential benefits and limitations. We therefore conducted a rapid review of AI applications for COVID-19. In our review, we sought to delineate the major categories of AI use, describe the limitations of AI, and identify areas for further development.

## Methods

We based our review on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [16], and extensively searched two databases (PubMed and EMBASE) for all English-language papers published from December 1, 2019, to March 31, 2020, using the search terms “novel coronavirus,” “2019 novel coronavirus,” “2019-nCoV,” “coronavirus disease 2019,” “COVID-19,” and “SARS-CoV-2.” The database search was supplemented by reference list checks. Papers reporting new data on AI applications for COVID-19

were included. Review papers and commentaries without new data were excluded.

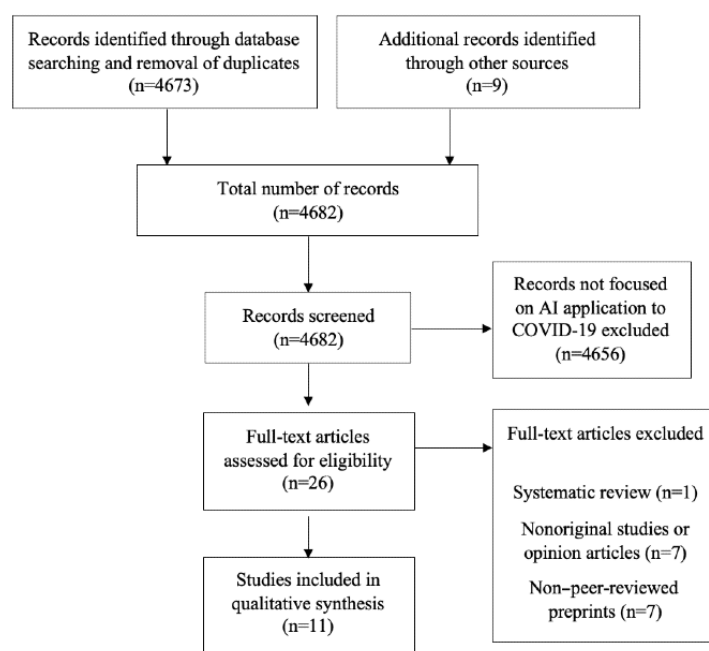
The quality of included studies was assessed using a modified TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement for adherence to reporting standards [17] and PROBAST (prediction model risk of bias assessment tool) for risk of bias [18]. A thematic analysis and narrative review of AI applications for COVID-19 was then conducted.

## Results

### Included Studies, Adherence to Reporting Standards, and Risk of Bias

Of 4682 articles, 11 articles were included for review (Figure 1, Table 1). The original TRIPOD statement consisted of a 22-category checklist with 37 items. However, items related to predictor variables were not relevant for studies assessing the performance of AI algorithms. The final components selected, as well as the degree of adherence in each category, are shown in Figure 2. Overall, adherence rates of publications to TRIPOD items ranged from 18.2% to 100%. Items concerned with data validation were reported in <50% of the 11 publications reviewed. Items related to clinical context, study methodology, and applicability were better reported, with >80% of the publications providing adequate information. For the assessment of risk of bias and applicability, we applied the PROBAST tool (Figure 3); as with TRIPOD, we did not assess predictors. Generally, risk of bias in the remaining three categories was low. We identified four areas where AI was applied to COVID-19: diagnosis, public health, clinical decision making, and therapeutics.

**Figure 1.** Study flow diagram.

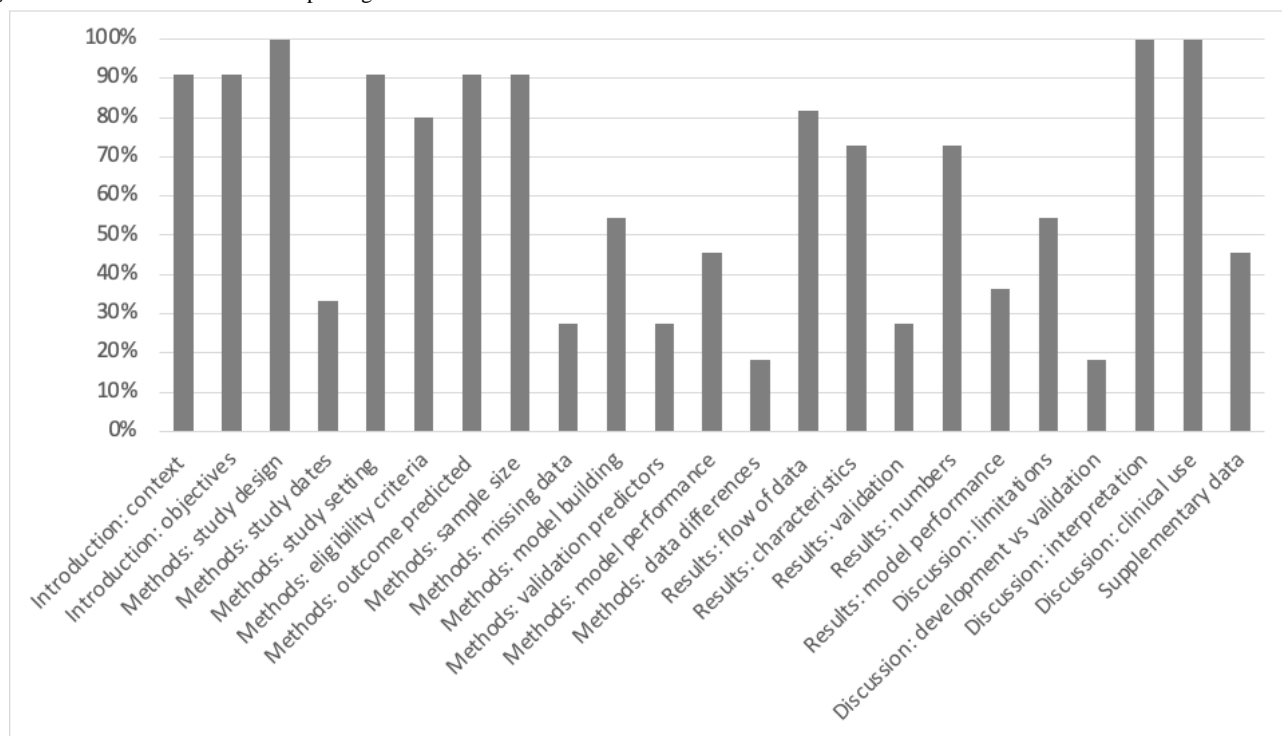
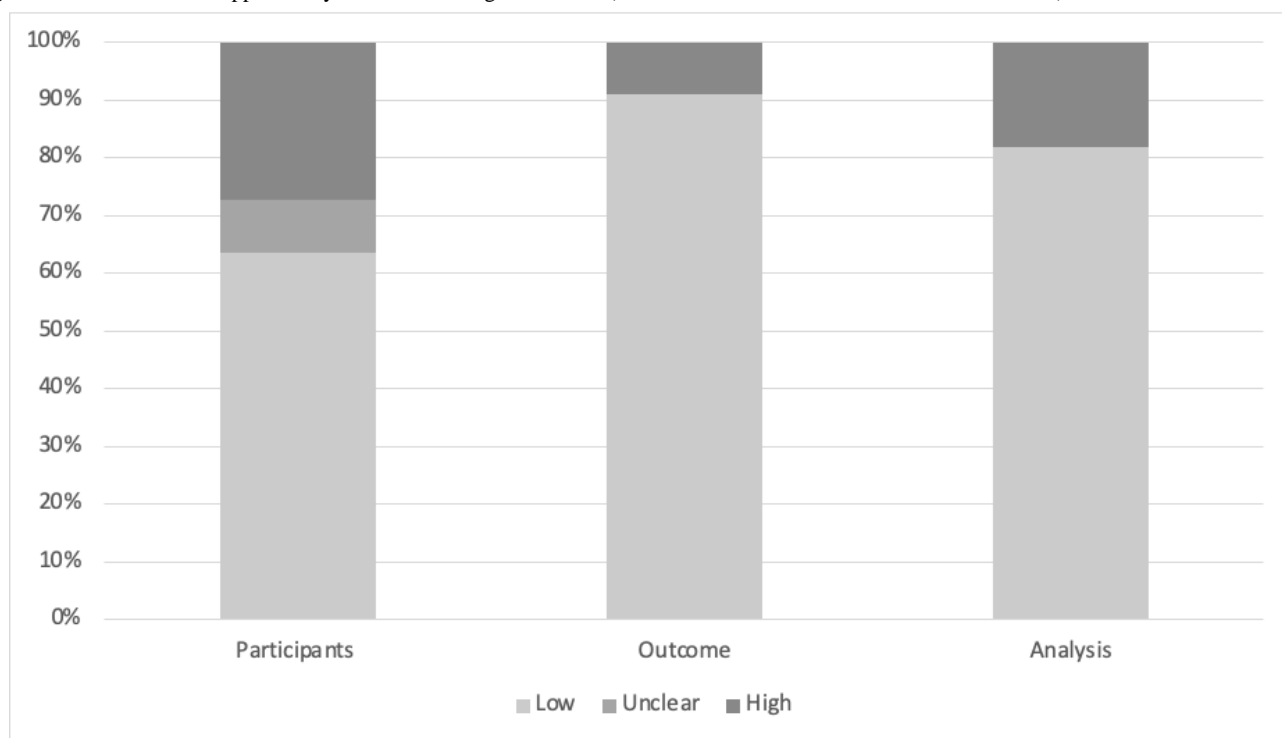




**Table 1.** Studies included in the review.

First author (year)	Area or specialty of AI <sup>a</sup> application	AI application method	Clinical benefit shown	Internal or external validation done
Hurt (2020) [19]	Diagnosis and clinical decision making	Using a deep learning approach to augment radiographs with color probability	Improved diagnostic accuracy of pneumonia and COVID-19 at point of care, triaged patient for CT <sup>b</sup> scan, helped physicians track evolution of pulmonary manifestation over length of hospitalization	No
Li (2020) [20]	Diagnosis	Deep learning-based computer-aided diagnostic system for pneumonia trained with CT scans of patients with COVID-19 suggested pneumonia in patients who received a negative reverse-transcription polymerase chain reaction test result	Improved accuracy of diagnosis	No
Li (2020) [21]	Diagnosis	AI 3D deep learning model to analyze CT scan	Improved diagnostic accuracy and differentiated from non-COVID-19 lung pathologies	Yes
Yang (2020) [22]	Public health	Recurrent neural network for AI-based prediction of epidemic trend	Good epidemiological modeling and prediction of trends relating to COVID-19	Yes
Al-Najjar (2020) [23]	Public health	AI-based classifier prediction model to determine the outcome of patients	Identified key factors influencing clinical outcome, guided public health decision making	No
Jiang (2020) [24]	Clinical decision making	Tool with AI capabilities that will predict patients at risk for more severe illnesses based on clinical parameters	AI tool predicted patients at risk for more severe illness on initial presentation, provided clinical decision support	Yes
Beck (2020) [25]	Therapeutics	Used pretrained deep learning-based system to identify commercially available drugs that could act on the viral proteins of SARS-CoV-2	Used AI to discover that atazanavir, an antiretroviral medication, is the best chemical compound due to its high inhibitory potency, among several other antiviral agents that could be used in the treatment of SARS-CoV-2	Yes
Kadioglu (2020) [26]	Therapeutics	AI combined with molecular docking to identify candidates suitable for drug repurposing via in silico methods	Supervised machine learning was used to study drug likeliness of candidate compounds, helped with evaluation of the potential of various agents	Yes
Richardson (2020) [27]	Therapeutics	Use of BenevolentAI's knowledge graph to search for approved drugs that can help treat COVID-19	Baricitinib was identified as a viable drug with tolerable side effects and potential therapeutic use in patients with COVID-19	No
Ton (2020) [28]	Therapeutics	Use of Deep Docking for accelerated screening of large chemical libraries for potential drugs against COVID-19	Screened through 1.3 billion compounds from the ZINC15 library to identify the top 1000 potential ligands against the main protease (Mpro) of SARS-CoV-2 and made them publicly available	Yes
Zhang (2020) [29]	Therapeutics	Use of AI-based dock analysis to determine whether the compounds listed in Traditional Chinese Medicine databases had potential for direct SARS-CoV-2 protein interaction	Identified 26 herbal plants containing compounds potentially active against SARS-CoV-2	No

<sup>a</sup>AI: artificial intelligence.<sup>b</sup>CT: computed tomography.

**Figure 2.** Adherence of studies to reporting standards.**Figure 3.** Risk of bias and applicability assessment using PROBAST (Prediction model Risk Of Bias ASsessment Tool).

## Diagnosis

Radiological diagnosis relies on imaging data and is amenable to deep-learning techniques [30]. For COVID-19, features seen on computed tomography (CT) images include bilateral ground-glass opacification and consolidation [31]. One fully automatic 3D deep-learning framework developed for the detection of COVID-19 (COVNet) was able to extract relevant information from both 2D and 3D images obtained from a CT

scan to generate a probability score to distinguish patients with COVID-19 from patients with non-COVID-19 community-acquired pneumonia [21]. Once trained, COVNet could process each CT scan with high sensitivity (90%) and specificity (96%) for COVID-19 identification. In contrast to CT, real-time reverse-transcription polymerase chain reaction (RT-PCR), a technique frequently used for COVID-19 diagnosis, had a much lower sensitivity (71%-80%) [20,32]. Another AI system (InferRead CT Pneumonia), a deep learning diagnostic

system, was able to identify features of coronavirus infection in CT scans of patients with false-negative RT-PCR results [33]. Even without CT images, using only plain chest radiographs, a deep learning approach could improve diagnosis of COVID-19 pneumonia by augmenting radiographs with color probability [19].

Apart from superior sensitivity compared to RT-PCR, AI coupled with radiological investigations could be more efficient for COVID-19 diagnosis. For instance, COVNet could process each CT scan in under 5 seconds on average [21]. This could both decrease physician workload and allow a greater volume of patients to be evaluated, facilitating more rapid case detection [34].

## Public Health

AI has proven to be a valuable tool in public health efforts, helping to characterize the epidemiology of COVID-19 and model disease transmission, even in the early stages of the pandemic. The application of AI to this area would be particularly helpful for policy makers, in contrast to clinical decision making by individual clinicians.

Perhaps one of the most prominent examples of AI used in public health, BlueDot has demonstrated its effectiveness in predicting and monitoring infectious diseases like COVID-19. Other applications of AI gathered from our rapid review are discussed as follows. Yang et al [22] developed an AI-based model that used a form of recurrent neural network (RNN) for epidemiological modelling. The model was trained using the 2003 severe acute respiratory syndrome (SARS) epidemic data from China, while incorporating the epidemiological parameters of COVID-19 and public health interventions like the lockdown of Hubei province. Using the model, it was predicted that the number of cases in Hubei would peak in early February. When the number of predicted infections was plotted against real-time epidemiologic trends, there was considerable similarity between the actual numbers and predictions made by the AI, supporting the accuracy of such models for forecasting disease development. Similar work is in progress using a wide range of data sources.

An artificial neural network has also been used to build a classifier prediction model for patient outcomes in South Korea [23]. The Korea Centers for Disease Control and Prevention collected patient characteristics like age and gender, extracted the independent predictors, and classified patients with COVID-19 into 2 groups: deceased and recovered. This helped policy makers target the most vulnerable patients, directing attention and resources toward their care.

## Clinical Decision Making

AI can assist in clinical decision making regarding patients with COVID-19, including triage decisions for the optimal use of limited health care resources [24]. Hurt et al [19] used an AI-augmented system for plain chest radiographs to track and predict the pulmonary progression of COVID-19 among hospitalized patients, which helped to identify patients that needed critical care. Early identification of at-risk cases could further guide clinicians toward earlier intervention, which may result in improved outcomes. For instance, Jiang et al [24] took

multiple clinical parameters and constructed an AI framework that could predict deterioration even at initial presentation, with superior accuracy compared to logistic regression.

## Therapeutics

In silico screening with AI can help identify potentially effective therapeutic agents among existing drugs (ie, drug repurposing). Deep learning technology has been used to screen 1.3 billion compounds from the ZINC15 library for drug repurposing [28]. Molecule transformer-drug target interaction (MT-DTI) was a natural language processing (NLP) tool used to predict binding affinity values between commercially available antiviral drugs and target proteins on SARS-CoV-2. This led to the identification of atazanavir, an antiretroviral medication, which may also be effective against SARS-CoV-2 [25]. Similarly, using BenevolentAI's knowledge graph and a library of structured medical information, machine learning uncovered baricitinib, a Janus kinase inhibitor used for rheumatoid arthritis, as a safe candidate drug that could inhibit SARS-CoV-2 viral entry [27]. In another study, a similar process of virtual drug screening identified antiviral agents against hepatitis C as drugs that had high binding affinities to target proteins on SARS-CoV-2 [26].

Apart from the identification of conventional Western medications that may be applied to COVID-19, in silico screening with AI has also been applied to Chinese herbal medicines. Zhang et al [29] used molecular docking analysis to determine whether natural compounds listed in the Traditional Chinese Medicine (TCM) Systems Pharmacology Database, Encyclopedia of Traditional Chinese Medicine, and SymMap could interact directly with SARS-CoV-2 proteins. They eventually shortlisted 26 herbal plants containing potential SARS-CoV-2 antivirals for further trials.

## Discussion

### Limitations of AI Identified

Despite the many benefits that AI can bring to the table, there are some significant limitations to its use.

### Difficult Data Collection

AI needs large amounts of training data to generate accurate predictive algorithms [35]. For instance, the European Centre for Disease Prevention and Control requires a dedicated team of epidemiologists to screen, validate, and collate data from multiple international, regional, and local sources [36]. Even data collection for smaller geographical areas is far from easy. In the city-state of Singapore, a Health Ministry–helmed national consortium has been required to assemble detailed data from various health systems, hospitals, and clinics [37].

However, even if financial and logistical resources are plentiful, large volumes of information may not be available during the early stages of disease outbreaks, which is ironically the time when prediction is most required. Hence, relying on AI early in a disease outbreak may be impractical. Nonetheless, even if large amounts of data are available, AI is not infallible, as demonstrated by the failure of GoogleFlu, a big data analysis tool for epidemiological trending for influenza [38]. Similarly,

diagnostic AI tools like those applied to radiology require large volumes of both image- and non-image-based clinical data to achieve high accuracy [11].

Apart from the quantity of data, obtaining high-quality inputs can be particularly challenging. Informal data sources and news reports [39] provide heterogeneous data with inherent noise, resulting in biased results [40]. This is particularly concerning in the context of epidemiological trending and disease prediction, when inaccurate forecasting leads to incorrect calibration of public health responses and health care resourcing. Furthermore, the collection of data from hospitals can be complicated. Some countries, such as Singapore, achieved success with data collection efforts by having a nationwide standardized protocol and rigorous contact tracing [37]. Other countries, such as the United States, rely on individual jurisdictions to report case information to the Centers for Disease Control and Prevention (CDC), but not all jurisdictions provide daily updates. As a result, the case counts may increase at different intervals [41]. Based on a report published by the CDC COVID-19 Response Team in March, less than 6% of cases reported had information regarding the patients' comorbidities and risk factors and this incomplete data collection has limited research efforts [42].

### ***Lack of Multimodal AI Assessments***

Many studies used singular data types to perform AI-driven tasks, for instance only using radiological images for the diagnosis of COVID-19 [20,21]. Assessments of patients based on a single data type may be skewed, thus highlighting the need for a multimodal AI framework capable of analyzing different data types [35]. As there is significant overlap in how the lungs respond to different pathological insults [33], and as radiological presentation often depends on an interplay of multiple factors [21], the full potential of CT-based AI algorithms may be better realized by including non-image-based clinical data. For now, based on recommendations by the Italian Society of Medical and Interventional Radiology, CT should be used as a screening tool only for symptomatic patients with specific indications, and the use of CT with AI for screening or as a first-line test is not supported [43].

### ***Delay in Realization of Benefits***

As useful as AI may be in helping to identify drugs and vaccines against COVID-19, such treatments are unlikely to be made widely available in the immediate future, precisely when such treatments are needed the most. There is considerable delay caused by rigorous medical trials required before approval can be given to drugs or vaccines, and it can take 12-18 months just to develop a vaccine [44]. Even with a vaccine on the market, supply shortage will likely be an issue in the face of massive demand and limited initial production [45].

### ***Poor Internal Validation***

Another issue we noticed was the poor internal validation done in some of the studies, which makes it hard to determine the clinical or incremental value of AI over conventional methods. Studies that discussed the performance of AI augmentation of CT scans [21] and plain radiographs [19] did not compare AI with radiologists' evaluations, and did not describe data sets

used for validation [33]. Furthermore, confounding factors can affect the internal validity of such studies, such as how variation in respiratory effort, image contrast, technique, and the resolution of radiological images may affect the accuracy of AI-based radiological interpretation frameworks [19]. After completion of our review, we found one study that addressed our validity concerns. Bai et al [46] compared AI with radiologists and demonstrated that AI alone had higher accuracy, sensitivity, and specificity than radiologists, whereas the performance of radiologists was enhanced when AI was used to augment their evaluations.

### ***Poor External Validation***

External validation could also be improved. Many studies were only applied to patients seen at single centers, or populations within the same geographical region [20,22,23]. This could mean that algorithms shown to be accurate for the population studied may fare less well in other settings. Increasing the diversity of data sets from different populations and demonstrating the reproducibility of AI-based algorithms in different settings would be required if we are to generalize the usage of AI tools [11].

### ***Inability to Be Used by Laypersons***

While the use of AI technology in the clinical context may seem simple, the underlying theory and operating mechanisms of these algorithms are often opaque to the untrained layperson, which includes health care professionals unfamiliar with AI. For example, one of the intrinsic drawbacks of deep learning is the lack of interpretability, as it is "impossible to determine what imaging features are being used to determine the output" [21]. Physicians may not trust AI to evaluate clinical scenarios, and would not be able to troubleshoot incorrect AI-based assessments [11]. To avoid hindering AI's uptake, implementation frameworks need to be designed in a way that make AI easily operable and clearly understood by most health care professionals.

### ***Inability to Be Used in Resource-Poor Settings***

AI-based methods of disease detection, surveillance, and prognostication often require access to digital resources such as CT scanners, mobile phones, and internet access. Such resources may not be widely or consistently available in less developed regions [47]. Thus, the benefits of AI-based approaches may not be realized in resource-poor settings. Effort is needed to shift the reliance of AI from expensive technologies to cheaper and more readily accessible alternatives such as chest X-ray, point-of-care ultrasound, or even vital signs data alone.

### ***Ethical Pitfalls***

The use of AI may require access to personal information to generate trends, make predictions, and conduct assessments. Moreover, individual patient information has been placed online on multiple platforms [48]. The sharing of such information may lead to the infringement of privacy and personal rights. Although it may be acceptable to process personal data for disease containment in a pandemic, problems arise when the data are used for sinister purposes. As such, there need to be proper ethical guidelines and laws in place to govern the use of AI and big data. For example, the Australian Human Rights



Commission has set out practical steps for researchers to control the use of AI [49]. Overall, health care organizations can be viewed by patients as trustworthy, but it is important to remember that the confidentiality of the data obtained from the public should be respected, and that there should be transparency in how institutions handle the data obtained [50].

A more subtle type of ethical pitfall concerns the need for human input when programming AI tools. AI's superiority in calculation and computation does not necessarily translate into good decision making unless the machine has values that enable it to make an ethical choice, which requires input from a human programmer [51]. In other words, a particular AI program will make choices based on the programmer's morality. This raises an ethical concern, as difficult public health decisions like resource allocation are often multifaceted and based on more than just a single set of ethical guidelines conforming to the beliefs of a single person or single group of people. On a smaller scale, the decision to rely solely on the computer to make a unilateral judgment on how or whether a patient should be treated based on the machine's calculated benefit of treatment is also fraught with ethical concerns as there are some who believe that medical treatment should be discussed together with the patient [50]. As such, it is prudent to use AI only as a guide in decision making, rather than relying on it wholly.

Like all prediction methods, AI may unwittingly single out ethnic minority groups as being at high risk for disease. Unethical interpretation of prediction results may exacerbate ethnic tensions and lead to the stigmatization of and discrimination against specific ethnic groups. To illustrate, the US CDC has reported that African Americans and Hispanics have higher rates of hospitalization and mortality from COVID-19 than the White population [52]. Even though such a finding is likely to be due to differential access to health care [53], the same finding may lead to misconceptions about race and inherent disease susceptibility.

### **Legal Barriers**

Finally, there may be legal liabilities associated with adverse outcomes when human physicians use AI technologies for the care of their patients. When malpractice cases involving medical AI applications are involved, the legal system needs to be clear on which party holds the liability [30]. This is particularly worrisome given that the usability of AI, especially for the care of patients with COVID-19, is still relatively unknown due to the lack of sufficient evidence of its effectiveness over traditional methods. Consequently, medical professionals may hesitate when asked to use AI for patient management.

### **Areas for Further Work**

There are several other areas in which AI has shown significant promise.

#### **Surveillance**

Infrared thermal cameras used to screen the public for fever have been paired with AI-powered facial recognition systems to determine if individuals are wearing surgical masks [54]. A US-based computer vision startup has started offering a software that uses camera images to observe for compliance with social

distancing rules [55]. With the proliferation of such technologies, it is increasingly evident that AI-based surveillance can greatly help with public health interventions that slow the spread of infection.

Blockchain technology is a digital method that can be used in conjunction with AI, and it refers to a verifiable permanent ledger system that can be used to store health care-related information [7]. The coupling of AI with blockchain for self-testing and tracking systems has been proposed for the surveillance of COVID-19. Not only can such a system of self-testing overcome supply chain limitations in resource-scarce countries and achieve a higher rate of testing [5], it also gives real-time feedback on population health and allows for risk stratification of suspect cases.

AI can also potentially predict the occurrence of disease outbreaks early on, thus giving health authorities more time to act. Effenberger et al [56] have demonstrated the effectiveness of using internet relative search volume (RSV) indices for the forecasting of COVID-19 outbreaks in different geographical areas. Maximum public interest, and therefore maximum RSV indices, preceded peaks in case numbers. This meant that trending RSV indices can help public health authorities predict and respond to local surges of COVID-19 cases.

### **Combination With Big Data**

AI-based techniques like machine learning can be used to gather data from multiple sources for processing, which can provide novel insights. We have discussed how AI's ability to analyze large amounts of data swiftly makes it possible to use the data to predict the likelihood of new outbreaks, model successful disease containment strategies, and to find the most effective treatment protocols [9]. However, there are some other areas of application that have yet to be explored. Sun et al [57] demonstrated the value of using a Chinese health care-oriented social network that streamed reports from local or national health authorities—along with several other international media outlets—for epidemiological studies of the disease; AI can be used to facilitate this collection.

AI, when paired with big data, can find new drugs to treat COVID-19. Integrated AI-based drug discovery methods for novel drug compounds, such as deep generative models, can use large data sets to train and generate new drugs with optimized chemical properties. Such methods are often more time-efficient than traditional computational methods [58].

In addition, AI combined with big data can also be used to assess the accuracy of online information available to the public. Stratifyd, a US-based data analytics company, scans social media posts, cross-references them with data from official sources, and alerts users when false information is identified [59]. It can thus be applied to the current pandemic to prevent misinformation about the disease from spreading online. AI can also be used to analyze international air travel data. By doing so, AI can track disease spread between countries, allowing for an assessment of importation risk at a particular location and the latter's capacity to respond [60], estimation of disease spread from the epicenter of the outbreak [61], and identification of



areas that may have undetected cases imported from other countries [62].

### **Operation of Other Core Clinical Services**

As the COVID-19 pandemic exerts pressure on health care resources, institutions have had to reduce the provision of clinical services. The American College of Surgeons, for example, has provided guidelines for the management of nonemergency operations, recommending hospitals to “thoughtfully review all scheduled elective procedures” [63]. In Singapore, nonurgent medical appointments have been rescheduled and staff have been redeployed to help manage patients with COVID-19 [64]. To mitigate the impact on non-COVID-19 patients, AI can be used to augment their care.

With regard to medical investigations, AI-based radiological interpretation algorithms can be applied to read scans for non-COVID-19 diseases. With regard to medical consultations, AI-based conversational chatbots can assume some of the duties of the physician, such as symptom screening and patient education. Chatbots are already being used to combat the pandemic. Symptoma, a symptom-to-disease digital health assistant using AI, has been shown to be highly accurate when screening for COVID-19 [65]. Similarly, chatbots have been used to rapidly screen health care workers for COVID-19, facilitating staff movements and minimizing nosocomial transmission [66]. By extension, AI-based chatbots can provide a platform for patients with non-COVID-19 conditions to receive medical care at a time when clinical resources are limited. These chatbots can potentially be augmented by smartwatch-based health monitors (eg, heart rate and electrocardiogram monitoring using the Apple Watch [67]).

### **Clinical Management of Patients With COVID-19**

Studies have already established that AI can help guide management of patients with COVID-19 in general practice.

AI can also potentially help in the management of critically ill patients with COVID-19. For instance, given the uncertainty over the optimal management of COVID-19-related acute respiratory distress syndrome (ARDS) [68], AI methods like reinforcement learning could be used to determine management choices to achieve the best possible clinical outcomes [69]. Besides therapeutics, machine learning strategies can be used for vaccine development against COVID-19. They have already been applied to SARS-CoV-2 proteomes, and nonstructural proteins were found to be potential vaccine candidates [70].

### **Conclusion**

AI is no longer new in the field of medicine, and many studies have explored how its potential to enhance medical care of patients could be realized. Even as we see the situation improving in some countries, others are still struggling to contain the spread of COVID-19. In the face of growing pressure on limited health care resources, the use of AI-driven techniques to aid in diagnosis, surveillance, finding therapeutics, and public health decision making may help improve the efficiency and effectiveness of human efforts to combat the pandemic. Another recently published review found that AI could contribute to the prevention and control of the spread of COVID-19 via several important approaches: detecting suspected cases, large-scale screening, monitoring, determining interactions with experimental therapies, pneumonia screening, using the Internet of Intelligent Things for data and information gathering and integration, allocating resources; making predictions, models, and simulations; and using robotics for medical quarantine [71]. We hope that our rapid review can help highlight additional areas for more robust AI applications and studies in the later phases of the ongoing pandemic.

---

### **Acknowledgments**

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work. JYC conducted the data collection and analysis, and was responsible for the drafting of the manuscript. KCS was responsible for the conception and design of the study and oversaw the critical revision of the article. The authors received no specific funding.

---

### **Conflicts of Interest**

None declared.

---

### **References**

1. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020 Feb 03;579(7798):270-273. [doi: [10.1038/s41586-020-2012-7](https://doi.org/10.1038/s41586-020-2012-7)]
2. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. *J Am Med Inform Assoc* 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: [10.1093/jamia/ocaa037](https://doi.org/10.1093/jamia/ocaa037)] [Medline: [32208481](https://pubmed.ncbi.nlm.nih.gov/32208481/)]
3. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *N Engl J Med* 2020 May 21;382(21):2049-2055 [FREE Full text] [doi: [10.1056/NEJMs2005114](https://doi.org/10.1056/NEJMs2005114)] [Medline: [32202722](https://pubmed.ncbi.nlm.nih.gov/32202722/)]
4. Yassine HM, Shah Z. How could artificial intelligence aid in the fight against coronavirus? *Expert Rev Anti Infect Ther* 2020 Jun 29;18(6):493-497. [doi: [10.1080/14787210.2020.1744275](https://doi.org/10.1080/14787210.2020.1744275)] [Medline: [32223349](https://pubmed.ncbi.nlm.nih.gov/32223349/)]

5. Mashamba-Thompson TP, Crayton ED. Blockchain and Artificial Intelligence Technology for Novel Coronavirus Disease-19 Self-Testing. *Diagnostics* (Basel) 2020 Apr 01;10(4):198 [FREE Full text] [doi: [10.3390/diagnostics10040198](https://doi.org/10.3390/diagnostics10040198)] [Medline: [32244841](https://pubmed.ncbi.nlm.nih.gov/32244841/)]
6. Siow WT, Liew MF, Shrestha BR, Muchtar F, See KC. Managing COVID-19 in resource-limited settings: critical care considerations. *Crit Care* 2020 Apr 22;24(1):167 [FREE Full text] [doi: [10.1186/s13054-020-02890-x](https://doi.org/10.1186/s13054-020-02890-x)] [Medline: [32321566](https://pubmed.ncbi.nlm.nih.gov/32321566/)]
7. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. *Nat Med* 2020 Apr 27;26(4):459-461 [FREE Full text] [doi: [10.1038/s41591-020-0824-5](https://doi.org/10.1038/s41591-020-0824-5)] [Medline: [32284618](https://pubmed.ncbi.nlm.nih.gov/32284618/)]
8. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism* 2017 Apr;69S:S36-S40. [doi: [10.1016/j.metabol.2017.01.011](https://doi.org/10.1016/j.metabol.2017.01.011)] [Medline: [28126242](https://pubmed.ncbi.nlm.nih.gov/28126242/)]
9. Hashimoto D, Witkowski E, Gao L, Meireles O, Rosman G. Artificial Intelligence in Anesthesiology: Current Techniques, Clinical Applications, and Limitations. *Anesthesiology* 2020 Feb;132(2):379-394 [FREE Full text] [doi: [10.1097/ALN.0000000000002960](https://doi.org/10.1097/ALN.0000000000002960)] [Medline: [31939856](https://pubmed.ncbi.nlm.nih.gov/31939856/)]
10. Long JB, Ehrenfeld JM. The Role of Augmented Intelligence (AI) in Detecting and Preventing the Spread of Novel Coronavirus. *J Med Syst* 2020 Feb 04;44(3):59 [FREE Full text] [doi: [10.1007/s10916-020-1536-6](https://doi.org/10.1007/s10916-020-1536-6)] [Medline: [32020374](https://pubmed.ncbi.nlm.nih.gov/32020374/)]
11. Kallianos K, Mongan J, Antani S, Henry T, Taylor A, Abuya J, et al. How far have we come? Artificial intelligence for chest radiograph interpretation. *Clin Radiol* 2019 May;74(5):338-345. [doi: [10.1016/j.crad.2018.12.015](https://doi.org/10.1016/j.crad.2018.12.015)] [Medline: [30704666](https://pubmed.ncbi.nlm.nih.gov/30704666/)]
12. Deo RC. Machine Learning in Medicine. *Circulation* 2015 Nov 17;132(20):1920-1930. [doi: [10.1161/CIRCULATIONAHA.115.001593](https://doi.org/10.1161/CIRCULATIONAHA.115.001593)] [Medline: [26572668](https://pubmed.ncbi.nlm.nih.gov/26572668/)]
13. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *J Travel Med* 2020 Mar 13;27(2):1 [FREE Full text] [doi: [10.1093/jtm/taaa008](https://doi.org/10.1093/jtm/taaa008)] [Medline: [31943059](https://pubmed.ncbi.nlm.nih.gov/31943059/)]
14. Bowles J. How Canadian AI start-up BlueDot spotted Coronavirus before anyone else had a clue. *Diginomica*. 2020 Mar 10. URL: <https://diginomica.com/how-canadian-ai-start-bluedot-spotted-coronavirus-anyone-else-had-clue> [accessed 2020-10-09]
15. Coronavirus: Singapore chatbot helps doctors monitor migrant workers' health in real time. *South China Morning Post*. 2020 Jun 11. URL: <https://www.scmp.com/news/asia/southeast-asia/article/3088547/coronavirus-singapore-chatbot-helps-doctors-monitor> [accessed 2020-10-09]
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009 Jul 21;339(jul21 1):b2535-b2535. [doi: [10.1136/bmj.b2535](https://doi.org/10.1136/bmj.b2535)] [Medline: [19622551](https://pubmed.ncbi.nlm.nih.gov/19622551/)]
17. Moons KHM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015 Jan 06;162(1):W1-73. [doi: [10.7326/M14-0698](https://doi.org/10.7326/M14-0698)] [Medline: [25560730](https://pubmed.ncbi.nlm.nih.gov/25560730/)]
18. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019 Jan 01;170(1):W1-W33. [doi: [10.7326/M18-1377](https://doi.org/10.7326/M18-1377)] [Medline: [30596876](https://pubmed.ncbi.nlm.nih.gov/30596876/)]
19. Hurt B, Kligerman S, Hsiao A. Deep Learning Localization of Pneumonia: 2019 Coronavirus (COVID-19) Outbreak. *J Thorac Imaging* 2020 May;35(3):W87-W89 [FREE Full text] [doi: [10.1097/RTI.0000000000000512](https://doi.org/10.1097/RTI.0000000000000512)] [Medline: [32205822](https://pubmed.ncbi.nlm.nih.gov/32205822/)]
20. Li D, Wang D, Dong J, Wang N, Huang H, Xu H, et al. False-Negative Results of Real-Time Reverse-Transcriptase Polymerase Chain Reaction for Severe Acute Respiratory Syndrome Coronavirus 2: Role of Deep-Learning-Based CT Diagnosis and Insights from Two Cases. *Korean J Radiol* 2020 Apr;21(4):505-508 [FREE Full text] [doi: [10.3348/kjr.2020.0146](https://doi.org/10.3348/kjr.2020.0146)] [Medline: [32174053](https://pubmed.ncbi.nlm.nih.gov/32174053/)]
21. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* 2020 Aug 19;296(2):E65-E71 [FREE Full text] [doi: [10.1148/radiol.2020200905](https://doi.org/10.1148/radiol.2020200905)] [Medline: [32191588](https://pubmed.ncbi.nlm.nih.gov/32191588/)]
22. Yang Z, Zeng Z, Wang K, Wong S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* 2020 Mar;12(3):165-174 [FREE Full text] [doi: [10.21037/jtd.2020.02.64](https://doi.org/10.21037/jtd.2020.02.64)] [Medline: [32274081](https://pubmed.ncbi.nlm.nih.gov/32274081/)]
23. Al-Najjar HN, Al-Rousan N. A classifier prediction model to predict the status of Coronavirus COVID-19 patients in South Korea. *Eur Rev Med Pharmacol Sci* 2020 Mar;24(6):3400-3403 [FREE Full text] [doi: [10.26355/eurrev\\_202003\\_20709](https://doi.org/10.26355/eurrev_202003_20709)] [Medline: [32271458](https://pubmed.ncbi.nlm.nih.gov/32271458/)]
24. Jiang X, Coffee M, Bari A. Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity. 2020 Mar 30;63(1):537-551. [doi: [10.32604/cmc.2020.010691](https://doi.org/10.32604/cmc.2020.010691)]
25. Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Comput Struct Biotechnol J* 2020;18:784-790 [FREE Full text] [doi: [10.1016/j.csbj.2020.03.025](https://doi.org/10.1016/j.csbj.2020.03.025)] [Medline: [32280433](https://pubmed.ncbi.nlm.nih.gov/32280433/)]
26. Kadioglu O, Saeed M, Gretten HJ, Efferth T. Identification of novel compounds against three targets of SARS CoV-2 coronavirus by combined virtual screening and supervised machine learning. *Bull World Health Organ* 2020 Mar 21:1-29 [FREE Full text] [doi: [10.2471/blt.20.255943](https://doi.org/10.2471/blt.20.255943)]

27. Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *The Lancet* 2020 Feb;395(10223):e30-e31. [doi: [10.1016/s0140-6736\(20\)30304-4](https://doi.org/10.1016/s0140-6736(20)30304-4)]
28. Ton A, Gentile F, Hsing M, Ban F, Cherkasov A. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. *Mol Inform* 2020 Aug 11;39(8):e2000028 [FREE Full text] [doi: [10.1002/minf.202000028](https://doi.org/10.1002/minf.202000028)] [Medline: [32162456](https://pubmed.ncbi.nlm.nih.gov/32162456/)]
29. Zhang D, Wu K, Zhang X, Deng S, Peng B. In silico screening of Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus. *J Integr Med* 2020 Mar;18(2):152-158 [FREE Full text] [doi: [10.1016/j.joim.2020.02.005](https://doi.org/10.1016/j.joim.2020.02.005)] [Medline: [32113846](https://pubmed.ncbi.nlm.nih.gov/32113846/)]
30. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* 2018 Oct 10;2(10):719-731. [doi: [10.1038/s41551-018-0305-z](https://doi.org/10.1038/s41551-018-0305-z)] [Medline: [31015651](https://pubmed.ncbi.nlm.nih.gov/31015651/)]
31. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: [10.1001/jama.2020.1585](https://doi.org/10.1001/jama.2020.1585)] [Medline: [32031570](https://pubmed.ncbi.nlm.nih.gov/32031570/)]
32. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* 2020 Aug 19;296(2):E115-E117 [FREE Full text] [doi: [10.1148/radiol.2020200432](https://doi.org/10.1148/radiol.2020200432)] [Medline: [32073353](https://pubmed.ncbi.nlm.nih.gov/32073353/)]
33. Li Y, Xia L. Coronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management. *American Journal of Roentgenology* 2020 Jun;214(6):1280-1286. [doi: [10.2214/ajr.20.22954](https://doi.org/10.2214/ajr.20.22954)]
34. Ramesh A, Kambhampati C, Monson J, Drew P. Artificial intelligence in medicine. *Ann R Coll Surg Engl* 2004 Sep 01;86(5):334-338 [FREE Full text] [doi: [10.1308/147870804290](https://doi.org/10.1308/147870804290)] [Medline: [15333167](https://pubmed.ncbi.nlm.nih.gov/15333167/)]
35. Santosh KC. AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data. *J Med Syst* 2020 Mar 18;44(5):93 [FREE Full text] [doi: [10.1007/s10916-020-01562-1](https://doi.org/10.1007/s10916-020-01562-1)] [Medline: [32189081](https://pubmed.ncbi.nlm.nih.gov/32189081/)]
36. European Centre for Disease Prevention and Control. How ECDC collects and processes COVID-19 data. URL: <https://www.ecdc.europa.eu/en/covid-19/data-collection> [accessed 2020-10-09]
37. Ng Y, Li Z, Chua YX, Chaw WL, Zhao Z, Er B, et al. Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore - January 2-February 29, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Mar 20;69(11):307-311 [FREE Full text] [doi: [10.15585/mmwr.mm6911e1](https://doi.org/10.15585/mmwr.mm6911e1)] [Medline: [32191691](https://pubmed.ncbi.nlm.nih.gov/32191691/)]
38. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014 Mar 14;343(6176):1203-1205. [doi: [10.1126/science.1248506](https://doi.org/10.1126/science.1248506)] [Medline: [24626916](https://pubmed.ncbi.nlm.nih.gov/24626916/)]
39. Lai Y, Yeung W, Celi LA. Urban Intelligence for Pandemic Response: Viewpoint. *JMIR Public Health Surveill* 2020 Apr 14;6(2):e18873 [FREE Full text] [doi: [10.2196/18873](https://doi.org/10.2196/18873)] [Medline: [32248145](https://pubmed.ncbi.nlm.nih.gov/32248145/)]
40. Crawford K. The hidden biases in big data. *Harvard Business Review*. 2013. URL: <https://hbr.org/2013/04/the-hidden-biases-in-big-data> [accessed 2020-10-09]
41. Centers for Disease Control and Prevention. About CDC COVID-19 Data. 2020 Jul 13. URL: <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/about-us-cases-deaths.html> [accessed 2020-10-09]
42. CDC COVID-19 Response Team. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 - United States, February 12-March 28, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Apr 03;69(13):382-386 [FREE Full text] [doi: [10.15585/mmwr.mm6913e2](https://doi.org/10.15585/mmwr.mm6913e2)] [Medline: [32240123](https://pubmed.ncbi.nlm.nih.gov/32240123/)]
43. Neri E, Miele V, Coppola F, Grassi R. Use of CT and artificial intelligence in suspected or COVID-19 positive patients: statement of the Italian Society of Medical and Interventional Radiology. *Radiol Med* 2020 May 29;125(5):505-508. [doi: [10.1007/s11547-020-01197-9](https://doi.org/10.1007/s11547-020-01197-9)] [Medline: [32350794](https://pubmed.ncbi.nlm.nih.gov/32350794/)]
44. Vanderslott S, Pollard A, Thomas T. Coronavirus vaccine: here are the steps it will need to go through during development. *The Conversation*. URL: <https://theconversation.com/coronavirus-vaccine-here-are-the-steps-it-will-need-to-go-through-during-development-134726> [accessed 2020-10-09]
45. Khamsi R. If a coronavirus vaccine arrives, can the world make enough? *Nature* 2020 Apr 9;580(7805):578-580. [doi: [10.1038/d41586-020-01063-8](https://doi.org/10.1038/d41586-020-01063-8)] [Medline: [32273621](https://pubmed.ncbi.nlm.nih.gov/32273621/)]
46. Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, et al. Artificial Intelligence Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Origin at Chest CT. *Radiology* 2020 Sep;296(3):E156-E165 [FREE Full text] [doi: [10.1148/radiol.2020201491](https://doi.org/10.1148/radiol.2020201491)] [Medline: [32339081](https://pubmed.ncbi.nlm.nih.gov/32339081/)]
47. Maru DS, Schwarz R, Jason A, Basu S, Sharma A, Moore C. Turning a blind eye: the mobilization of radiology services in resource-poor regions. *Global Health* 2010 Oct 14;6(1):18 [FREE Full text] [doi: [10.1186/1744-8603-6-18](https://doi.org/10.1186/1744-8603-6-18)] [Medline: [20946643](https://pubmed.ncbi.nlm.nih.gov/20946643/)]
48. Naudé W. Artificial intelligence vs COVID-19: limitations, constraints and pitfalls. *AI Soc* 2020 Apr 28;35(3):1-5 [FREE Full text] [doi: [10.1007/s00146-020-00978-0](https://doi.org/10.1007/s00146-020-00978-0)] [Medline: [32346223](https://pubmed.ncbi.nlm.nih.gov/32346223/)]
49. Santow E. Emerging from AI utopia. *Science* 2020 Apr 03;368(6486):9-9. [doi: [10.1126/science.abb9369](https://doi.org/10.1126/science.abb9369)] [Medline: [32241926](https://pubmed.ncbi.nlm.nih.gov/32241926/)]
50. McCradden MD, Baba A, Saha A, Ahmad S, Boparai K, Fadaiefard P, et al. Ethical concerns around use of artificial intelligence in health care research from the perspective of patients with meningioma, caregivers and health care providers: a qualitative study. *CMAJ Open* 2020 Feb 18;8(1):E90-E95. [doi: [10.9778/cmajo.20190151](https://doi.org/10.9778/cmajo.20190151)] [Medline: [32071143](https://pubmed.ncbi.nlm.nih.gov/32071143/)]

51. Braga A, Logan R. The Emperor of Strong AI Has No Clothes: Limits to Artificial Intelligence. *Information* 2017 Nov 27;8(4):156. [doi: [10.3390/info8040156](https://doi.org/10.3390/info8040156)]
52. Centers for Disease Control and Prevention. COVID-19 in Racial and Ethnic Minority Groups. 2020 Jul 24. URL: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html> [accessed 2020-10-09]
53. Selden TM, Berdahl TA. COVID-19 And Racial/Ethnic Disparities In Health Risk, Employment, And Household Composition. *Health Aff (Millwood)* 2020 Sep 01;39(9):1624-1632. [doi: [10.1377/hlthaff.2020.00897](https://doi.org/10.1377/hlthaff.2020.00897)] [Medline: [32663045](https://pubmed.ncbi.nlm.nih.gov/32663045/)]
54. Chun A. In a time of coronavirus, China's investment in AI is paying off in a big way. *South China Morning Post*. 2020 Mar 18. URL: <https://www.scmp.com/comment/opinion/article/3075553/time-coronavirus-chinas-investment-ai-paying-big-way> [accessed 2020-10-09]
55. Maslan C. Social distancing detection for COVID-19. URL: <https://blog.camio.com/social-distancing-detection-for-covid-19-a5c9e8d6c760> [accessed 2020-10-09]
56. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends Analysis. *Int J Infect Dis* 2020 Jun;95:192-197 [FREE Full text] [doi: [10.1016/j.ijid.2020.04.033](https://doi.org/10.1016/j.ijid.2020.04.033)] [Medline: [32305520](https://pubmed.ncbi.nlm.nih.gov/32305520/)]
57. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digit Health* 2020 Apr;2(4):e201-e208 [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30026-1](https://doi.org/10.1016/S2589-7500(20)30026-1)] [Medline: [32309796](https://pubmed.ncbi.nlm.nih.gov/32309796/)]
58. Zhavoronkov A, Aladinskiy V, Zhebrak A. Potential COVID-2019 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches. *ChemRxiv* 2020 Feb 18:1. [doi: [10.26434/chemrxiv.11829102.v2](https://doi.org/10.26434/chemrxiv.11829102.v2)]
59. Heaven WD. AI could help with the next pandemic—but not with this one. *MIT Technology Review*. 2020 Mar 12. URL: <https://www.technologyreview.com/2020/03/12/905352/ai-could-help-with-the-next-pandemic-but-not-with-this-one/> [accessed 2020-10-09]
60. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boëlle P, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *The Lancet* 2020 Mar;395(10227):871-877. [doi: [10.1016/s0140-6736\(20\)30411-6](https://doi.org/10.1016/s0140-6736(20)30411-6)]
61. Lai S, Bogoch I, Ruktanonchai N, Watts A, Lu X, Yang W, et al. Assessing spread risk of Wuhan novel coronavirus within and beyond China, January-April 2020: a travel network-based modelling study. *medRxiv* 2020 Feb 05:0479 [FREE Full text] [doi: [10.1101/2020.02.04.20020479](https://doi.org/10.1101/2020.02.04.20020479)] [Medline: [32511631](https://pubmed.ncbi.nlm.nih.gov/32511631/)]
62. De Salazar PM, Niehus R, Taylor A, Buckee C, Lipsitch M. Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases. *medRxiv* 2020 Feb 11:0495 [FREE Full text] [doi: [10.1101/2020.02.04.20020495](https://doi.org/10.1101/2020.02.04.20020495)] [Medline: [32511458](https://pubmed.ncbi.nlm.nih.gov/32511458/)]
63. COVID-19: Recommendations for management of elective surgical procedures. *American College of Surgeons*. 2020 Mar 13. URL: <https://www.facs.org/covid-19/clinical-guidance/elective-surgery> [accessed 2020-10-09]
64. COVID-19: Essential healthcare services to continue operating, other services to be scaled down. *Channel News Asia*. 2020 Apr 04. URL: <https://www.channelnewsasia.com/news/singapore/healthcare-services-clinics-closed-covid-19-coronavirus-12609404> [accessed 2020-10-09]
65. Martin A, Nateqi J, Gruarin S, Munsch N, Abdurahmane I, Knapp B. An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot. *bioRxiv* 2020 Apr 06:1. [doi: [10.1101/2020.03.25.008805](https://doi.org/10.1101/2020.03.25.008805)]
66. Judson T, Odisho A, Young J, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. *J Am Med Inform Assoc* 2020 Jul 01;27(9):1450-1455 [FREE Full text] [doi: [10.1093/jamia/ocaa130](https://doi.org/10.1093/jamia/ocaa130)] [Medline: [32531066](https://pubmed.ncbi.nlm.nih.gov/32531066/)]
67. Owen M. Apple Watch ECG detects heart condition in German woman. *Apple Insider*. 2020. URL: <https://appleinsider.com/articles/20/05/03/apple-watch-ecg-detects-heart-condition-in-german-woman> [accessed 2020-10-09]
68. Gattinoni L, Chiumello D, Rossi S. COVID-19 pneumonia: ARDS or not? *Crit Care* 2020 Apr 16;24(1):154 [FREE Full text] [doi: [10.1186/s13054-020-02880-z](https://doi.org/10.1186/s13054-020-02880-z)] [Medline: [32299472](https://pubmed.ncbi.nlm.nih.gov/32299472/)]
69. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* 2018 Nov 22;24(11):1716-1720. [doi: [10.1038/s41591-018-0213-5](https://doi.org/10.1038/s41591-018-0213-5)] [Medline: [30349085](https://pubmed.ncbi.nlm.nih.gov/30349085/)]
70. Ong E, Wong MU, Huffman A, He Y. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *bioRxiv* 2020 Mar 21:1 [FREE Full text] [doi: [10.1101/2020.03.20.000141](https://doi.org/10.1101/2020.03.20.000141)] [Medline: [32511333](https://pubmed.ncbi.nlm.nih.gov/32511333/)]
71. Adly AS, Adly AS, Adly MS. Approaches Based on Artificial Intelligence and the Internet of Intelligent Things to Prevent the Spread of COVID-19: Scoping Review. *J Med Internet Res* 2020 Aug 10;22(8):e19104 [FREE Full text] [doi: [10.2196/19104](https://doi.org/10.2196/19104)] [Medline: [32584780](https://pubmed.ncbi.nlm.nih.gov/32584780/)]

## Abbreviations

**AI:** artificial intelligence

**CDC:** Centers for Disease Control and Prevention



**CT:** computed tomography

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**PROBAST:** prediction model risk of bias assessment tool

**RSV:** relative search volume

**TCM:** Traditional Chinese Medicine

**TRIPOD:** transparent reporting of a multivariable prediction model for individual prognosis or diagnosis

*Edited by G Eysenbach, G Fagherazzi; submitted 16.06.20; peer-reviewed by M Coffee, D Gunasekaran; comments to author 13.07.20; revised version received 25.07.20; accepted 15.09.20; published 27.10.20.*

*Please cite as:*

*Chen J, See KC*

*Artificial Intelligence for COVID-19: Rapid Review*

*J Med Internet Res 2020;22(10):e21476*

URL: <http://www.jmir.org/2020/10/e21476/>

doi: [10.2196/21476](https://doi.org/10.2196/21476)

PMID: [32946413](https://pubmed.ncbi.nlm.nih.gov/32946413/)

©Jiayang Chen, Kay Choong See. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 27.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing

Jose Luis Izquierdo<sup>1</sup>, MD; Julio Ancochea<sup>2</sup>, MD; Savana COVID-19 Research Group<sup>3</sup>; Joan B Soriano<sup>2</sup>, MD, PhD

<sup>1</sup>Hospital Universitario de Guadalajara, Guadalajara, Spain

<sup>2</sup>Hospital Universitario de La Princesa, Madrid, Spain

<sup>3</sup>Savana, Madrid, Spain

**Corresponding Author:**

Joan B Soriano, MD, PhD

Hospital Universitario de La Princesa

Diego de León 62

Madrid, 28005

Spain

Phone: 34 618867769

Email: [jbsoriano2@gmail.com](mailto:jbsoriano2@gmail.com)

## Abstract

**Background:** Many factors involved in the onset and clinical course of the ongoing COVID-19 pandemic are still unknown. Although big data analytics and artificial intelligence are widely used in the realms of health and medicine, researchers are only beginning to use these tools to explore the clinical characteristics and predictive factors of patients with COVID-19.

**Objective:** Our primary objectives are to describe the clinical characteristics and determine the factors that predict intensive care unit (ICU) admission of patients with COVID-19. Determining these factors using a well-defined population can increase our understanding of the real-world epidemiology of the disease.

**Methods:** We used a combination of classic epidemiological methods, natural language processing (NLP), and machine learning (for predictive modeling) to analyze the electronic health records (EHRs) of patients with COVID-19. We explored the unstructured free text in the EHRs within the Servicio de Salud de Castilla-La Mancha (SESCAM) Health Care Network (Castilla-La Mancha, Spain) from the entire population with available EHRs (1,364,924 patients) from January 1 to March 29, 2020. We extracted related clinical information regarding diagnosis, progression, and outcome for all COVID-19 cases.

**Results:** A total of 10,504 patients with a clinical or polymerase chain reaction–confirmed diagnosis of COVID-19 were identified; 5519 (52.5%) were male, with a mean age of 58.2 years (SD 19.7). Upon admission, the most common symptoms were cough, fever, and dyspnea; however, all three symptoms occurred in fewer than half of the cases. Overall, 6.1% (83/1353) of hospitalized patients required ICU admission. Using a machine-learning, data-driven algorithm, we identified that a combination of age, fever, and tachypnea was the most parsimonious predictor of ICU admission; patients younger than 56 years, without tachypnea, and temperature <39 degrees Celsius (or >39 °C without respiratory crackles) were not admitted to the ICU. In contrast, patients with COVID-19 aged 40 to 79 years were likely to be admitted to the ICU if they had tachypnea and delayed their visit to the emergency department after being seen in primary care.

**Conclusions:** Our results show that a combination of easily obtainable clinical variables (age, fever, and tachypnea with or without respiratory crackles) predicts whether patients with COVID-19 will require ICU admission.

(*J Med Internet Res* 2020;22(10):e21801) doi:[10.2196/21801](https://doi.org/10.2196/21801)

**KEYWORDS**

artificial intelligence; big data; COVID-19; electronic health records; tachypnea; SARS-CoV-2; predictive model

## Introduction

The unprecedented global spread of SARS-CoV-2, the virus that causes COVID-19, requires innovative approaches that deliver real-time results [1,2]. To date, big data analytics have been primarily used to assess SARS-CoV-2 transmission [3] and to indirectly estimate COVID-19 incidence using data from social media [4]. However, many factors involved in the onset and temporal distribution of the ongoing COVID-19 pandemic remain unknown. Similarly, both the individual and population burdens of COVID-19 are only beginning to be elucidated. Although big data analytics and artificial intelligence (AI) are widely used in the realms of health and medicine [5-7], researchers are only beginning to use these tools to explore the clinical characteristics and predictive factors of patients with COVID-19, including mortality [8-11].

Considering the unprecedented spread and severity of the ongoing COVID-19 outbreak, focus has been given to hospitals' unmet needs, particularly their ICU requirements [8,9,12]. Indeed, health systems have been or currently are near collapse, and independent modelling efforts have aimed at forecasting a number of epidemiological estimators, including ICU use [13-15].

Previously, our team reported that a combination of big data analytics and machine learning techniques helped better determine the quality of diagnosis and treatment of chronic obstructive pulmonary disease (COPD) via an analysis of hospital electronic health records (EHRs) using natural language processing (NLP) and validated algorithms [16,17].

As part of the BigCOVIData study, our primary objectives are to describe the clinical characteristics and determine the factors that predict ICU admission of patients with COVID-19. Determining these factors using a well-defined population can increase our understanding of the real-world epidemiology of the disease. To achieve this aim, we used a combination of classic epidemiological methods [18], NLP, and machine

learning (for predictive modeling) to analyze the clinical information contained in the EHRs of patients with COVID-19.

## Methods

The BigCOVIData study was conducted in compliance with legal and regulatory requirements and followed generally accepted research practices described in the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) Guideline for Good Clinical Practice, the latest edition of the Helsinki Declaration, the Guidelines for Good Pharmacoepidemiology Practices, and applicable local regulations. This study was classified as a "non-postauthorization study" by the Spanish Agency of Medicines and Health Products, and it was approved by the Research Ethics Committee at the University Hospital of Guadalajara (Spain). We and endorsed the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidance for reporting observational research [19].

### Study Design and Data Source

This was a multicenter, noninterventional, retrospective study using data captured in the EHRs of the participating hospitals within the Servicio de Salud de Castilla-La Mancha (SESCAM) Health Care Network in Castilla-La Mancha, Spain (Figure 1). Data captured in EHRs were collected from all available departments, including inpatient hospital, outpatient hospital, and emergency department (ED), for virtually all types of provided services in each participating hospital. The study period was January 1 to March 29, 2020.

The study database was fully anonymized in a structured format and contained no personal information from patients. Likewise, personal information was not accessed during either the application of automated and algorithmic methods (ie, NLP) or the conversion of unstructured data into the structured database. Importantly, given that clinical information was handled in an aggregate, anonymized, and irreversibly dissociated manner, patient consent regulations do not apply to the present study.

**Figure 1.** Map of the Castilla-La Mancha region (red) within the Spanish (blue line) and European territories. From a general source population of 2,035,000 inhabitants, we collected and analyzed the clinical information in the EHRs of 1,364,924 patients within the Servicio de Salud de Castilla-La Mancha (SESCAM) Health Care Network. EHR: electronic health record.



## Study Sample

The study sample included all patients in the source population who were diagnosed with COVID-19. Patients were identified on the basis of clinical diagnosis or microbiological test results. Clinical confirmation of COVID-19 cases was determined by observed symptomatology, imaging (mostly chest X-ray), and laboratory results, as captured in the unstructured, free-text information in the EHRs. Microbiological test result confirmation of COVID-19 cases involved reverse transcriptase–polymerase chain reaction (RT-PCR) or similar available tests. Our decision to consider cases confirmed both clinically and by RT-PCR was justified by the limited availability of routinely administered RT-PCR tests in the region during the study period and supported by recent discussions on the far-from-optimal sensitivity of RT-PCR for COVID-19 (ie, a single negative result from a single specimen cannot exclude the disease in suspected cases) [20,21]. Indeed, recent reports highlight the clinical validity and relatively high sensitivity of

symptom- and imaging-based identification of patients with COVID-19, especially in early stages of the disease [20,22,23].

## EHRead

To meet the study objectives, we used EHRead [24], a technology developed by Savana that applies NLP, machine learning, and deep learning to analyze the unstructured free-text information written in millions of deidentified EHRs. This technology enables the extraction of information from all types of EHRs and subsequent normalization of the extracted clinical entities to a unique terminology. This process enables further analysis of a descriptive or predictive nature. Originally based on Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) terminology, our unique body of terminology comprises more than 400,000 medical concepts, acronyms, and laboratory parameters aggregated over the course of five years of free-text mining, targeting the most common diseases (eg, respiratory diseases, cardiovascular diseases, and diabetes).

Using a combination of regular expression rules and machine learning models, the terminology entities are detected in the unstructured text and later classified based on sections typically contained in EHRs, hospital services, and other clinical specifications. Importantly, each detected term is described in terms of negative, speculative, or affirmative clinical statements; this is achieved by using deep learning convolutional neural network classification methods that rely on word embeddings and context information (for a similar methodological approach, see [25]). The limitations of case-by-case detection are also overcome with a similar approach to ensure that the detected concepts are used within the appropriate context for the descriptive and predictive analysis.

For particular cases where extra specifications are required (ie, to differentiate COVID-19 cases from other mentions of the term related to fear of the disease or to potential contact), the detection output was manually reviewed in more than 5000 reports to avoid any possible ambiguity associated with free-text reporting. All NLP deep learning models used in this study were validated using the standard training/validation/testing approach; we used a 75/12/13 split ratio in the available annotated data (between 2000 and 3000 records, depending on the model) to ensure efficient generalization on unseen cases. For all developed models, we obtained F scores greater than 0.89.

### Data Analyses

All categorical variables (eg, comorbidities, symptoms) are shown in frequency tables, whereas continuous variables (eg, age) are described via summary tables that include the mean, SD, median, minimum and maximum, and quartiles of each variable. To test for possible statistically significant differences in the distribution of categorical variables between study groups (ie, male vs female, ICU admission vs no ICU admission), we used Yates-corrected chi-square tests. For continuous variables, mean differences were tested using *t* tests. Given our general population approach and our unusually large sample size, we were interested in exploring sex-related differences in patients with COVID-19; therefore, most results were stratified by sex [26]. All statistical inferences were performed at the 5% significance level using 2-sided tests or 2-sided CIs.

### Predictive Model

We developed a decision tree to classify patients with COVID-19 according to their risk of being admitted to the ICU. The two types of patients or *classes* considered in the model were therefore “admitted to the ICU” and “not admitted to the ICU.” The model maps the characteristics of the patients (the *variables*) to their class in the shape of a tree. From a clinical perspective, this model contemplates all patient variables upon

admission; therefore, it is predictive from symptom debut until outcome. The tree is composed of nodes that branch to subsequent child nodes depending on the patient’s variables. The tree is built in such a way that each branch separates the two classes as much as possible. This separation is measured as the Shannon entropy, where a node with an entropy of zero indicates that the classification is perfect (either all or none of the patients were admitted to the ICU) and an entropy of one is the worst possible mix (50%/50%) [27].

### Model Training and Validation

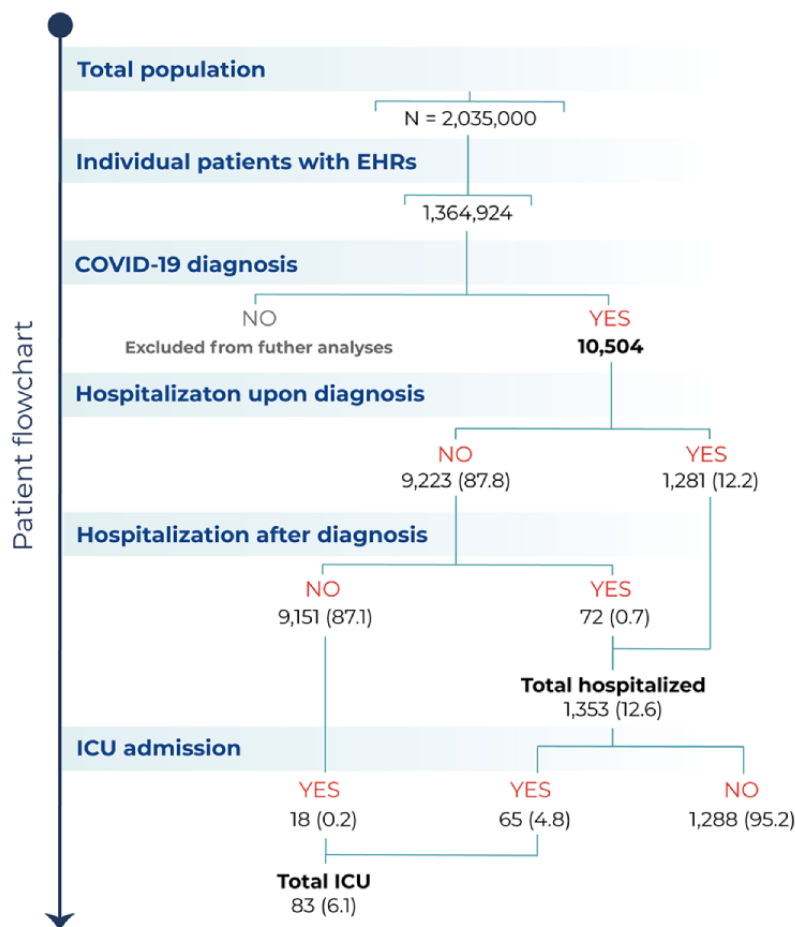
The model was developed and tested on the available data from hospitalized patients who had or had not been admitted to the ICU; the latter were either discharged from the hospital or died in the course of the disease. This amounted to a total of 900 patients. We validated our algorithm by splitting our COVID-19 sample into a 70% training set and a 30% validation set. This means that the model was trained with 630 patients (582 who did not require intensive care vs 48 who did) and validated over the remaining 270 patients. Because the two classes were unbalanced (far fewer patients required ICU admission), we used the standard technique of oversampling the lower class to guarantee a balance of accuracy and recall (ie, the tradeoff between false positives vs false negatives). Further, we sought to replicate the results of this validation in an a posteriori sensitivity analysis, as per recent recommendations for predictive modeling in COVID-19 [28] and TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guidance [29]. For this second validation, we trained the model with data from the provinces of Ciudad Real and Guadalajara (38% of the study sample from Castilla-La Mancha), and we used an independent set with combined data set from the other three provinces, namely Toledo, Cuenca, and Albacete, for validation.

Additional details regarding the development and validation of the predictive algorithm are included in the Supplementary Methods in [Multimedia Appendix 1](#). The workflow used for the generation of the predictive algorithm is summarized in Figure S1 in [Multimedia Appendix 1](#).

## Results

From a source general population of 2,035,000 inhabitants, we used NLP and machine learning to analyze the clinical information contained in the EHRs of 1,364,924 anonymous patients ([Figure 1](#)). Among these, we identified a total of 10,504 patients diagnosed with COVID-19 ([Figure 2](#)). The flowchart of participation in the study up to hospital admission, ICU admission, or discharge is presented in [Figure 2](#).

**Figure 2.** Patient flowchart depicting the total number of inhabitants in the source population, the number (%) of patients with available EHRs analyzed, the number of patients diagnosed with COVID-19, and of those, the number of hospitalizations and ICU admissions. EHR: electronic health record; ICU: intensive care unit.



Of the patients with COVID-19, 52.5% (5519/10,504) were male, with a mean age of 58.2 years (SD 19.7) (Table 1). Most patients with COVID-19 were aged  $\geq 50$  years (Figure 3). Upon diagnosis, the most common symptoms reported were cough, fever, and dyspnea (Table 1); notably, less than half of patients presented with all three of these symptoms. Further, respiratory crackles, myalgia, and diarrhea were identified in  $\geq 5\%$  of cases, while other respiratory and nonrespiratory signs and symptoms

were less common. Sex-dependent differences in sign and symptom frequencies upon diagnosis are shown in Table 1. Of note, we observed subtle increases in the frequency of diarrhea, myalgia, headache, chest pain, and anosmia in female patients with COVID-19, while male patients showed significant increases in fever, dyspnea, respiratory crackles, rhonchus, lymphopenia, and tachypnea (all  $P < .05$ ).



**Table 1.** Baseline demographics and clinical data of the patients in the study upon diagnosis (N=10,504).

Characteristic	Female	Male	Total	P value <sup>a</sup>
Sex <sup>b</sup> , n (%)	4984 (47.4)	5519 (52.5)	10,504 (100)	N/A <sup>c</sup>
<b>Age (years)</b>				<.001
Mean (SD)	57.4 (20.0)	59.0 (19.5)	58.2 (19.7)	
Median (minimum-maximum)	58.0 (0.0-100.0)	60.0 (0.0-102.0)	59.0 (0.0-102.0)	
Q1-Q3	44.0-73.0	46.0-74.0	45.0-73.0	
<b>Signs and symptoms, n (%)</b>				
Cough	2482 (49.8)	2760 (50.0)	5243 (49.9)	.85
Fever	2120 (42.5)	2783 (50.4)	4904 (46.7)	<.001
Dyspnea	1476 (29.6)	1818 (32.9)	3294 (31.4)	<.001
Respiratory crackles	849 (17.0)	1085 (19.7)	1934 (18.4)	<.001
Diarrhea	556 (11.2)	543 (9.8)	1099 (10.5)	.03
Myalgia	467 (9.4)	451 (8.2)	919 (8.7)	.03
Headache	462 (9.3)	302 (5.5)	764 (7.3)	<.001
Rhonchus	279 (5.6)	414 (7.5)	693 (6.6)	<.001
Chest pain	287 (5.8)	267 (4.8)	554 (5.3)	.04
Lymphopenia	196 (3.9)	346 (6.3)	542 (5.2)	<.001
Wheezing	194 (3.9)	195 (3.5)	389 (3.7)	.36
Tachypnea	135 (2.7)	203 (3.7)	338 (3.2)	.006
Anosmia	166 (3.3)	134 (2.4)	300 (2.9)	.007
Sore throat	69 (1.4)	57 (1.0)	127 (1.2)	.12
Ageusia	33 (0.7)	32 (0.6)	65 (0.6)	.68
Dysphagia	19 (0.4)	28 (0.5)	47 (0.4)	.41
Neuralgia	19 (0.4)	22 (0.4)	41 (0.4)	>.99
Splenomegaly	8 (0.2)	14 (0.3)	22 (0.2)	.41
Hepatomegaly	2 (0.0)	6 (0.1)	8 (0.1)	.36
<b>Comorbidities<sup>d</sup>, n (%)</b>				
<b>Cardiovascular disease</b>	2253 (45.2)	2805 (50.8)	5058 (48.2)	<.001
Hypertension	1552 (31.1)	1975 (35.8)	3527 (33.6)	<.001
Ischemic stroke	91 (1.8)	163 (3.0)	254 (2.4)	<.001
<b>Heart disease</b>	1100 (22.1)	1539 (27.9)	2639 (25.1)	<.001
Ischemic heart disease	152 (3.0)	475 (8.6)	627 (6.0)	<.001
Heart failure	243 (4.9)	309 (5.6)	552 (5.3)	.11
Diabetes	689 (13.8)	957 (17.3)	1646 (15.7)	<.001
Obesity	479 (9.6)	457 (8.3)	936 (8.9)	.02
<b>Renal dysfunction</b>	271 (5.4)	493 (8.9)	764 (7.3)	<.001
CKD <sup>e</sup>	171 (3.4)	323 (5.9)	494 (4.7)	<.001
Depression	484 (9.7)	219 (4.0)	703 (6.7)	<.001
<b>Chronic respiratory disease</b>	242 (4.9)	646 (11.7)	888 (8.5)	<.001
Asthma	496 (10.0)	263 (4.8)	759 (7.2)	<.001
COPD <sup>f</sup>	126 (2.5)	549 (9.9)	675 (6.4)	<.001
OSA <sup>g</sup>	69 (1.4)	143 (2.6)	212 (2.0)	<.001

Characteristic	Female	Male	Total	<i>P</i> value <sup>a</sup>
Bronchiectasis	42 (0.8)	87 (1.6)	129 (1.2)	<.001
<b>Chronic liver disease</b>	36 (0.7)	75 (1.4)	111 (1.1)	.002
Cirrhosis	16 (0.3)	35 (0.6)	51 (0.5)	.03
HIV	12 (0.2)	22 (0.4)	34 (0.3)	.21

<sup>a</sup>*P* values from Yates-corrected chi-square test on percentage difference of female vs male COVID-19 patients. All tests were performed individually for each variable (sign, symptom, or comorbidity, where applicable). For numerical values (ie, age), *t* tests of difference between means were used.

<sup>b</sup>The sex of one patient was listed as Unknown.

<sup>c</sup>N/A: not applicable.

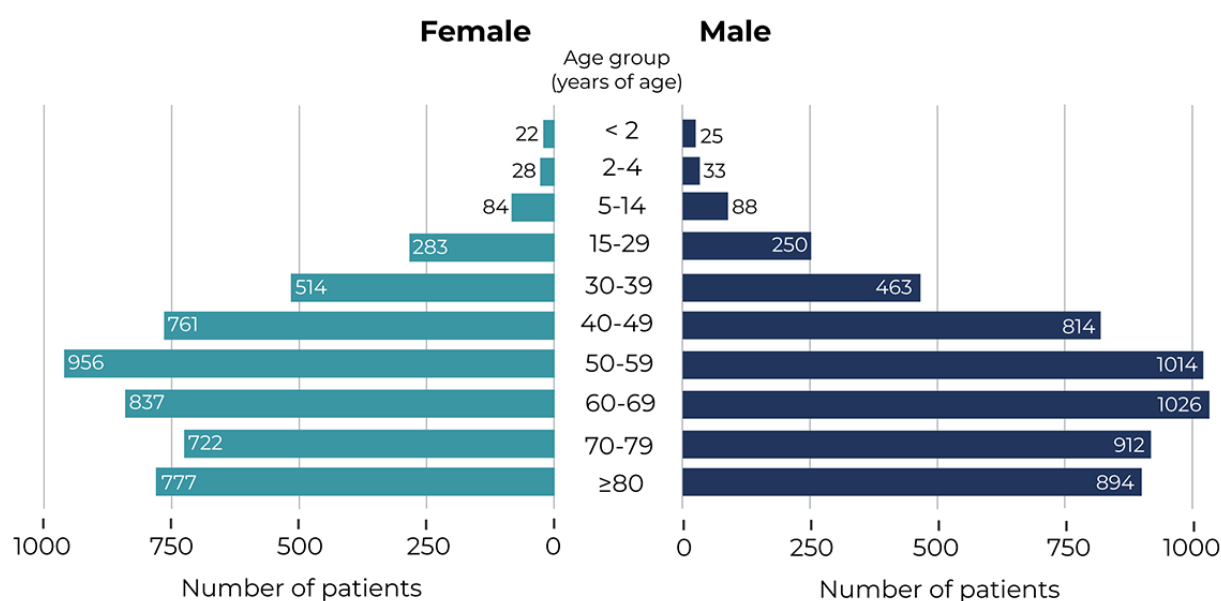
<sup>d</sup>List of medical conditions according to Systematized Nomenclature of Medicine Clinical Terms terminology.

<sup>e</sup>CKD: chronic kidney disease.

<sup>f</sup>COPD: chronic obstructive pulmonary disease.

<sup>g</sup>OSA: obstructive sleep apnea.

**Figure 3.** Age distribution of incident cases of COVID-19 in females (left) and males (right) in the study population for the period comprised between Jan 1, 2020 and March 29, 2020.



Similarly, the most frequent comorbidities among the 10,504 patients were cardiovascular disease ( $n=5058$ , 48.2%), mainly arterial hypertension ( $n=3527$ , 33.6%); heart disease ( $n=2639$ , 25.1%); and diabetes ( $n=1646$ , 15.7%) (Table 1). Regarding respiratory diseases, COPD was present in 6.4% (675), asthma in 7.2% (759), obstructive sleep apnea (OSA) in 2% (212), and bronchiectasis in 1.2% (129) of the 10,504 patients. Sex-dependent differences in comorbidities upon diagnosis are also shown in Table 1; except for asthma, the frequency of all comorbidities was significantly higher in male than in female patients with COVID-19 (all  $P<.05$ ).

Next, we explored whether the distribution of comorbidities and signs and symptoms captured in the patients' EHRs upon diagnosis differed between patients with COVID-19 who were and were not admitted to the ICU (Table 2). Regarding comorbidities, diabetes, obesity, cardiovascular disease (mainly hypertension), heart disease (mainly ischemic heart disease), and renal dysfunction were more common among patients who were admitted to the ICU (all  $P<.01$ ). As for signs and symptoms, cough, fever, dyspnea, respiratory crackles, diarrhea, tachypnea, lymphopenia, and rhonchus were more frequent among ICU patients (all  $P<.05$ ). Interestingly, respiratory diseases were not more frequent among patients who were admitted to the ICU (Table 2).

**Table 2.** Associations of signs and symptoms and comorbidities with ICU admission upon diagnosis in patients with COVID-19 (N=10,504).

Variable	Not admitted to ICU <sup>a</sup> (n=10,421), n (%)	Admitted to ICU (n=83), n (%)	P value <sup>b</sup>
<b>Signs and symptoms</b>			
Cough	5181 (49.7)	62 (74.7)	<.001
Fever	4849 (46.5)	55 (66.3)	<.001
Dyspnea	3246 (31.1)	48 (57.8)	<.001
Respiratory crackles	1904 (18.3)	30 (36.1)	<.001
Myalgia	908 (8.7)	11 (13.3)	.21
Diarrhea	1084 (10.4)	15 (18.1)	.04
Dysphagia	47 (0.5)	0 (0)	>.99
Wheezing	383 (3.7)	6 (7.2)	.16
Tachypnea	311 (3)	27 (32.5)	<.001
Chest pain	546 (5.2)	8 (9.6)	.12
Lymphopenia	524 (5)	18 (21.7)	<.001
Headache	757 (7.3)	7 (8.4)	.84
Rhonchus	676 (6.5)	17 (20.5)	<.001
Hepatomegaly	8 (0.1)	0 (0)	>.99
Anosmia	297 (2.9)	3 (3.6)	.93
Ageusia	65 (0.6)	0 (0)	.98
Neuralgia	41 (0.4)	0 (0)	1
Sore throat	126 (1.2)	1 (1.2)	1
Splenomegaly	21 (0.2)	1 (1.2)	.43
<b>Comorbidities<sup>c</sup></b>			
Diabetes	1613 (15.5)	33 (39.8)	<.001
Obesity	917 (8.8)	19 (22.9)	<.001
<b>Chronic respiratory disease</b>	883 (8.5)	5 (6)	.55
COPD <sup>d</sup>	673 (6.5)	2 (2.4)	.20
Asthma	750 (7.2)	9 (10.8)	.29
OSA <sup>e</sup>	211 (2)	1 (1.2)	.89
Bronchiectasis	129 (1.2)	0 (0)	.60
<b>Cardiovascular disease</b>	4998 (48)	60 (72.3)	<.001
Hypertension	3487 (33.5)	40 (48.2)	.007
Ischemic stroke	253 (2.4)	1 (1.2)	.72
<b>Heart disease</b>	2604 (25)	35 (42.2)	<.001
Ischemic heart disease	616 (5.9)	11 (13.3)	.01
Heart failure	548 (5.3)	4 (4.8)	>.99
<b>Renal dysfunction</b>	748 (7.2)	16 (19.3)	<.001
CKD <sup>f</sup>	488 (4.7)	6 (7.2)	.41
<b>Chronic liver disease</b>	109 (1)	2 (2.4)	.50
Cirrhosis	51 (0.5)	0 (0)	>.99
Depression	699 (6.7)	4 (4.8)	.64
HIV	33 (0.3)	1 (1.2)	.65

<sup>a</sup>ICU: intensive care unit.

<sup>b</sup>P values from Yates-corrected chi-square tests of differences between percentages of patients in either outcome group. All tests were performed individually for each variable (sign, symptom, or comorbidity, where applicable).

<sup>c</sup>List of medical conditions according to Systematized Nomenclature of Medicine Clinical Terms terminology.

<sup>d</sup>COPD: chronic obstructive pulmonary disease.

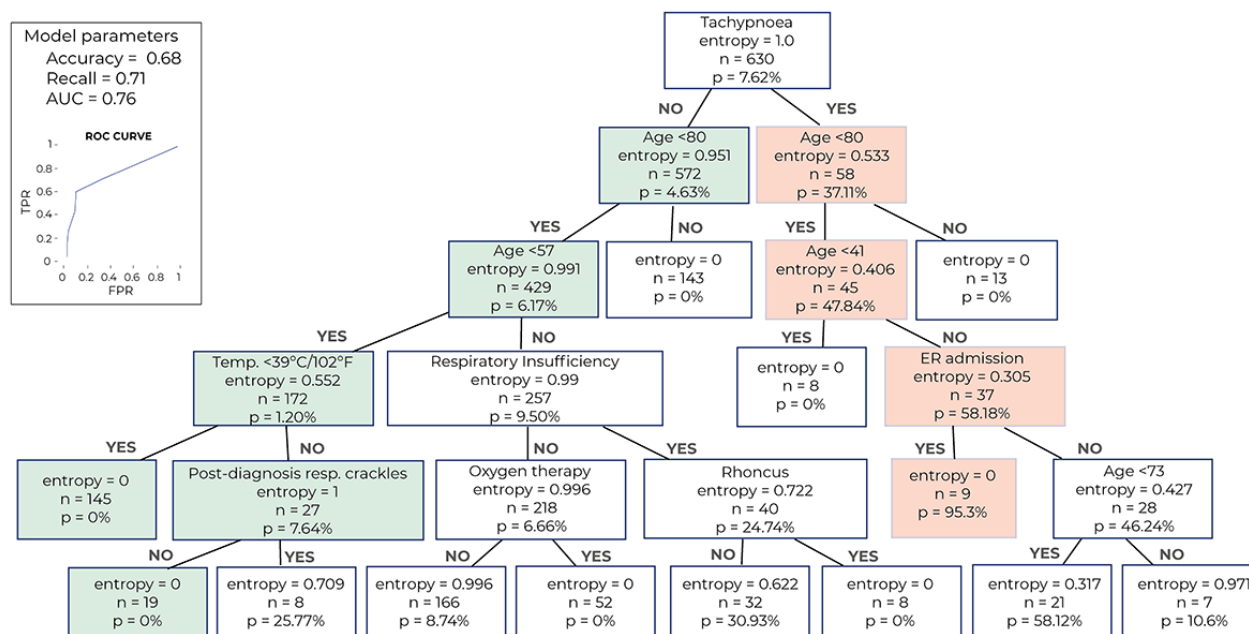
<sup>e</sup>OSA: obstructive sleep apnea.

<sup>f</sup>CKD: chronic kidney disease.

Finally, by using a machine-learning, data-driven algorithm, we identified that a combination of three easily available clinical variables, namely age, temperature, and respiratory frequency, was the most parsimonious predictor of ICU admission among patients with COVID-19 (Figure 4). For this model, age and temperature were captured as continuous variables, whereas tachypnea (yes/no) was defined as a respiratory frequency of more than 20 breaths per minute. With accuracy, recall, and area under the curve (AUC) values of 0.68, 0.71, and 0.76, respectively, the presented model reached optimal balance in terms of positive and negative predictive values for ICU admission. On the one hand, patients younger than 56 years, without tachypnea, and with temperature <39 °C (entropy=0, n=145) or >39 °C without respiratory crackles (entropy=0, n=18) were not admitted to the ICU. On the other hand, patients with COVID-19 aged 40 to 70 years were likely to be admitted in the ICU if they presented with tachypnea and delayed their visit

to the ED after being seen in primary care (entropy=0, n=104). As stated in the Methods section, we performed an additional sensitivity analysis with different data sets to further validate the results of our predictive model. The independent data set of two provinces (Ciudad Real and Guadalajara, including a total of 753,408 individual patients, or 38% of the entire study sample from Castilla-La Mancha; Figure 1 and Supplemental Table S1 in Multimedia Appendix 1), was used to retrain our algorithm to identify ICU admission at onset; validation was performed in the remaining three provinces. As shown in Supplemental Figure S2 in Multimedia Appendix 1, the new decision tree identified the same relevant clinical variables, that is age, tachypnea, temperature, and respiratory crackles/rhonchus, with similar (but not identical) thresholds in some variables. This additional model achieved accuracy, recall, and AUC values of 0.85, 0.57, and 0.84, respectively, providing additional proof of validity for our main findings.

**Figure 4.** Decision tree of relevant clinical variables for the prediction of ICU admission in patients with COVID-19. The combination of three easily available clinical variables, namely age, temperature, and respiratory frequency, was the most parsimonious predictor of ICU admission among COVID-19 patients. The number of patients, probability (p) of ICU admission predicted by the model, and level of entropy (a measure indicating how mixed or pure the classification is, where 0 indicates optimal separation of classes) are indicated in each box. The green pathway indicates that patients with no tachypnea, age <56 years, and temperature <39 °C (OR >39 °C without respiratory crackles) did not require ICU admission. In contrast, the red pathway indicates that patients aged 40-79 years, who presented with tachypnea, and who delayed their visit to the emergency department after being seen in primary care were likely to be admitted in the ICU. For this model, we obtained accuracy, recall, and AUC values of 0.68, 0.71, and 0.76, respectively (top right panel). AUC: area under the curve; FPR: false positive rate; resp.: respiratory; ROC: receiver operating characteristic; TPR: true positive rate.



## Discussion

### Principal Findings

By accessing the clinical information of more than 10,000 anonymous patients with COVID-19 (a number that largely

surpasses samples included in recent reports about the disease [30,31]), we were able to describe their demographic and clinical characteristics, their clinical journey, and the statistical relationship between the most common symptoms and comorbidities on admission, and COVID-19 prognosis (ie, ICU admission). There were subtle differences in clinical symptoms

at onset by sex, while all comorbidities (except asthma) were significantly more frequent in male than female patients with COVID-19. The variables identified in our ICU admission model (ie, age, temperature, and tachypnea) are clinically relevant, as they are readily available and easily observable in everyday practice for patients with COVID-19. Although tachypnea is not an exclusive manifestation of COVID-19 and can be present in patients suffering from other respiratory diseases (ie, pneumonia), our model suggests that this sign (in combination with age and temperature) is a more reliable predictor of ICU use than other common symptoms and signs, such as cough, dyspnea, or respiratory crackles.

The reported sex-dependent differences in the symptomatology of COVID-19 at onset have been further confirmed by our group using similar methods [32] and should be interpreted in light of data suggesting that female teenagers and young adult women are significantly more affected by the disease than their male counterparts [32]. In this regard, our results warrant further investigations aimed at closing the gender gap in the ongoing pandemic [33].

Given that the stability and capacity of ICUs worldwide is threatened by the rapid spread of COVID-19, the identification of individual factors that predict ICU admission may not only improve patient management but also optimize health care resource use and planning. Thus, recent studies using big data and machine learning have explored the prognostic factors of the disease, including ICU transfer, discharge, and mortality [8-11]. In line with our results, respiratory rate has also been identified as an important predictor of ICU transfer in patients with COVID-19 [9].

If further applied to other national and international health care networks, the tools and methodology presented in this study can potentially characterize and predict the prognosis of COVID-19 in a timely and unprecedented manner. As demonstrated in recent studies [34,35], there may be value in the application of AI to the current COVID-19 pandemic, not only to predict outbreaks [36] or read chest computed tomography scans [37] but also to elucidate the clinical onset and natural history of COVID-19 almost in real time. Indeed, classical methods would require months of questionnaire-based data collection and questionnaire validation along with multiple Ethics Board approvals and other practical hurdles; these steps can be avoided with our current approach.

In the race against COVID-19 [38], where the goal is to curb the pandemic, it is imperative to leverage big data and intelligent analytics for the betterment of public health. However, it is of the utmost importance not to neglect privacy and public trust, to apply best practices, and to maintain responsible standards for data collection and data processing on a global scale [39].

## Strengths and Limitations

To our knowledge, this is one of the first attempts to combine NLP and machine learning to access and analyze unstructured, free-text real-world data from EHRs in a large series of patients with COVID-19. Although recent studies have used machine learning to predict ICU admission in patients with COVID-19 [9], our approach takes this methodology one step further by

applying NLP to exclusively analyze unstructured information. Indeed, our state-of-the-art methodology enabled rapid analysis of the unstructured free-text narratives in the EHRs of 1 million patients from the general population of the region of Castilla-La Mancha (Spain).

Our methodology combined modules for sentence segmentation, tokenization, text normalization, acronym disambiguation, negation detection, and a multidimensional ranking scheme; the latter involved linguistic knowledge, statistical evidence, and continuous vector representations of words and documents learned via shallow neural networks. When applied to EHRs, NLP enables both access to longitudinal health records for *all* patients in the target population and the implementation of exploratory analysis to clarify associations between variables that have remained undetected with traditional research methods. By considering all possible patients with the target disease, the information and analyses used here (ie, real-world data and free-scale statistics) remained unbiased by the research question or the observers. Unlike classical statistical methods (eg, logistic regression), the main advantage associated with the use of machine learning in this context is that it enables the automatic detection of meaningful relationships between variables. For instance, if a given symptom (ie, fever) is only relevant for certain patients (ie, older than 50 years), techniques such as the classification trees used here are suitable to uncover this relationship. In this context, although the total number of patients who required ICU use in the training set was somewhat low (48 patients), the number of variables considered in the model was also very limited. In addition, the inclusion of a validation stage reduces the likelihood of overfitting. Ultimately, the use of classification trees in this study (as opposed to other models, such as artificial neural networks) is especially appropriate in the clinical context because they are easily interpretable.

Regarding the geographical location of our participating hospital sites, it is worth mentioning that with a total of 1,364,924 patients from the region of Castilla-La Mancha (SESCAM Health Care Network), our sample is representative of the Spanish population. Spain is among the countries that have been most affected by the pandemic in terms of both total cases and mortality rates [40,41]; the Castilla-La Mancha region in particular is the third most affected region in the country, just behind Madrid and Catalonia. For this reason, we anticipate that the clinical conclusions drawn here will be relevant for clinicians worldwide. Of note, the ICU capacity in the region during the study period had not yet been compromised, which protects against possible bias in our training data (all patients requiring intensive care were indeed admitted to the ICU).

The results and conclusions of the present study should be interpreted in light of the following limitations. First, we did not distinguish COVID-19 cases confirmed by laboratory results (ie, RT-PCR) from those exclusively diagnosed through clinical observation (ie, symptomatology, imaging and laboratory results). However, it should be noted that PCR and other rapid laboratory tests for the detection of SARS-CoV-2 were not routinely administered in Spain during the study period. In addition, this decision is supported by recent discussions on the clinical validity and relatively high sensitivity of symptom- and



imaging-based identification of patients with COVID-19, especially in early stages of the disease [20,22,23]. Second, independent replications by different research groups in larger patient sets are needed to further support the clinical validity of our results.

Finally, future reports from the BigCOVIData study may incorporate laboratory results and treatments and may contextualize the results presented here in a larger clinical picture [28].

## Conclusion

In this study, we found that in the largest series of patients with COVID-19 attended during the first three months of the pandemic in Spain, 6.1% of all hospitalized patients (83/1353) required ICU admission. We also found that a combination of easily obtained clinical variables, namely age, fever, and tachypnea, predicts whether patients with COVID-19 will require ICU admission.

## Acknowledgments

The members of the Savana COVID-19 Research Group are Ignacio H Medrano, MD; Jorge Tello; Alberto Porras, MD, PhD; Marisa Serrano, PhD; Sara Lumbreras, PhD; Universidad Pontificia Comillas; Carlos Del Rio-Bermudez, PhD; Stephanie Marchesseau, PhD; Ignacio Salcedo; Andrea Martínez; Claudia Maté, MD; Sergio Collazo, MD; Jesús Barea, MD; María Villamayor, MD; Antonio Urda, MD, PhD; Carolina de la Pinta, MD; Imanol Zubizarreta; Yolanda González, PhD; and Sebastian Menke, PhD. We thank all the Savaners for helping accelerate health science with their daily work. This study would have not been possible without every single team member. This study was sponsored by Savana [42]. We also thank the SESCAM Health Care Network in Castilla-La Mancha, Spain, for its participation in the study and for supporting the development of cutting-edge technology in real time.

## Conflicts of Interest

JLI has received consulting or speaking fees from AstraZeneca, Bayer, Boehringer Ingelheim, Chiesi, Glaxo, Grifols, Smith Kline, Menarini, Novartis, Orion, Pfizer, Sandoz, and Teva.

## Multimedia Appendix 1

Supplementary Methods/Supplementary Figures and Tables.

[DOC File, 366 KB - [jmir\\_v22i10e21801\\_app1.doc](#)]

## References

1. Coronavirus (COVID-19). US Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/coronavirus/2019-ncov/index.html> [accessed 2020-04-08]
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020 Feb 20;382(8):727-733. [doi: [10.1056/nejmoa2001017](#)]
3. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: [10.1126/science.abb6936](#)] [Medline: [32234805](#)]
4. Qin L, Sun Q, Wang Y, Wu K, Chen M, Shia B, et al. Prediction of Number of Cases of 2019 Novel Coronavirus (COVID-19) Using Social Media Search Index. *Int J Environ Res Public Health* 2020 Mar 31;17(7):2365 [FREE Full text] [doi: [10.3390/ijerph17072365](#)] [Medline: [32244425](#)]
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019 Jan;25(1):44-56. [doi: [10.1038/s41591-018-0300-7](#)] [Medline: [30617339](#)]
6. Divita G, Carter M, Redd A, Zeng Q, Gupta K, Trautner B, et al. Scaling-up NLP Pipelines to Process Large Corpora of Clinical Notes. *Methods Inf Med* 2018 Jan 23;54(06):548-552. [doi: [10.3414/me14-02-0018](#)]
7. Burgner D, Jamieson SE, Blackwell JM. Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *Lancet Infect Dis* 2006 Oct;6(10):653-663. [doi: [10.1016/s1473-3099\(06\)70601-6](#)]
8. Liu Y, Mao B, Liang S, Yang J, Lu H, Chai Y, et al. Association Between Age and Clinical Characteristics and Outcomes of Coronavirus Disease 2019. SSRN Journal Preprint posted online on March 31, 2020. [doi: [10.2139/ssrn.3556689](#)]
9. Cheng F, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients. *J Clin Med* 2020 Jun 01;9(6):1668 [FREE Full text] [doi: [10.3390/jcm9061668](#)] [Medline: [32492874](#)]
10. Nemati M, Ansary J, Nemati N. Machine-Learning Approaches in COVID-19 Survival Analysis and Discharge-Time Likelihood Prediction Using Clinical Data. *Patterns (NY)* 2020 Aug 14;1(5):100074 [FREE Full text] [doi: [10.1016/j.patter.2020.100074](#)] [Medline: [32835314](#)]
11. Darabi HR, Tsinis D, Zecchini K, Whitcomb WF, Liss A. Forecasting Mortality Risk for Patients Admitted to Intensive Care Units Using Machine Learning. *Procedia Computer Science* 2018;140:306-313. [doi: [10.1016/j.procs.2018.10.313](#)]

12. Horton R. Offline: COVID-19—what countries must do now. *Lancet* 2020 Apr;395(10230):1100. [doi: [10.1016/s0140-6736\(20\)30787-x](https://doi.org/10.1016/s0140-6736(20)30787-x)]
13. Xu B, Gutierrez B, Mekaru S, Sewalk K, Goodwin L, Loskill A, et al. Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci Data* 2020 Mar 24;7(1):106 [FREE Full text] [doi: [10.1038/s41597-020-0448-0](https://doi.org/10.1038/s41597-020-0448-0)] [Medline: [32210236](https://pubmed.ncbi.nlm.nih.gov/32210236/)]
14. IHME COVID-19 health service utilization forecasting team, Murray CJL. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *medRxiv Preprint* posted online on March 30, 2020. [doi: [10.1101/2020.03.27.20043752](https://doi.org/10.1101/2020.03.27.20043752)]
15. Sotgiu G, Gerli AG, Centanni S, Miozzo M, Canonica GW, Soriano JB, et al. Advanced forecasting of SARS-CoV-2-related deaths in Italy, Germany, Spain, and New York State. *Allergy* 2020 Jul;75(7):1813-1815 [FREE Full text] [doi: [10.1111/all.14327](https://doi.org/10.1111/all.14327)] [Medline: [32306406](https://pubmed.ncbi.nlm.nih.gov/32306406/)]
16. Izquierdo J, Morena D, González Y, Paredero J, Pérez B, Graziani D, et al. Clinical Management of COPD in a Real-World Setting. A Big Data Analysis. *Arch Bronconeumol* 2020 Feb 22;in press. [doi: [10.1016/j.arbres.2019.12.025](https://doi.org/10.1016/j.arbres.2019.12.025)] [Medline: [32098727](https://pubmed.ncbi.nlm.nih.gov/32098727/)]
17. Sociedad EDNYCT. Chart Review of Patients With COPD, Using Electronic Medical Records and Artificial Intelligence (BigCOPData). *ClinicalTrials.gov*. 2019 Dec 20. URL: <https://clinicaltrials.gov/ct2/show/NCT04206098> [accessed 2020-04-13]
18. Koo D, Thacker SB. In snow's footsteps: Commentary on shoe-leather and applied epidemiology. *Am J Epidemiol* 2010 Sep 15;172(6):737-739. [doi: [10.1093/aje/kwq252](https://doi.org/10.1093/aje/kwq252)] [Medline: [20720100](https://pubmed.ncbi.nlm.nih.gov/20720100/)]
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Bull World Health Organ* 2007 Nov;85(11):867-872 [FREE Full text] [doi: [10.2471/blt.07.045120](https://doi.org/10.2471/blt.07.045120)] [Medline: [18038077](https://pubmed.ncbi.nlm.nih.gov/18038077/)]
20. Long C, Xu H, Shen Q, Zhang X, Fan B, Wang C, et al. Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? *Eur J Radiol* 2020 May;126:108961 [FREE Full text] [doi: [10.1016/j.ejrad.2020.108961](https://doi.org/10.1016/j.ejrad.2020.108961)] [Medline: [32229322](https://pubmed.ncbi.nlm.nih.gov/32229322/)]
21. Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* 2020 May 12;323(18):1843-1844 [FREE Full text] [doi: [10.1001/jama.2020.3786](https://doi.org/10.1001/jama.2020.3786)] [Medline: [32159775](https://pubmed.ncbi.nlm.nih.gov/32159775/)]
22. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* 2020 Aug;296(2):E32-E40 [FREE Full text] [doi: [10.1148/radiol.2020200642](https://doi.org/10.1148/radiol.2020200642)] [Medline: [32101510](https://pubmed.ncbi.nlm.nih.gov/32101510/)]
23. Xu J, Wu R, Huang H, Zheng W, Ren X, Wu N, et al. Computed Tomographic Imaging of 3 Patients With Coronavirus Disease 2019 Pneumonia With Negative Virus Real-time Reverse-Transcription Polymerase Chain Reaction Test. *Clin Infect Dis* 2020 Jul 28;71(15):850-852 [FREE Full text] [doi: [10.1093/cid/ciaa207](https://doi.org/10.1093/cid/ciaa207)] [Medline: [32232429](https://pubmed.ncbi.nlm.nih.gov/32232429/)]
24. Hernandez Medrano I, Tello Guijarro J, Belda C, Urena A, Salcedo I, Espinosa-Anke L, et al. Savana: Re-using Electronic Health Records with Artificial Intelligence. *IJIMAI* 2018;4(7):8. [doi: [10.9781/ijimai.2017.03.001](https://doi.org/10.9781/ijimai.2017.03.001)]
25. Yang Z, Dehmer M, Yli-Harja O, Emmert-Streib F. Combining deep learning with token selection for patient phenotyping from electronic health records. *Sci Rep* 2020 Jan 29;10(1):1432 [FREE Full text] [doi: [10.1038/s41598-020-58178-1](https://doi.org/10.1038/s41598-020-58178-1)] [Medline: [31996705](https://pubmed.ncbi.nlm.nih.gov/31996705/)]
26. The Lancet. The gendered dimensions of COVID-19. *Lancet* 2020 Apr;395(10231):1168. [doi: [10.1016/s0140-6736\(20\)30823-0](https://doi.org/10.1016/s0140-6736(20)30823-0)]
27. Quinlan JR. Induction of decision trees. *Mach Learn* 1986 Mar;1(1):81-106. [doi: [10.1007/bf00116251](https://doi.org/10.1007/bf00116251)]
28. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020 Apr 07;369:m1328. [doi: [10.1136/bmj.m1328](https://doi.org/10.1136/bmj.m1328)] [Medline: [32265220](https://pubmed.ncbi.nlm.nih.gov/32265220/)]
29. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015 Jan 06;162(1):W1. [doi: [10.7326/m14-0698](https://doi.org/10.7326/m14-0698)]
30. Lescure F, Bouadma L, Nguyen D, Parisey M, Wicky P, Behillil S, et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. *Lancet Infect Dis* 2020 Jun;20(6):697-706. [doi: [10.1016/s1473-3099\(20\)30200-0](https://doi.org/10.1016/s1473-3099(20)30200-0)]
31. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 2020 Mar 26;382(13):1199-1207. [doi: [10.1056/nejmoa2001316](https://doi.org/10.1056/nejmoa2001316)]
32. Ancochea J, Izquierdo J, Savana COVID-19 Research Group, Soriano J. Evidence of gender bias in the diagnosis and management of COVID-19 patients: A Big Data analysis of Electronic Health Records. *medRxiv Preprint* posted online on July 26, 2020. [doi: [10.1101/2020.07.20.20157735](https://doi.org/10.1101/2020.07.20.20157735)]
33. Wenham C, Smith J, Morgan R. COVID-19: the gendered impacts of the outbreak. *Lancet* 2020 Mar;395(10227):846-848. [doi: [10.1016/s0140-6736\(20\)30526-2](https://doi.org/10.1016/s0140-6736(20)30526-2)]
34. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. *Lancet Digit Health* 2020 Apr;2(4):e166-e167. [doi: [10.1016/s2589-7500\(20\)30054-6](https://doi.org/10.1016/s2589-7500(20)30054-6)]

35. Du RH, Liang LR, Yang CQ, Wang W, Cao TZ, Li M, et al. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study. *Eur Respir J* 2020 May;55(5):2000524. [doi: [10.1183/13993003.00524-2020](https://doi.org/10.1183/13993003.00524-2020)] [Medline: [32269088](https://pubmed.ncbi.nlm.nih.gov/32269088/)]
36. Ayyoubzadeh S, Ayyoubzadeh S, Zahedi H, Ahmadi M, Kalhori SRN. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. *JMIR Public Health Surveill* 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: [10.2196/18828](https://doi.org/10.2196/18828)] [Medline: [32234709](https://pubmed.ncbi.nlm.nih.gov/32234709/)]
37. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* 2020 Aug;296(2):E65-E71 [FREE Full text] [doi: [10.1148/radiol.202000905](https://doi.org/10.1148/radiol.202000905)] [Medline: [32191588](https://pubmed.ncbi.nlm.nih.gov/32191588/)]
38. Editorial. The race against COVID-19. *Nat Nanotechnol* 2020 Apr 17;15(4):239-240. [doi: [10.1038/s41565-020-0680-y](https://doi.org/10.1038/s41565-020-0680-y)] [Medline: [32303704](https://pubmed.ncbi.nlm.nih.gov/32303704/)]
39. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat Med* 2020 Apr 27;26(4):463-464 [FREE Full text] [doi: [10.1038/s41591-020-0832-5](https://doi.org/10.1038/s41591-020-0832-5)] [Medline: [32284619](https://pubmed.ncbi.nlm.nih.gov/32284619/)]
40. Coronavirus disease 2019 (COVID-19) Situation Report 64. World Health Organization. 2020 Mar 24. URL: [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200324-sitrep-64-covid-19.pdf?sfvrsn=723b221e\\_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200324-sitrep-64-covid-19.pdf?sfvrsn=723b221e_2) [accessed 2020-03-24]
41. Situación de COVID-19 en España. Ministerio de Sanidad 2020. URL: <https://covid19.isciii.es> [accessed 2020-04-13]
42. Savana. URL: <https://www.savanaamed.com/> [accessed 2020-10-23]

## Abbreviations

**AI:** artificial intelligence

**AUC:** area under the curve

**COPD:** chronic obstructive pulmonary disease

**ED:** emergency department

**EHR:** electronic health record

**ICH:** International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use

**ICU:** intensive care unit

**NLP:** natural language processing

**OSA:** obstructive sleep apnea

**RT-PCR:** reverse transcriptase–polymerase chain reaction

**SESCAM:** Servicio de Salud de Castilla-La Mancha

**SNOMED CT:** Systematized Nomenclature of Medicine Clinical Terms

**STROBE:** STrengthening the Reporting of OBservational studies in Epidemiology

**TRIPOD:** Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis

*Edited by G Eysenbach; submitted 29.06.20; peer-reviewed by C Fincham, F Palmieri, I Mircheva, J Li; comments to author 21.07.20; revised version received 28.07.20; accepted 20.10.20; published 28.10.20.*

*Please cite as:*

*Izquierdo JL, Ancochea J, Savana COVID-19 Research Group, Soriano JB*

*Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing*

*J Med Internet Res* 2020;22(10):e21801

URL: <http://www.jmir.org/2020/10/e21801/>

doi: [10.2196/21801](https://doi.org/10.2196/21801)

PMID: [33090964](https://pubmed.ncbi.nlm.nih.gov/33090964/)

©Jose Luis Izquierdo, Julio Ancochea, Savana COVID-19 Research Group, Joan B Soriano. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Determinants of Patients' Intention to Use the Online Inquiry Services Provided by Internet Hospitals: Empirical Evidence From China

Dehe Li<sup>1</sup>, BA; Yinhan Hu<sup>1</sup>, PhD; Holger Pfaff<sup>2</sup>, PhD; Liuming Wang<sup>3</sup>, MA; Lu Deng<sup>1</sup>, MA; Chuntao Lu<sup>4</sup>, MA; Shixiao Xia<sup>1</sup>, BA; Siyu Cheng<sup>1</sup>, BA; Ximin Zhu<sup>1</sup>, BA; Xiaoyue Wu<sup>1</sup>, BA

<sup>1</sup>School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

<sup>2</sup>Center for Health Services Research Cologne, University of Cologne, Cologne, Germany

<sup>3</sup>Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

<sup>4</sup>Jingmen No. 2 People's Hospital, Jingmen, Hubei, China

**Corresponding Author:**

Yinhan Hu, PhD

School of Medicine and Health Management

Tongji Medical College

Huazhong University of Science and Technology

No. 13, Hangkong Road

Wuhan, Hubei,

China

Phone: 86 13554285879

Fax: 86 2783692727

Email: [hyh288@hotmail.com](mailto:hyh288@hotmail.com)

## Abstract

**Background:** Internet hospitals show great potential for adequately fulfilling people's demands for high-quality outpatient services, and with the normalization of the epidemic prevention and control of COVID-19, internet hospitals play an increasingly important role in delivering health services to the public. However, the factors that influence patients' intention to use the online inquiry services provided by internet hospitals remain unclear. Understanding the patients' behavioral intention is necessary to support the development of internet hospitals in China and promote patients' intention to use online inquiry services provided by internet hospitals during the prevention and control of the COVID-19 epidemic.

**Objective:** The purpose of this study is to identify the determinants of patients' intention to use the online inquiry services provided by internet hospitals based on the theory of planned behavior (TPB).

**Methods:** The hypotheses of our research model were developed based on the TPB. A questionnaire was developed through patient interviews, verified using a presurvey, and used for data collection for this study. The cluster sampling technique was used to include respondents with chronic diseases. Structural equation modeling was used to test the research hypotheses.

**Results:** A total of 638 valid responses were received from patients with chronic diseases. The goodness-of-fit indexes corroborated that the research model was a good fit for the collected data. The model explained 45.9% of the variance in attitude toward the behavior and 60.5% of the variance in behavioral intention. Perceived behavioral control and perceived severity of disease had the strongest total effects on behavioral intention ( $\beta=.624$ ,  $P=.004$  and  $\beta=.544$ ,  $P=.003$ , respectively). Moreover, perceived convenience, perceived information risk, emotional preference, and health consciousness had indirect effects on behavioral intention, and these effects were mediated by attitude toward the behavior. Among the four constructs, perceived convenience had the highest indirect effect on behavioral intention ( $\beta=.207$ ;  $P=.001$ ).

**Conclusions:** Perceived behavioral control and perceived severity of disease are the most important determinants of patients' intention to use the online inquiry services provided by internet hospitals. Therefore, internet hospitals should further optimize the design of online service delivery and ensure a reasonable assembly of high-quality experts, which will benefit the promotion of patients' adoption intention toward online inquiry services for health purposes. Perceived convenience, emotional preference, and perceived risks also have effects on behavioral intention. Therefore, the relevant quality control standards and regulations for internet hospitals should be further developed and improved, and the measures to protect personal information should be



strengthened to ensure the patient safety. Our study supports the use of the TPB in explaining patients' intention to use online inquiry services provided by internet hospitals.

(*J Med Internet Res* 2020;22(10):e22716) doi:[10.2196/22716](https://doi.org/10.2196/22716)

## KEYWORDS

internet hospital; health care-seeking intention; online inquiry; theory of planned behavior; chronic disease; structural equation modeling; China; COVID-19; intention; online service; eHealth; behavior; modeling

## Introduction

### Background

With the development of human society, an aging population with chronic and age-related diseases, and its increasing health care requirements have become a challenge for the governments of all countries worldwide, especially China [1,2]. In China, many people complain about the “difficulty of seeing a doctor” and tend to go to high-level hospitals even for mild symptoms [3]. It is difficult for patients in rural and remote areas to access high-quality health services owing to an uneven distribution of medical resources [4,5]; moreover, the time spent by physicians on communicating with each patient is limited, usually less than 3 minutes in outpatient departments in large hospitals [6]. A latest national health services survey showed the rate of chronic diseases in Chinese residents was up to 245.2 person per thousand population, and the majority of the common chronic diseases were hypertension, diabetes, and heart disease [7]. Patients with chronic diseases such as diabetes receive little education or information about their diseases in such a short time [8]; thus, their health service requirements are rarely fulfilled. Under such circumstances, inability of traditional medical resources to meet the increasing demands of health services for the public contributes to the emergence of internet hospitals [9].

Internet hospitals, a novel approach to health service delivery via internet technology [3,9,10], were developed as a potential solution to optimize the allocation of health resources and to meet people's demands for high-quality health services [3]. Compared with mobile health (mHealth), a technology integration of mobile computing, medical sensors, and portable devices for meeting the needs of consumers about health care or health information services [11], internet hospitals refer to one-stop online service platforms based on traditional hospitals, which integrate consultation, prescription, payment, and drug delivery; internet hospitals enable patients to communicate with skilled experts anytime and anywhere using a website or a smartphone app and gain access to health-related services such as online disease counseling, online guidance, electronic prescriptions, medical information, remote diagnosis, treatment and rehabilitation, chronic disease follow-up, and health education [3,9]. Several studies have shown that compared with traditional medical services, internet hospitals increased the time that the physicians spent with each patient, improved the accessibility to high-quality health care, alleviated the difficulty of “seeing a doctor” [3,12], promoted the rational distribution of health resources, and improved the balance between supply and demand of health resources in China [4,9]. In addition, the Chinese government has implemented an ambitious national

strategy called “Health China” since 2016, and this strategy is expected to solve the problem of “difficulty and expense of seeing a doctor” and improve the ability of medical resources to meet the increasing demands of the public health services in China [13]. Thus, the development of internet hospitals is crucial for implementing this national strategy.

Internet hospitals have also shown great potential for the epidemic prevention and control of COVID-19 in China [10,14]. During the COVID-19 epidemic, internet hospitals, which provide free COVID-19 consultation services and offer essential medical support to the public, have effectively reduced the probability of nosocomial cross-infections and fulfilled patients' demands for outpatient services, especially for patients with chronic diseases [10], and thus, have been further widely accepted by society. However, a latest research report in China in 2020 has highlighted that the users' acceptance of online inquiry services provided by different types of internet hospitals was relatively poor, with significant differences [15]. Among enterprise-led internet hospitals such as WeDoctor, Hao Yisheng, and AliHealth, more than half (53.4%) provided services to less than 1000 patients per day, whereas approximately one-third (26.6%) provided services to more than 3000 patients per day. Moreover, the number of daily outpatient visits of medical institution-led internet hospitals such as the internet hospital of People's Hospital of Zhejiang Province and the internet hospital of No.2 People's Hospital of Guangdong Province was generally lower, with two-thirds (66.7%) of the internet hospitals providing services to less than 100 patients per day.

With the normalization of prevention and control of the COVID-19 epidemic, internet hospitals play an increasingly important role in delivering health services to the public. Moreover, issues such as whether patients accept internet hospitals and what factors influence patients' decisions regarding the online inquiry services provided by internet hospitals should be considered for the long-term development of internet hospitals. Understanding the factors that affect patients' intention to use online inquiry services will help internet hospitals to further improve their service design, promote their development, and thus promote patients' health care-seeking intention to use online inquiry services during the prevention and control of the COVID-19 epidemic. However, the factors influencing patients' health care-seeking intention to use online inquiry services provided by internet hospitals remain unclear.

Previous studies on patients' health care-seeking behaviors of online inquiry focus on mHealth services, and theoretical models have been applied to identify the determinants of adoption intentions for mHealth services [8,13,16-18] or health technology [19-21]; however, few studies focus on patients'



health care-seeking intention to use the online inquiry service provided by internet hospitals. A theoretical model must be identified and tested to provide a context-related understanding of patients' intention to use the online inquiry services provided by internet hospitals. Internet hospitals have unique service delivery characteristics, and patients exhibit different degrees of acceptance for online inquiry services for different types of internet hospitals [15]. Thus, it is necessary to analyze the factors influencing the intention to use the online inquiry services provided by internet hospitals. To date, relevant theoretical models have not been applied to patients' intention to use the online inquiry services in the context of internet hospitals.

### Theory of Planned Behavior

The theory of planned behavior (TPB), an expansion of the theory of reasoned action, has been widely used as a classical theory in explaining and predicting various types of human behaviors [22,23]. The TPB mainly consists of five variables: attitude toward the behavior, subjective norm (SN), perceived behavioral control (PBC), behavioral intention, and actual behavior. According to TPB, whether individuals perform a certain behavior depends on their behavioral intention to perform the behavior. Behavioral intention is determined by the following factors: attitude toward the behavior, SN, and PBC, which is also called adoption intention [23]. Compared with the technology acceptance model, which has been widely applied to the acceptance and use of information technology [13,16], TPB considers both the subjective and social factors affecting patients' behavioral intention as well as external control conditions that are not affected by personal willingness [22,23]. In health care settings, patients' intention to perform a behavior for health purposes is often the result of an interaction between internal and external factors. The TPB has been frequently applied in the field of mHealth services [24-27], and these studies found that attitude toward the behavior, SN, and PBC are the major determinants of behavioral intention toward mHealth service adoption. However, to date, TPB has not been used in the study of the factors influencing patients' behavioral intention to use online inquiry services in the context of internet hospitals. Considering the uniqueness of mHealth service, Zhang et al [27] proposed the integration of TPB and protection motivation theory in the context of mHealth services and found an indirect effect of perceived severity of disease (PSD) on behavioral intention. Thus, patients' PSD should be considered. To date, China has not established an effective quality control system for internet hospitals [12,15]. Potential medical risks of the online inquiry services provided by internet hospitals threaten patients' health safety. Therefore, medical risks should also be considered.

Thus, in this study, we combined the TPB and patients' PSD and perceived medical risks (PMRs) to develop a theoretical model to identify the factors influencing patients' behavioral intention to use the online inquiry services provided by internet hospitals.

### Research Hypotheses and Model

In this study, the TPB model was adopted as the theoretical framework, where attitude toward the behavior refers to the

extent to which a patient has a favorable or unfavorable opinion of online inquiry behavior; SN is defined as a patient's perceived social pressure when performing online inquiry; and PBC is defined as a patient's perception of the degree of ease or difficulty when using the online inquiry services provided by internet hospitals, and it is generally equivalent to self-efficacy [23,27]. Several studies have shown that attitude toward the behavior, SN, and PBC are the positive determinants of patients' adoption intention toward mHealth services [24,26,27]. Therefore, three hypotheses were formulated as follows:

- Hypothesis (H)1: Attitude toward the behavior has a positive effect on the behavioral intention of patients to use the online inquiry services provided by internet hospitals.
- H2: PBC has a positive effect on the behavioral intention of patients to use the online inquiry services provided by internet hospitals.
- H3: SN has a positive effect on the behavioral intention of patients to use the online inquiry services provided by internet hospitals.

A previous study on the use of mobile technology demonstrated that perceived convenience (PC) positively affected attitude toward the behavior [28]. Internet hospitals not only save the time and effort that patients spend on seeking health services but also conveniently provide patients with health services via the internet anywhere and anytime. In this study, we define PC as patients' perceptions of saving time and effort when using the online inquiry services provided by internet hospitals. If patients find it convenient to access health service while having to spend less time and effort, they will have a positive attitude toward participating in telemedicine [29]. Thus, we posed the following hypothesis:

- H4: PC has a positive effect on the attitude of patients toward using online inquiry in internet hospitals.

A study concerning mobile commerce demonstrated the effect of perceived outcome (PO) on behavioral intention, and this effect was mediated by behavioral attitude [30]. In health care settings, the value of PO affects patients' choices to a certain extent, and patients in China tend to go to high-level hospitals rather than primary health care institutions even for mild symptoms [3]. If patients find that the service outcomes obtained from internet hospitals are consistent with or even better than those from traditional large hospitals, they may have a positive attitude toward online inquiry. In this study, PO refers to patients' perceptions of the degree to which online inquiry achieves the expected health outcome improvement. Thus, we posed the following hypothesis:

- H5: PO has a positive effect on the attitude of patients toward online inquiry in internet hospitals.

Although mHealth services may promote the accessibility of high-quality health services, alleviate the difficulty in "seeing a doctor," and thus improve patients' quality of life, they also lead to issues regarding quality control, information security, and privacy protection [31,32]. Likewise, internet hospitals were potentially puzzled by these issues. The potential risks of using the online inquiry services provided by internet hospitals include medical risk and information risk. Patients may want

to gain access to health services from internet hospitals but may not want to get involved in medical disputes for undesired outcomes or disclose their personal information resulting in online fraud. To date, China has not established an effective quality control system for internet hospitals [12,15], and their potential medical risks threaten patient safety, thereby possibly affecting patients' intention to use the online inquiry services of internet hospitals. In this study, we define perceived information risk (PIR) as patients' perception of lack of control over their personal information (ie, privacy information on health care-seeking data) after they have adopted the online inquiry services provided by internet hospitals. Information privacy concern is a core predictor of PIR [33]. A study by Deng et al [13] found that privacy risk was negatively correlated with adoption intention toward mHealth services, and a study by Zhang et al [8] on the intention to use diabetes management apps found that perceived privacy risk negatively influenced behavioral intention. Therefore, we formulated the following hypotheses:

- H6a: PMR has a negative effect on the behavioral intention of patients to use online inquiry in internet hospitals.
- H6b: PIR has a negative effect on the attitude of patients toward online inquiry in internet hospitals.

Emotional preference (EP) could be defined as emotions that individuals are motivated to experience; these emotions strongly influence various human behaviors [34]. In health care settings, a patient's EP needs to be respected and be inaccessible protection from society. Several studies regarding patients with mental health problems suggest that the requirements of patients associated with their EPs, owing to issues such as stigma and social discrimination, affected their intention toward seeking help in time [35,36]. In internet hospitals, patients equipped with mobile devices can communicate with physicians without the constraints of time and space, which help to meet the requirements associated with their EPs, for example, privacy protection and avoiding embarrassment and stress to a certain extent; thus, those with EP may have a positive attitude toward online inquiry services of internet hospitals. Therefore, we established the following hypothesis:

- H7: EP has a positive effect on the attitude of patients toward online inquiry in internet hospitals.

With the rapid development of internet hospitals, the issue of their medical liability comes to the fore, threatening patient safety [3,37,38]. In addition to patients, physicians, and hospitals, the stakeholders of internet hospitals also involve a third-party medical platform, which could cause a more complex medical dispute. However, to date, the complaint system established by the Chinese government for internet hospitals regarding medical accidents and medical disputes is imperfect,

and the related regulations from traditional hospitals are often used for the management of medical behaviors in internet hospitals [12,15]. If patients find it challenging to defend their rights and interests in internet hospitals, they may have a negative attitude toward online inquiry. Therefore, we posed the following hypothesis:

- H8: Perceived medical liability (PML) has a negative effect on the attitude of patients toward online inquiry in internet hospitals.

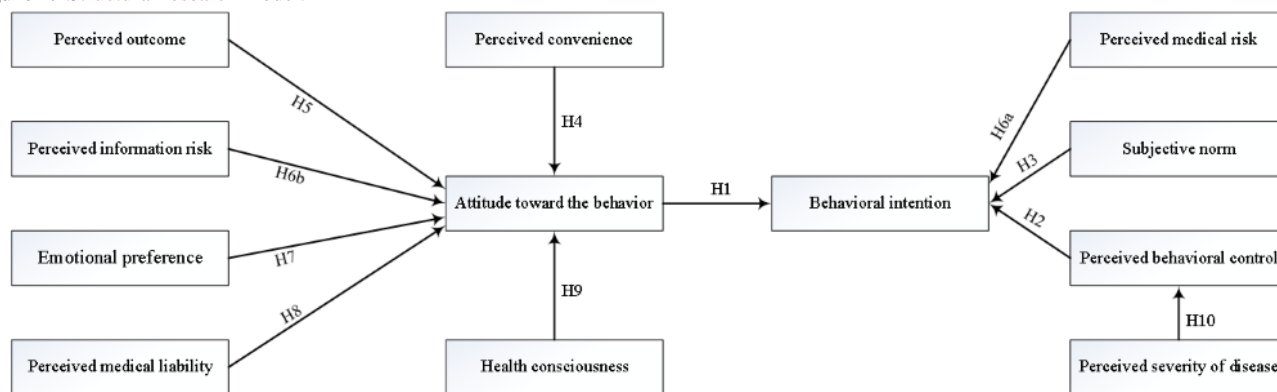
Health consciousness is defined as the degree to which health concerns are integrated into an individual's daily life, thus creating a motivation for health improvement [39]. A study by Guo et al [40] found such a moderation effect of health consciousness on behavioral attitude toward continuous use of mHealth services [40], and an investigation by Ahadzadeh et al [20] found that the partial effect of health consciousness on the intention toward health-related internet use was mediated by behavioral attitude [20]. Thus, we posed the following hypothesis:

- H9: Health consciousness has a positive effect on the attitude of patients toward online inquiry in internet hospitals.

As well as health consciousness, PSD is a driving factor for health-seeking behaviors [18]. PSD could be defined as patients' perception of the degree of seriousness of their untreated diseases, including opinions of clinical consequences and potential social consequences [27]. A higher perceived health risk of the disease consequences leads to a greater motivation to adopt a health-oriented behavior for access to health services. This hypothesis has been tested in several studies [19,20]. Several other studies on mHealth services also found that the PSD has a positive effect on behavioral intention [18,27]. In China, people usually tend to go to traditional high-level hospitals even for mild symptoms, owing to their lack of confidence in the quality of health care provided by primary care institutions [3]. When individuals perceive a more serious health threat, they might tend to consider the online inquiry service provided by internet hospitals challenging for the purpose of diagnosing or treating their diseases, and thus, they are likely to seek high-quality health services in traditional large hospitals rather than in internet hospitals. Thus, we posed the following hypothesis:

- H10: PSD has a negative effect on the patients' PBC of online inquiry in internet hospitals.

The 11 research hypotheses, with reference to the results of eliciting salient beliefs from patient interviews, are summarized in the research model (Figure 1).

**Figure 1.** Structural research model.

## Methods

### Development of the Study Questionnaire

We developed a survey questionnaire for the research model. We interviewed 19 patients effectively according to the connotation requirements of the TPB [41]; the results of eliciting

salient beliefs from patient interviews were used to design the questionnaire items, and the items were also discussed and approved by a focus group. The questionnaire items (Textbox 1) were measured with a 5-point Likert scale ranging from “strongly disagree” (1) to “strongly agree” (5). We then performed a presurvey to validate the designed questionnaire.

**Textbox 1.** Measurement items of the constructs.**Perceived convenience**

1. I think it is convenient to use the online inquiry services provided by internet hospitals anywhere without going to the hospital in person.
2. I think that internet hospitals could provide a vast amount of medical information and help me find well-known hospitals or doctors across the country.
3. I think it is time-saving to make an online inquiry through internet hospitals.
4. I think that using the online inquiry services provided by internet hospitals could be helpful to save expenses such as transportation and lodging fees.
5. I think that using the online inquiry services provided by internet hospitals could be helpful to decrease losses such as the loss of normal income due to sick leaves.

**Perceived outcome**

1. I think that the online inquiry services provided by internet hospitals are helpful to accurately understand the conditions.
2. I think that effective treatment recommendations could be obtained through the online inquiry services provided by internet hospitals.
3. I think that effective medication guidance could be obtained through the online inquiry services provided by internet hospitals.

**Perceived medical risk**

1. I am worried that physicians in internet hospitals pay less attention to my conditions.
2. I have doubts about the professionalism of internet hospitals or their physicians.
3. I have doubts about the authenticity of internet hospitals or their physicians.
4. I am worried that I may not be able to clearly explain my conditions to physicians when using online inquiry services provided by internet hospitals.
5. I am worried that, because the online inquiry services provided by internet hospitals cannot provide medical examinations, it may result in misdiagnosis and missed diagnosis.

**Perceived information risk**

1. I am worried that I would be deceived during the use of the online inquiry services provided by internet hospitals.
2. I am worried that the personal information could be disclosed after using the online inquiry services provided by internet hospitals.

**Emotional preference**

1. I think that using the online inquiry services provided by internet hospitals could avoid the tension associated with face-to-face communication with physicians.
2. I think that using the online inquiry services provided by internet hospitals could be helpful to protect personal information and avoid the embarrassment associated with face-to-face communication with physicians.

**Perceived medical liability**

1. I think the relevant laws and regulations about the online inquiry services provided by internet hospitals are at an imperfect stage.
2. I think it is difficult to divide the liability of medical accidents encountered between patients, physicians, hospitals, and network platforms when using the online inquiry services provided by internet hospitals.
3. I think it is difficult to protect my legal rights when a medical accident occurs while using the online inquiry services provided by internet hospitals.

**Attitude toward the behavior**

1. Overall, I think using the online inquiry services provided by internet hospitals is a better experience than those provided by traditional hospitals.
2. Overall, I think using the online inquiry services provided by internet hospitals is helpful for disease treatment.
3. Overall, I think it is meaningful to using the online inquiry services provided by internet hospitals.

**Subjective norm**

1. Both family and friends around me think that I should use the online inquiry services provided by internet hospitals.
2. Patients around me think that I should use the online inquiry services provided by internet hospitals.
3. Web friends think that I should use the online inquiry services provided by internet hospitals.

4. The government has been advocating the use of online inquiry services provided by internet hospitals.

#### Health consciousness

1. I am aware of and very concerned about my health problems.
2. I will try to manage and improve my wellness.

#### Perceived severity of disease

1. I do not think my health problems are serious.

#### Perceived behavioral control

1. I have appropriate hardware equipment (such as smartphone and computer) to use the online inquiry services provided by internet hospitals.
2. I could independently complete the online operation process of using the online inquiry services provided by internet hospitals.
3. I could clearly distinguish the valid and invalid information about the online inquiry services provided by internet hospitals.

#### Behavioral intention

1. If I am sick, I will choose an internet hospital for online inquiry.
2. If someone around me requires health services, I will recommend the online inquiry services provided by internet hospitals to them.

In the presurvey analysis, we selected the outpatients who visited a tertiary public hospital, patients who bought medicines in a large drugstore, and patients from communication groups with chronic diseases (including diabetes, hypertension, and heart disease) as respondents. The presurvey sample size was often suggested to be 5-10 times the number of observed variables of the scale, or greater than or equal to 100 [42], and thus, we selected 150 patients by convenience sampling, and 128 patients were volunteered to participate in the presurvey during December 2019. Finally, a validation sample of 117 patients was used for testing reliability and validity of the designed questionnaire. Reliability was measured using Cronbach alpha. If the Cronbach alpha of a construct increases significantly after the removal of an item, indicating that the item reduces the internal consistency of its measurement, the item should be further revised or removed. The Cronbach alpha of all constructs should be higher than .70 [43,44]. We used exploratory factor analysis with varimax-rotated components to measure the validity of the designed questionnaire, and its cumulative total of variance and factor loadings were used to assess the construct validity. If the cumulative total of variance of the principal components selected accounted for more than 70% of the total variance, the composition of the principal components as exogenous constructs was consistent with the constructs of the designed questionnaire, and if each item had a factor loading value of 0.50 or higher on one of the principal components but factor loading values below 0.50 on others, the validity of the designed questionnaire was considered acceptable, and no items were removed.

No items were revised or removed on the basis of the analysis result of the presurvey, and thus, the designed questionnaire with acceptable reliability and validity could be further used for data collection. Structural equation modeling (SEM) was then used to verify the research model. Data on participants' demographic characteristics such as gender, age, education level, marital status, living alone or with others, residence,

monthly average income, types of medical insurance, and online inquiry experience were also collected.

#### Data Collection

The target population included patients with chronic diseases including diabetes, hypertension, and heart disease, selected using the cluster sampling technique. We used the developed questionnaire to create an electronic questionnaire using the web-based survey tool wenjuanxing [45], a widely used website for conducting surveys in China. To collect enough sample data, the sample size was determined by general guidelines for SEM, and the samples were recommended to be at least 10-15 times as many as the items of the scale [46]. From February 12-29, 2020, we sent the survey link to Tencent QQ groups, WeChat groups, and Baidu Post Bar of patients with the aforementioned chronic diseases. Prior to the beginning of the survey, an informed consent of the managers of these patient social platforms was also requested and obtained. Moreover, we selected some general practitioners working in the community health center and physicians working in the tertiary hospital via convenience sampling, and these doctors were devoted to the treatment and management of the aforementioned chronic diseases. We then sent the survey website link to the doctors, and subsequently, they sent the website link to the targeted patient communication groups via WeChat. Participants could click on the website link using their phones to access and complete the electronic questionnaire. Before the survey, we introduced the purpose of the survey, provided the definition of internet hospitals, and guaranteed that the survey data would not be used for other purposes. After an individual's consent was obtained, the survey was conducted accordingly. A WeChat, Tencent QQ, or Baidu Post Bar account and mobile Internet Protocol address could be used to complete the questionnaire only once. The study was approved by the Medical Ethics Committee of Tongji Medical College of Huazhong University of Science & Technology.



## Statistical Analysis

Descriptive statistics were used to analyze the demographic characteristics of the respondents. Patients' intention to use the online inquiry services provided by internet hospitals was measured using two items (Textbox 1), with a higher score indicating elevated adoption intention. These respondents were divided into a younger group (younger than 50 years) and an older group. An independent *t* test was further used to evaluate the differences among adoption intention scores between the younger and the older groups.

SEM was used to verify the research model. Before evaluating the structural model, we assessed the measurement model to evaluate the reliability, convergent validity, and data fit indexes. We measured the reliability using the Cronbach alpha and composite reliability (CR). Values of .70 or higher for Cronbach alpha and values of 0.6 or higher for CR for all constructs were considered acceptable [8,43,44]. Convergent validity was measured using the average variance extracted (AVE) and standard loadings. The AVE of each construct and all the standard loadings should be greater than 0.50 [8,16,19]. The model fit was generally considered acceptable when the indexes met the criteria including  $\chi^2 / df < 3$ , root mean square error of approximation  $< 0.08$ , goodness-of-fit index (GFI)  $> 0.90$ , adjusted GFI  $> 0.90$ , Tucker-Lewis index  $> 0.90$ , comparative fit index  $> 0.90$ , and normed fit index  $> 0.90$  [13,27,47,48].

Otherwise, the research model was further revised. Moreover, we performed a bootstrap analysis with 1000 bootstrap bias-corrected samples to assess the direct, indirect, and total effect of variables (constructs) on the endogenous variables (ie, attitude and behavioral intention toward using online inquiry) [8,48]. Data analysis was done using SPSS V.24 and AMOS V.21 (IBM Corp).

## Results

### Sample Characteristics

In total, 812 responses (online survey) were received; of these, 174 responses were excluded because they contained identical answers to all questions, displayed certain logical contradictions, or because the time taken to answer the questionnaire was less than 100 seconds. The demographic characteristics of the 638 participants are shown in Table 1. Among these participants, 332 (52.0%) were aged 19-29 years, and 197 (30.9%) experienced the online inquiry services provided by internet hospitals. On average, the patients' intention to use the online inquiry services provided by internet hospitals (min 1, max 5) was moderate, with a mean score of 3.21 (SD 0.84), and differences were not observed between younger patients and older patients (mean 3.19, SD 0.84 vs mean 3.36, SD 0.82;  $P=.10$ ).

**Table 1.** Demographic characteristics of the participants (N=638).

Characteristics	Value, n (%)
<b>Gender</b>	
Male	241 (37.8)
Female	397 (62.2)
<b>Age (years)</b>	
≤18	14 (2.2)
19-29	332 (52.0)
30-39	109 (17.1)
40-49	110 (17.2)
50-59	62 (9.7)
≥60	11 (1.7)
<b>Education level</b>	
Up to secondary school	241 (37.8)
College	85 (13.3)
Undergraduate	206 (32.3)
Postgraduate and higher	106 (16.6)
<b>Marital status</b>	
Married	325 (50.9)
Unmarried	313 (49.1)
<b>Living alone or with others</b>	
Living alone	70 (11.0)
Living with family	510 (79.9)
Living with friends	58 (9.1)
<b>Residence</b>	
Rural	263 (41.2)
Urban	375 (58.8)
<b>Monthly average income, ¥ (US \$)</b>	
<2000 (285.40)	252 (39.5)
2000 (285.40)-5000 (713.60)	243 (38.1)
5001 (713.80)-8000 (1141.80)	91 (14.3)
>8000 (1141.80)	52 (8.2)
<b>Types of medical insurance</b>	
Basic medical insurance for urban and rural residents	362 (56.7)
Basic medical insurance for urban employees	196 (30.7)
Commercial medical insurance	24 (3.8)
Others	56 (8.8)
<b>Online inquiry experience</b>	
Yes	197 (30.9)
No	441 (69.1)

## Measurement Model Testing

Table 2 shows the results of the measurement model. The Cronbach alpha and CR of each construct were higher than the

recommended values, indicating excellent reliability. The AVEs of constructs, except PMR, SN, and PBC, and all the standard loadings were greater than the recommended values, indicating acceptable convergent validity. Specifically, for example, the

values of Cronbach alpha and CR in PC were .866 and .877, respectively, and higher than the recommended values of .70, indicating the measurement model had excellent reliability; moreover, the AVE was 0.597 and greater than the threshold value of 0.50, indicating the measurement model had acceptable convergent validity.

The goodness-of-fit results of the research model are shown in [Table 3](#). All fit indexes were below the recommended values, indicating that the data collected did not fit with the research model, and thus further revision of the research model was needed.

**Table 2.** Statistical results of the research model

Constructs and items	Standard loadings	<i>P</i> value	Score, mean (SD)	Cronbach alpha	CR <sup>a</sup>	AVE <sup>b</sup>
<b>PC<sup>c</sup></b>				.866	0.877	0.597
PC1 <sup>d</sup>	0.762	N/A <sup>e</sup>	3.82 (0.94)			
PC2	0.711	<.001	3.97 (0.77)			
PC3	0.782	<.001	3.98 (0.83)			
PC4	0.757	<.001	4.11 (0.75)			
PC5	0.753	<.001	3.91 (0.79)			
<b>PO<sup>f</sup></b>				.850	0.773	0.533
PO1	0.755	N/A	3.22 (0.96)			
PO2	0.848	<.001	3.53 (0.86)			
PO3	0.833	<.001	3.57 (0.82)			
<b>PMR<sup>g</sup></b>				.847	0.794	0.438
PMR1	0.663	N/A	2.59 (0.94)			
PMR2	0.816	<.001	2.75 (0.91)			
PMR3	0.832	<.001	2.76 (0.93)			
PMR4	0.648	<.001	2.35 (0.94)			
PMR5	0.659	<.001	2.12 (0.83)			
<b>PIR<sup>h</sup></b>				.796	0.703	0.545
PIR1	0.888	N/A	2.54 (0.96)			
PIR2	0.747	<.001	2.30 (0.90)			
<b>EP<sup>i</sup></b>				.770	0.712	0.562
EP1	0.649	<.001	3.41 (0.90)			
EP2	0.946	N/A	3.48 (0.89)			
<b>PML<sup>j</sup></b>				.840	0.763	0.518
PML1	0.718	<.001	2.09 (0.79)			
PML2	0.845	N/A	2.15 (0.81)			
PML3	0.832	<.001	2.13 (0.81)			
<b>AB<sup>k</sup></b>				.813	0.887	0.723
AB1	0.680	N/A	3.61 (0.72)			
AB2	0.735	<.001	3.77 (0.65)			
AB3	0.755	<.001	3.81 (0.66)			
<b>SN<sup>l</sup></b>				.858	0.796	0.496
SN1	0.763	N/A	3.28 (0.85)			
SN2	0.880	<.001	3.11 (0.90)			
SN3	0.802	<.001	3.11 (0.91)			
SN4	0.658	<.001	3.25 (0.86)			
<b>HC<sup>m</sup></b>				.801	0.790	0.661
HC1	1.000	<.001	4.05 (0.69)			
HC2	0.670	N/A	4.12 (0.63)			
<b>PSD<sup>n</sup></b>				— <sup>o</sup>	0.630	0.630
PSD1	0.744	N/A	2.54 (0.91)			

Constructs and items	Standard loadings	<i>P</i> value	Score, mean (SD)	Cronbach alpha	CR <sup>a</sup>	AVE <sup>b</sup>
<b>PBC<sup>p</sup></b>				.731	0.680	0.422
PBC1	0.641	N/A	3.87 (0.82)			
PBC2	0.863	<.001	3.74 (0.84)			
PBC3	0.599	<.001	3.35 (0.86)			
<b>BI<sup>q</sup></b>				.846	0.695	0.533
BI1	0.796	N/A	3.19 (0.88)			
BI2	0.831	<.001	3.23 (0.92)			

<sup>a</sup>CR: composite reliability.

<sup>b</sup>AVE: average variance extracted.

<sup>c</sup>PC: perceived convenience.

<sup>d</sup>Numbers refer to the numbered lists under each construct in [Textbox 1](#).

<sup>e</sup>N/A: not applicable.

<sup>f</sup>PO: perceived outcome.

<sup>g</sup>PMR: perceived medical risk.

<sup>h</sup>PIR: perceived information risk.

<sup>i</sup>EP: emotional preference.

<sup>j</sup>PML: perceived medical liability.

<sup>k</sup>AB: attitude toward the behavior.

<sup>l</sup>SN: subjective norm.

<sup>m</sup>HC: health consciousness.

<sup>n</sup>PSD: perceived severity of disease.

<sup>o</sup>Not available because of only an item.

<sup>p</sup>PBC: perceived behavioral control.

<sup>q</sup>BI: behavioral intention.

**Table 3.** Goodness-of-fit results of the research model.

Fit indexes	$\chi^2$ / df	RMSEA <sup>a</sup>	GFI <sup>b</sup>	AGFI <sup>c</sup>	TLI <sup>d</sup>	CFI <sup>e</sup>	NFI <sup>f</sup>
Research model	5.185	0.0813	0.767	0.732	0.786	0.803	0.768
Recommended value	<3	<0.08	>0.90	>0.90	>0.90	>0.90	>0.90

<sup>a</sup>RMSEA: root mean square error of approximation.

<sup>b</sup>GFI: goodness-of-fit index.

<sup>c</sup>AGFI: adjusted goodness-of-fit index.

<sup>d</sup>TLI: Tucker-Lewis index.

<sup>e</sup>CFI: comparative fit index.

<sup>f</sup>NFI: normed fit index.

## Model Modification

To improve the goodness of fit of the research model, hypothesis paths of H3, H5, H8, and H10 were removed. Furthermore, the constructs including PML, PO, and SN were removed. Considering that the AVEs of the constructs PMR and PBC were considerably lower than the recommended values, indicating the measurement indexes with interference, the items including PMR2, PMR3, and PBC3 were removed. Moreover, the items including attitude toward behavior 1, PC4, and PC5

were removed on the basis of a focus group discussion. Additionally, the relationships between PBC and PC, between PBC and PMR, between PSD and PMR, between PSD and PBC, between PMR and PIR, between PMR and EP, and between EP and PC were established.

As shown in [Table 4](#), after modification, the goodness of fit of the research model was improved. All fit indexes were greater than the recommended values, indicating that the eight-construct model was a good fit for the data collected.



**Table 4.** Goodness-of-fit results of the revised research model.

Fit indexes	$\chi^2 / df$	RMSEA <sup>a</sup>	GFI <sup>b</sup>	AGFI <sup>c</sup>	TLI <sup>d</sup>	CFI <sup>e</sup>	NFI <sup>f</sup>
Research model	2.837	0.054	0.942	0.920	0.933	0.946	0.920
Recommended value	<3	<0.08	>0.90	>0.90	>0.90	>0.90	>0.90

<sup>a</sup>RMSEA: root mean square error of approximation.<sup>b</sup>GFI: goodness-of-fit index.<sup>c</sup>AGFI: adjusted goodness-of-fit index.<sup>d</sup>TLI: Tucker-Lewis index.<sup>e</sup>CFI: comparative fit index.<sup>f</sup>NFI: normed fit index.

## Structural Model Testing

Table 5 shows that three (H3, H5, and H10) of the eleven research hypotheses were rejected. The standardized path

coefficients of all other relationships were significant at  $P < .05$ . The relationships among constructs in the final research model are illustrated in Figure 2.

**Table 5.** Hypothesis testing results of the research model

Hypothesis paths	Standardized path coefficients	<i>P</i> value	Results
H1 <sup>a</sup> Attitude toward the behavior (+ <sup>b</sup> ) → Behavioral intention	0.394	<.001	H1 supported
H2 Perceived behavioral control (+) → Behavioral intention	0.624	<.001	H2 supported
H3 Subjective norm (+) → Behavioral intention	— <sup>c</sup>	—	H3 not supported
H4 Perceived convenience (+) → Attitude toward the behavior	0.525	<.001	H4 supported
H5 Perceived outcome (+) → Attitude toward the behavior	—	—	H5 not supported
H6a Perceived medical risk (– <sup>d</sup> ) → Behavioral intention	–0.192	<.001	H6a supported
H6b Perceived information risk (–) → Attitude toward the behavior	–0.182	<.001	H6b supported
H7 Emotional preference (+) → Attitude toward the behavior	0.206	<.001	H7 supported
H8 Perceived medical liability (–) → Attitude toward the behavior	—	—	H8 not supported
H9 Health consciousness (+) → Attitude toward the behavior	0.243	<.001	H9 supported
H10 Perceived severity of disease (–) → Perceived behavioral control	—	—	H10 not supported

<sup>a</sup>H: hypothesis.<sup>b</sup>+: positive effect.<sup>c</sup>Hypothesis paths of H3, H5, H8, and H10 were removed in the revised model.<sup>d</sup>–: negative effect.

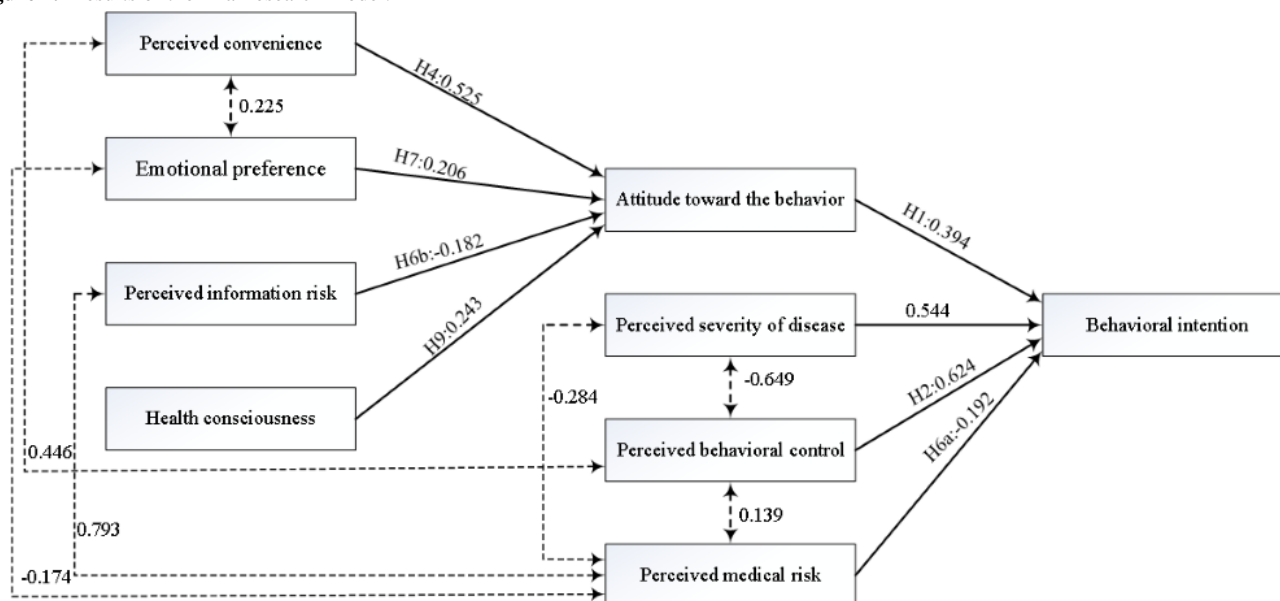
**Figure 2.** Results of the final research model.

Table 6 shows the total, direct, and indirect effects (standardized path coefficients) of the model variables (constructs) on attitude toward the behavior and behavioral intention, respectively. PC, PIR, EP, and health consciousness explained 45.9% of the variance in attitude toward the behavior. Among these four variables, PC had the strongest effect on attitude toward the behavior ( $\beta=.525$ ,  $P=.002$ ). EP and health consciousness had moderate effects on attitude toward the behavior ( $\beta=.206$ ,  $P=.002$  and  $\beta=.243$ ,  $P=.002$ , respectively). Additionally, PIR had a slightly negative direct effect on attitude toward the behavior ( $\beta=-.182$ ,  $P=.001$ ).

Further, PC, PIR, EP, and health consciousness had indirect effects on behavioral intention, and these effects were mediated by attitude toward the behavior. PC had the highest indirect effect on behavioral intention ( $\beta=.207$ ,  $P=.001$ ) among the four constructs.

Overall, attitude toward the behavior, PMR, PSD, and PBC explained 60.5% of the variance in behavioral intention. PBC and PSD had the highest total effects on behavioral intention ( $\beta=.624$ ,  $P=.004$  and  $\beta=.544$ ,  $P=.003$ , respectively). Attitude toward the behavior had a moderate total effect on behavioral intention ( $\beta=.394$ ,  $P=.002$ ), whereas PMR had a slightly negative total effect on behavioral intention ( $\beta=-.192$ ,  $P=.005$ ).

Within expectations, PMR had the strongest positive relationship with PIR ( $r=0.793$ ,  $P=.001$ ). Moreover, PMR had a slightly negative relationship with EP and PSD ( $r=-0.174$ ,  $P=.002$  and  $r=-0.284$ ,  $P=.002$ , respectively) but a positive relationship with PBC ( $r=0.139$ ,  $P=.002$ ). Additionally, PBC had a moderate positive relationship with PC ( $r=0.446$ ,  $P=.001$ ) but was strongly and negatively correlated with PSD ( $r=-0.649$ ,  $P=.002$ ). PC also had a slightly positive relationship with EP ( $r=0.225$ ,  $P=.002$ ).

**Table 6.** Total, direct, and indirect effects of the model variables.

Variables	AB <sup>a</sup> ( $R^2=45.9\%$ )		BI <sup>b</sup> ( $R^2=60.5\%$ )	
	$\beta$	<i>P</i> value	$\beta$	<i>P</i> value
<b>PC<sup>c</sup></b>				
Direct	.525	.002	— <sup>d</sup>	—
Indirect	—	—	.207	.001
Total	.525	.002	.207	.001
<b>PMR<sup>e</sup></b>				
Direct	—	—	-.192	.005
Indirect	—	—	—	—
Total	—	—	-.192	.005
<b>PIR<sup>f</sup></b>				
Direct	-.182	.001	—	—
Indirect	—	—	-.072	.001
Total	-.182	.001	-.072	.001
<b>EP<sup>g</sup></b>				
Direct	.206	.002	—	—
Indirect	—	—	.081	.001
Total	.206	.002	.081	.001
<b>HC<sup>h</sup></b>				
Direct	.243	.002	—	—
Indirect	—	—	.096	.001
Total	.243	.002	.096	.001
<b>PSD<sup>i</sup></b>				
Direct	—	—	.544	.003
Indirect	—	—	—	—
Total	—	—	.544	.003
<b>PBC<sup>j</sup></b>				
Direct	—	—	.624	.004
Indirect	—	—	—	—
Total	—	—	.624	.004
<b>AB</b>				
Direct	—	—	.394	.002
Indirect	—	—	—	—
Total	—	—	.394	.002

<sup>a</sup>AB: attitude toward behavior.<sup>b</sup>BI: behavioral intention.<sup>c</sup>PC: perceived convenience.<sup>d</sup>Not available.<sup>e</sup>PMR: perceived medical risk.<sup>f</sup>PIR: perceived information risk.<sup>g</sup>EP: emotional preference.<sup>h</sup>HC: health consciousness.<sup>i</sup>PSD: perceived severity of disease.

<sup>j</sup>PBC: perceived behavioral control.

## Discussion

### Principal Findings

Our study shows that the patients' intention to use the online inquiry services provided by internet hospitals was relatively low, with room for development. Internet hospitals in China show an emerging trend for adequately fulfilling people's demands for health services [9,10]. Thus, the determinants of patients' intention to use online inquiry services provided by internet hospitals are needed that help provide opportunities to promote the development of internet hospitals. Our findings show the patients' intention to use the online inquiry services provided by internet hospitals was determined by the following factors: attitude toward the behavior, PBC, PSD, PC, PIR, PMR, health consciousness, and EP.

Our study found that two TPB factors (ie, attitude toward the behavior and PBC) had significant effects on patients' intention to use the online inquiry services provided by internet hospitals, whereas the factor (ie, SN) had an insignificant effect on behavioral intention. PBC was the most important determinant of patients' intention to use the online inquiry services provided by internet hospitals. Several studies on mHealth services also indicated PBC was the major determinant of behavioral intention [24,26,27]. If patients find it easy to use online inquiry services, their willingness toward online inquiry services provided by internet hospitals will be stronger. Many studies found that the perceived ease of use of mHealth technology is a determinant of patients' adoption intention toward mHealth services [8,21,49,50], especially among older adults [16]. Similar to mHealth service, the ease of using online inquiry services provided by internet hospitals must be promoted, thereby potentially improving the PBC of users in using these services. Therefore, internet hospitals should be patient-centered, further optimize online service delivery process to improve the ease of using online inquiry services, and provide patients with multiple convenient and easy-to-use channels for online inquiry, such as mHealth service apps and WeChat official accounts.

PSD had a strong positive total effect on patients' intention to use the online inquiry services provided by internet hospitals, and this effect was also direct but not mediated by PBC. This finding is consistent with the findings of a study by Zhang et al [27] regarding mHealth service adoption in China. Another study by Liu et al [51] also found a positive effect of PSD on propensity of users in online health care communities to seek health information. This finding suggests that perceived disease severity motivates patients to adopt online inquiry services provided by internet hospitals. Internet hospitals, owing to the integration of a large number of high-quality medical resources across the country could competently fulfill the increasing patient demands for high-quality health services. It is difficult for patients in rural and remote areas of China to access high-quality health services [4,5]. Thus, internet hospitals should be dedicated to enhance the awareness of individuals regarding disease vulnerability and emphasize the advantages of online inquiry services to change the indifferent attitude and distrust of certain people toward the online inquiry services provided

by internet hospitals. With the increasing health service requirements of patients, physicians tend to have a heavier workload [44], and unreasonable expectations regarding online services may further increase the physicians' workload. Therefore, internet hospitals should also ensure a reasonable assembly of high-quality experts and coordinate their online and offline medical resources. Ensuring a reasonable assembly of high-quality medical resources in internet hospitals will not only optimize the physicians' workload but also meet the increasing patient demands for renowned experts, thereby promoting patients' willingness toward online inquiry services provided by internet hospitals.

PC also had a moderate positive total effect on patients' intention to use online inquiry services provided by internet hospitals. This finding is in line with the findings of a study of Valikodath et al [29] on patients' willingness toward telemedicine for diabetic retinopathy. Another study by Clevenger et al [52] also demonstrated that PC was positively associated with acceptance of adolescent vaccination outside of the traditional medical home. This finding suggests the need for internet hospital managers to identify the gaps in convenience of health care service delivery and further improve the organizational service delivery design. During the special survey time called "Wuhan on lockdown," this policy might contribute to a greater intention of patients in the Wuhan area to use online inquiry services provided by internet hospitals owing to the inconvenience of seeking health services in traditional hospitals. With the normalization of the prevention and control of the COVID-19 epidemic, service convenience has become an important topic in the context of health care, and internet hospitals play an increasingly important role in delivering health services to the public. Therefore, internet hospitals should also promote and achieve one-step health care services via the integration of medical service, medicine, and medical insurance. Improving service convenience will help people reduce the risks of nosocomial cross-infection and, thus, benefit the promotion of patients' intention to use the online inquiry services provided by internet hospitals.

Although the indirect effect of PIR on patients' intention to use online inquiry services provided by internet hospitals was weak (and this effect was mediated by attitude toward the behavior), PMR had a significantly negative stronger direct effect on behavioral intention. Similarly, several studies on mHealth services found the negative effect of perceived privacy risk on behavioral intention [13,53]. Another study by Zhang et al [8] also demonstrated the negative effect of perceived privacy risk on patients' intention to use diabetes management apps. With the rapid development of health information technology, patients are increasingly aware of and concerned about their personal information protection such as privacy protection on health care-seeking data. Although our research found that PIR had only a weak effect on patients' intention to use the online inquiry services provided by internet hospitals, solid patient measures to protect patient information are needed. Additionally, our research found that PMR had a stronger effect on patients' intention to use online inquiry services provided by internet

hospitals. However, to date, the Chinese government has not established an effective quality control system for internet hospitals, and related regulations from traditional hospitals are often used for the management of medical behaviors in internet hospitals, although the government have realized the problem [12,15], which could lead to an increase in probability of medical risks, threatening patient safety [38], thereby potentially affecting risk-conscious patients' intention to use the online inquiry services. Thus, to ensure patient safety regarding the use of online inquiry services provided by internet hospitals, the relevant quality control standards and regulations for internet hospitals should be further developed and improved.

Health consciousness had slightly positive indirect effects on patients' intention to use the online inquiry services provided by internet hospitals, and these effects were mediated by attitude toward the behavior. A previous study by Ahadzadeh et al [20] suggested that health consciousness positively affected patients' intention of the internet use for health-related purposes. Several studies concerning disease information-seeking behaviors also demonstrated an effect of health consciousness on behavioral intention [54,55]. Patients with a higher level of health consciousness are more likely to be aware of and concerned about their wellness [56], and if they believe that their health service requirements can be adequately fulfilled by the online inquiry services provided by internet hospitals, their willingness toward online inquiry will be stronger. With the increasing health awareness of people in China, their willingness toward online inquiry services provided by internet hospitals will be further improved.

Our study also found that EP had a slightly positive total effect on patients' intention to use the online inquiry services provided by internet hospitals. Emotion is an essential element of human behavior. Regarding health behaviors, some patients with genital diseases or patients who are sensitive about a certain disease have more awareness about self-precautionary measures and privacy protection or are more unwilling to disclose their condition to others, eventually affecting their help-seeking intentions. Such unfavorable results have been found in several studies on patients with mental health problems [35,36]. With internet hospitals, such patients could be provided with a more confidential approach of seeking health services through graph-text, voice, and video services [9], which protects their privacy and fulfills their emotional requirements to a certain extent; thus, patients with EPs would be more willing to gain access to the online services provided by internet hospitals.

## Limitations

This study has several limitations. First, among the twelve constructs, PSD included only one item, which may have resulted in a measurement error. Second, although the research model was developed based on the TPB, some factors influencing patients' intention to use online inquiry services provided by internet hospitals may not have been identified; for example, "medical expenses" was included in PC but deleted in the revised model; the medical service price is an important determinant of patients' intention to accept a medical service in traditional hospitals, especially whether this service is covered by medical insurance payment [57], but was not as a factor included into the research model, and thus, continuous research is needed to help further explore these determinants. Third, online survey was used as the data collection method owing to the COVID-19 epidemic, and the target population included individuals with chronic diseases such as diabetes, hypertension, and heart disease; although the age of patients with chronic diseases was usually old, the younger patients in targeted patient communication groups might be more willing to participate in the online survey; moreover, we used cluster sampling rather than proportional sampling, and although differences were not observed between younger patients and older patients, our results may have been biased by the age distributions. Further, data collection was self-reported by patients via online survey, which might have a recall bias. In addition, to improve the goodness of fit of the research model, a few constructs including PML, PO, and SN were removed, which may have produced an insignificant result; therefore, further research is necessary to identify whether these constructs significantly affect patients' intention to use online inquiry services provided by internet hospitals.

## Conclusions

PBC and PSD are the most important determinants of patients' intention to use online inquiry services provided by internet hospitals. Therefore, internet hospitals should further optimize the design of online service delivery and ensure a reasonable assembly of high-quality experts, which will benefit the promotion of patients' adoption intention toward online inquiry services for health purposes. PC, EP, and perceived risks also have effects on behavioral intention. Therefore, the relevant quality control standards and regulations for internet hospitals should be further developed and improved, and the measures to protect personal information should be strengthened to ensure patient safety. Our study supports the use of the planned behavior theory in explaining patients' intention to use the online inquiry services provided by internet hospitals.

## Acknowledgments

We thank all respondents for voluntarily participating in the survey. This study was supported by the National Natural Science Foundation of China (Grant Number 71774062) and the Fundamental Research Funds for the Central Universities (Grant Number 2018KFYYXJJ065).

## Conflicts of Interest

None declared.



## References

1. Global action plan for the prevention and control of NCDs 2013-2020. World Health Organization. 2020. URL: [http://www.who.int/nmh/events/ncd\\_action\\_plan/en/](http://www.who.int/nmh/events/ncd_action_plan/en/) [accessed 2020-04-12]
2. Sun J, Guo Y, Wang X, Zeng Q. mHealth for aging China: opportunities and challenges. *Aging Dis* 2016 Jan;7(1):53-67 [FREE Full text] [doi: [10.14336/AD.2015.1011](https://doi.org/10.14336/AD.2015.1011)] [Medline: [26816664](https://pubmed.ncbi.nlm.nih.gov/26816664/)]
3. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. *Lancet Glob Health* 2015 Aug;3(8):e445-e446 [FREE Full text] [doi: [10.1016/S2214-109X\(15\)00042-X](https://doi.org/10.1016/S2214-109X(15)00042-X)] [Medline: [26187488](https://pubmed.ncbi.nlm.nih.gov/26187488/)]
4. Ye Q, Deng Z, Chen Y, Liao J, Li G, Lu Y. How resource scarcity and accessibility affect patients' usage of mobile health in China: resource competition perspective. *JMIR mHealth uHealth* 2019 Aug 09;7(8):e13491 [FREE Full text] [doi: [10.2196/13491](https://doi.org/10.2196/13491)] [Medline: [31400104](https://pubmed.ncbi.nlm.nih.gov/31400104/)]
5. Yip W, Hsiao W. Harnessing the privatisation of China's fragmented health-care delivery. *Lancet* 2014 Aug 30;384(9945):805-818 [FREE Full text] [doi: [10.1016/S0140-6736\(14\)61120-X](https://doi.org/10.1016/S0140-6736(14)61120-X)] [Medline: [25176551](https://pubmed.ncbi.nlm.nih.gov/25176551/)]
6. Li Q, Xie P. Outpatient workload in China. *Lancet* 2013 Jun 08;381(9882):1983-1984. [doi: [10.1016/S0140-6736\(13\)61198-8](https://doi.org/10.1016/S0140-6736(13)61198-8)] [Medline: [23746895](https://pubmed.ncbi.nlm.nih.gov/23746895/)]
7. National Health Commission of the People's Republic of China. *China Health Statistics Yearbook*. Beijing, China: Chinese Peking Union Medical College Press; 2019.
8. Zhang Y, Liu C, Luo S, Xie Y, Liu F, Li X, et al. Factors influencing patients' intentions to use diabetes management apps based on an extended unified theory of acceptance and use of technology model: web-based survey. *J Med Internet Res* 2019 Aug 13;21(8):e15023 [FREE Full text] [doi: [10.2196/15023](https://doi.org/10.2196/15023)] [Medline: [31411146](https://pubmed.ncbi.nlm.nih.gov/31411146/)]
9. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: cross-sectional survey. *J Med Internet Res* 2017 Jul 04;19(7):e239 [FREE Full text] [doi: [10.2196/jmir.7854](https://doi.org/10.2196/jmir.7854)] [Medline: [28676472](https://pubmed.ncbi.nlm.nih.gov/28676472/)]
10. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet hospitals help prevent and control the epidemic of COVID-19 in China: multicenter user profiling study. *J Med Internet Res* 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: [10.2196/18908](https://doi.org/10.2196/18908)] [Medline: [32250962](https://pubmed.ncbi.nlm.nih.gov/32250962/)]
11. Iwaya LH, Gomes MAL, Simplício MA, Carvalho TCMB, Dominicini CK, Sakuragui RRM, et al. Mobile health in emerging countries: a survey of research initiatives in Brazil. *Int J Med Inform* 2013 May;82(5):283-298. [doi: [10.1016/j.ijmedinf.2013.01.003](https://doi.org/10.1016/j.ijmedinf.2013.01.003)] [Medline: [23410658](https://pubmed.ncbi.nlm.nih.gov/23410658/)]
12. Qiu Y, Liu Y, Ren W, Qiu Y, Ren J. Internet-based and mobile-based general practice: cross-sectional survey. *J Med Internet Res* 2018 Sep 25;20(9):e266 [FREE Full text] [doi: [10.2196/jmir.8378](https://doi.org/10.2196/jmir.8378)] [Medline: [30257819](https://pubmed.ncbi.nlm.nih.gov/30257819/)]
13. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mHealth services in China: empirical study. *JMIR mHealth uHealth* 2018 Aug 29;6(8):e172 [FREE Full text] [doi: [10.2196/mhealth.9316](https://doi.org/10.2196/mhealth.9316)] [Medline: [30158101](https://pubmed.ncbi.nlm.nih.gov/30158101/)]
14. Sun S, Yu K, Xie Z, Pan X. China empowers internet hospital to fight against COVID-19. *J Infect* 2020 Jul;81(1):e67-e68 [FREE Full text] [doi: [10.1016/j.jinf.2020.03.061](https://doi.org/10.1016/j.jinf.2020.03.061)] [Medline: [32251688](https://pubmed.ncbi.nlm.nih.gov/32251688/)]
15. China Internet hospital development research report. Institute of Healthcare. 2020. URL: <http://zk.cn-healthcare.com/doc-show-39773.html> [accessed 2020-04-12]
16. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. *Int J Med Inform* 2017 May;101:75-84. [doi: [10.1016/j.ijmedinf.2017.02.002](https://doi.org/10.1016/j.ijmedinf.2017.02.002)] [Medline: [28347450](https://pubmed.ncbi.nlm.nih.gov/28347450/)]
17. Guo X, Sun Y, Wang N, Peng Z, Yan Z. The dark side of elderly acceptance of preventive mobile health services in China. *Electron Markets* 2012 Dec 11;23(1):49-61. [doi: [10.1007/s12525-012-0112-4](https://doi.org/10.1007/s12525-012-0112-4)]
18. Deng Z. Understanding public users' adoption of mobile health service. *Int J Mobile Commun* 2013;11(4):351. [doi: [10.1504/IJMC.2013.055748](https://doi.org/10.1504/IJMC.2013.055748)]
19. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. *JMIR mHealth uHealth* 2017 Dec 06;5(12):e177 [FREE Full text] [doi: [10.2196/mhealth.7886](https://doi.org/10.2196/mhealth.7886)] [Medline: [29212629](https://pubmed.ncbi.nlm.nih.gov/29212629/)]
20. Ahadzadeh AS, Pahlevan Sharif S, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. *J Med Internet Res* 2015 Feb 19;17(2):e45 [FREE Full text] [doi: [10.2196/jmir.3564](https://doi.org/10.2196/jmir.3564)] [Medline: [25700481](https://pubmed.ncbi.nlm.nih.gov/25700481/)]
21. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mHealth technology use. *Stud Health Technol Inform* 2015;216:467-471 [FREE Full text] [Medline: [26262094](https://pubmed.ncbi.nlm.nih.gov/26262094/)]
22. Armitage C, Conner M. Efficacy of the Theory of Planned Behaviour: a meta-analytic review. *Br J Soc Psychol* 2001 Dec;40(Pt 4):471-499. [doi: [10.1348/014466601164939](https://doi.org/10.1348/014466601164939)] [Medline: [11795063](https://pubmed.ncbi.nlm.nih.gov/11795063/)]
23. Ajzen I. The theory of planned behavior. *Organizational Behav Hum Decision Processes* 1991 Dec;50(2):179-211. [doi: [10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)]
24. Jen W, Hung M. An empirical study of adopting mobile healthcare service: the family's perspective on the healthcare needs of their elderly members. *Telemed J E Health* 2010;16(1):41-48. [doi: [10.1089/tmj.2009.0093](https://doi.org/10.1089/tmj.2009.0093)] [Medline: [20070162](https://pubmed.ncbi.nlm.nih.gov/20070162/)]
25. Wu I, Li J, Fu C. The adoption of mobile healthcare by hospital's professionals: an integrative perspective. *Decision Support Syst* 2011 Jun;51(3):587-596. [doi: [10.1016/j.dss.2011.03.003](https://doi.org/10.1016/j.dss.2011.03.003)]

26. Deng Z, Mo X, Liu S. Comparison of the middle-aged and older users' adoption of mobile health services in China. *Int J Med Inform* 2014 Mar;83(3):210-224. [doi: [10.1016/j.ijmedinf.2013.12.002](https://doi.org/10.1016/j.ijmedinf.2013.12.002)] [Medline: [24388129](https://pubmed.ncbi.nlm.nih.gov/24388129/)]
27. Zhang X, Liu S, Wang L, Zhang Y, Wang J. Mobile health service adoption in China: integration of theory of planned behavior, protection motivation theory and personal health differences. *Online Inf Rev* 2019 Aug 29;44(1):1-23. [doi: [10.1108/oir-11-2016-0339](https://doi.org/10.1108/oir-11-2016-0339)]
28. Chang C, Yan C, Tseng J. Perceived convenience in an extended technology acceptance model: mobile technology and English learning for college students. *Australasian J Educ Technol* 2012 Jul 26;28(5):809-826. [doi: [10.14742/ajet.818](https://doi.org/10.14742/ajet.818)]
29. Valikodath NG, Leveque TK, Wang SY, Lee PP, Newman-Casey PA, Hansen SO, et al. Patient attitudes toward telemedicine for diabetic retinopathy. *Telemed J E Health* 2017 Mar;23(3):205-212 [FREE Full text] [doi: [10.1089/tmj.2016.0108](https://doi.org/10.1089/tmj.2016.0108)] [Medline: [27336678](https://pubmed.ncbi.nlm.nih.gov/27336678/)]
30. Wu Z. Chinese customer's attitude and adopt intention on mobile commerce. : IEEE Presented at: 6th International Conference on Service Systems and Service Management; 2009; Xiamen, China p. 731-734. [doi: [10.1109/icsssm.2009.5174977](https://doi.org/10.1109/icsssm.2009.5174977)]
31. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy assessment in mobile health apps: scoping review. *JMIR mHealth uHealth* 2020 Jul 02;8(7):e18868 [FREE Full text] [doi: [10.2196/18868](https://doi.org/10.2196/18868)] [Medline: [32459640](https://pubmed.ncbi.nlm.nih.gov/32459640/)]
32. Kotz D, Gunter CA, Kumar S, Weiner JP. Privacy and security in mobile health: a research agenda. *Computer (Long Beach Calif)* 2016 Jun;49(6):22-30 [FREE Full text] [doi: [10.1109/MC.2016.185](https://doi.org/10.1109/MC.2016.185)] [Medline: [28344359](https://pubmed.ncbi.nlm.nih.gov/28344359/)]
33. Libaque-Sáenz CF, Wong SF, Chang Y, Ha YW, Park M. Understanding antecedents to perceived information risks: an empirical study of the Korean telecommunications market. *Inf Dev* 2014 Mar 21;32(1):91-106. [doi: [10.1177/0266666913516884](https://doi.org/10.1177/0266666913516884)]
34. Kim H, Choi Y. Exploring emotional preference for smartphone applications. : IEEE; 2012 Presented at: 2012 IEEE Consumer Communications and Networking Conference (CCNC); 2012; Las Vegas, NV, USA p. 245-249. [doi: [10.1109/ccnc.2012.6181095](https://doi.org/10.1109/ccnc.2012.6181095)]
35. Schnyder N, Panczak R, Groth N, Schultze-Lutter F. Association between mental health-related stigma and active help-seeking: systematic review and meta-analysis. *Br J Psychiatry* 2017 Apr;210(4):261-268. [doi: [10.1192/bjp.bp.116.189464](https://doi.org/10.1192/bjp.bp.116.189464)] [Medline: [28153928](https://pubmed.ncbi.nlm.nih.gov/28153928/)]
36. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychol Med* 2015 Jan;45(1):11-27. [doi: [10.1017/S0033291714000129](https://doi.org/10.1017/S0033291714000129)] [Medline: [24569086](https://pubmed.ncbi.nlm.nih.gov/24569086/)]
37. Wang A, Zheng X. Discussion and analysis of network medical treatment. *Chin Hospitals* 2015;19(10):51-53. [doi: [10.3969/j.issn.1671-0592.2015.10.019](https://doi.org/10.3969/j.issn.1671-0592.2015.10.019)]
38. Xu L. Internet hospital: challenges and opportunities in China. 2016 Presented at: 2016 5th International Conference on Health Information Science (HIS); 2016; Shanghai, China p. 85-90. [doi: [10.1007/978-3-319-48335-1\\_9](https://doi.org/10.1007/978-3-319-48335-1_9)]
39. Jayanti RK, Burns AC. The antecedents of preventive health care behavior: an empirical study. *J Acad Marketing Sci* 1998 Jan 01;26(1):6-15. [doi: [10.1177/0092070398261002](https://doi.org/10.1177/0092070398261002)]
40. Guo X, Chen S, Zhang X, Ju X, Wang X. Exploring patients' intentions for continuous usage of mHealth services: elaboration-likelihood perspective study. *JMIR mHealth uHealth* 2020 Apr 06;8(4):e17258 [FREE Full text] [doi: [10.2196/17258](https://doi.org/10.2196/17258)] [Medline: [32250277](https://pubmed.ncbi.nlm.nih.gov/32250277/)]
41. Sutton S, French DP, Hennings SJ, Mitchell J, Wareham NJ, Griffin S, et al. Eliciting salient beliefs in research on the theory of planned behaviour: the effect of question wording. *Curr Psychol* 2003 Sep;22(3):234-251. [doi: [10.1007/s12144-003-1019-1](https://doi.org/10.1007/s12144-003-1019-1)]
42. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin J. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes* 2014 Dec 09;12:176 [FREE Full text] [doi: [10.1186/s12955-014-0176-2](https://doi.org/10.1186/s12955-014-0176-2)] [Medline: [25492701](https://pubmed.ncbi.nlm.nih.gov/25492701/)]
43. Kline RB. Principles and Practice of Structural Equation Modeling. 3rd edition. New York: Guilford publications; 2015:69-70.
44. Lu C, Hu Y, Fu Q, Governor S, Wang L, Li C, et al. Physician mental workload scale in China: development and psychometric evaluation. *BMJ Open* 2019 Oct 16;9(10):e030137. [doi: [10.1136/bmjopen-2019-030137](https://doi.org/10.1136/bmjopen-2019-030137)] [Medline: [31619422](https://pubmed.ncbi.nlm.nih.gov/31619422/)]
45. Wjx.cn. URL: <https://www.wjx.cn/>
46. Grimm L, Yarnold PR, editors. Reading and Understanding More Multivariate Statistics. Washington, DC: American Psychological Association; 2000:261-283.
47. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 6th edition. Boston: Pearson; 2013:720-726.
48. Orkibi H, Ronen T. Basic psychological needs satisfaction mediates the association between self-control skills and subjective well-being. *Front Psychol* 2017;8:936. [doi: [10.3389/fpsyg.2017.00936](https://doi.org/10.3389/fpsyg.2017.00936)] [Medline: [28638362](https://pubmed.ncbi.nlm.nih.gov/28638362/)]
49. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. *J Med Internet Res* 2016 Mar 02;18(3):e49 [FREE Full text] [doi: [10.2196/jmir.5069](https://doi.org/10.2196/jmir.5069)] [Medline: [26935646](https://pubmed.ncbi.nlm.nih.gov/26935646/)]
50. Koivumäki T, Pekkarinen S, Lappi M, Väisänen J, Juntunen J, Pikkarainen M. Consumer adoption of future MyData-based preventive eHealth services: an acceptance model and survey study. *J Med Internet Res* 2017 Dec 22;19(12):e429 [FREE Full text] [doi: [10.2196/jmir.7821](https://doi.org/10.2196/jmir.7821)] [Medline: [29273574](https://pubmed.ncbi.nlm.nih.gov/29273574/)]
51. Liu N, Tong Y, Chan HC. Information seeking in online healthcare communities: the dual influence from social self and personal self. *IEEE Trans Eng Manage* 2017 Nov;64(4):529-538. [doi: [10.1109/tem.2017.2703121](https://doi.org/10.1109/tem.2017.2703121)]

52. Clevenger L, Pyrzanowski J, Curtis CR, Bull S, Crane LA, Barrow JC, et al. Parents' acceptance of adolescent immunizations outside of the traditional medical home. *J Adolesc Health* 2011 Aug;49(2):133-140. [doi: [10.1016/j.jadohealth.2011.04.012](https://doi.org/10.1016/j.jadohealth.2011.04.012)] [Medline: [21783044](https://pubmed.ncbi.nlm.nih.gov/21783044/)]
53. Guo X, Zhang X, Sun Y. The privacy–personalization paradox in mHealth services acceptance of different age groups. *Electronic Commerce Res Applications* 2016 Mar;16:55-65. [doi: [10.1016/j.elerap.2015.11.001](https://doi.org/10.1016/j.elerap.2015.11.001)]
54. Yun EK, Park H. Consumers' disease information-seeking behaviour on the internet in Korea. *J Clin Nurs* 2010 Oct;19(19-20):2860-2868. [doi: [10.1111/j.1365-2702.2009.03187.x](https://doi.org/10.1111/j.1365-2702.2009.03187.x)] [Medline: [20598000](https://pubmed.ncbi.nlm.nih.gov/20598000/)]
55. Lee J. "Self" takes it all in mental illness: examining the dynamic role of health consciousness, negative emotions, and efficacy in information seeking. *Health Commun* 2019 Jul;34(8):848-858. [doi: [10.1080/10410236.2018.1437528](https://doi.org/10.1080/10410236.2018.1437528)] [Medline: [29461112](https://pubmed.ncbi.nlm.nih.gov/29461112/)]
56. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. *J Med Internet Res* 2014 May 09;16(5):e125 [FREE Full text] [doi: [10.2196/jmir.3283](https://doi.org/10.2196/jmir.3283)] [Medline: [24824062](https://pubmed.ncbi.nlm.nih.gov/24824062/)]
57. Xiong X, Zhang Z, Ren J, Zhang J, Pan X, Zhang L, et al. Impact of universal medical insurance system on the accessibility of medical service supply and affordability of patients in China. *PLoS One* 2018;13(3):e0193273 [FREE Full text] [doi: [10.1371/journal.pone.0193273](https://doi.org/10.1371/journal.pone.0193273)] [Medline: [29513712](https://pubmed.ncbi.nlm.nih.gov/29513712/)]

## Abbreviations

**AVE:** average variance extracted  
**CR:** composite reliability  
**EP:** emotional preference  
**GFI:** goodness-of-fit index  
**H:** hypothesis  
**mHealth:** mobile health  
**PBC:** perceived behavioral control  
**PC:** perceived convenience  
**PIR:** perceived information risk  
**PML:** perceived medical liability  
**PMR:** perceived medical risk  
**PO:** perceived outcome  
**PSD:** perceived severity of disease  
**SEM:** structural equation modeling  
**SN:** subjective norm  
**TPB:** theory of planned behavior

*Edited by G Eysenbach; submitted 21.07.20; peer-reviewed by Y He, S Song, J op den Buijs; comments to author 27.08.20; revised version received 15.09.20; accepted 29.09.20; published 29.10.20.*

### *Please cite as:*

Li D, Hu Y, Pfaff H, Wang L, Deng L, Lu C, Xia S, Cheng S, Zhu X, Wu X

*Determinants of Patients' Intention to Use the Online Inquiry Services Provided by Internet Hospitals: Empirical Evidence From China*

*J Med Internet Res* 2020;22(10):e22716

URL: <http://www.jmir.org/2020/10/e22716/>

doi: [10.2196/22716](https://doi.org/10.2196/22716)

PMID: [33006941](https://pubmed.ncbi.nlm.nih.gov/33006941/)

©Dehe Li, Yinhuan Hu, Holger Pfaff, Liuming Wang, Lu Deng, Chuntao Lu, Shixiao Xia, Siyu Cheng, Ximin Zhu, Xiaoyue Wu. Originally published in the *Journal of Medical Internet Research* (<http://www.jmir.org>), 29.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the *Journal of Medical Internet Research*, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Rapid Serological Assays and SARS-CoV-2 Real-Time Polymerase Chain Reaction Assays for the Detection of SARS-CoV-2: Comparative Study

Angelo Virgilio Paradiso<sup>1\*</sup>, MD; Simona De Summa<sup>2\*</sup>, PhD; Daniela Loconsole<sup>3</sup>, PhD; Vito Procacci<sup>4</sup>, PhD; Anna Sallustio<sup>5</sup>, PhD; Francesca Centrone<sup>3</sup>, PhD; Nicola Silvestris<sup>6,7</sup>, PhD; Vito Cafagna<sup>8</sup>, MSc; Giuseppe De Palma<sup>9</sup>, PhD; Antonio Tufaro<sup>9</sup>, MSc; Vito Michele Garrisi<sup>8\*</sup>, PhD; Maria Chironna<sup>3,4,5,10\*</sup>, PhD

<sup>1</sup>Science Direction, IRCCS Istituto Tumori Giovanni Paolo II, Bari, Italy

<sup>2</sup>Molecular Diagnostics and Pharmacogenetics Unit, IRCCS Istituto Tumori Giovanni Paolo II, Bari, Italy

<sup>3</sup>Department of Biomedical Sciences and Human Oncology-Hygiene Section, University of Bari, Bari, Italy

<sup>4</sup>Emergency Department, Policlinico Hospital, Bari, Italy

<sup>5</sup>Regional Epidemiological Observatory, Apulia Region, Bari, Italy

<sup>6</sup>Unit of Internal Medicine Guido Baccelli, Department of Biomedical Sciences and Human Oncology, University of Bari Medical School, Bari, Italy

<sup>7</sup>Medical Oncology Unit, IRCCS Istituto Tumori Giovanni Paolo II, Bari, Italy

<sup>8</sup>Clinical Pathology Laboratory, IRCCS Istituto Tumori Giovanni Paolo II, Bari, Italy

<sup>9</sup>Experimental Oncology and BioBank Management Unit, Institutional BioBank, IRCCS Istituto Tumori Giovanni Paolo II, Bari, Italy

<sup>10</sup>Hygiene Unit, Policlinico Hospital, Bari, Italy

\* these authors contributed equally

## Corresponding Author:

Angelo Virgilio Paradiso, MD

Science Direction

IRCCS Istituto Tumori Giovanni Paolo II

Viale Orazio Flacco 65

Bari, 70124

Italy

Phone: 39 805555900

Email: [a.paradiso@oncologico.bari.it](mailto:a.paradiso@oncologico.bari.it)

## Abstract

**Background:** Real-time polymerase chain reaction (RT-PCR) testing for the identification of viral nucleic acid is the current standard for the diagnosis of SARS-CoV-2 infection, but technical issues limit its utilization for large-scale screening. Serological immunoglobulin M (IgM)/IgG testing has been proposed as a useful tool for detecting SARS-CoV-2 exposure.

**Objective:** The objective of our study was to compare the results of the rapid serological VivaDiag test for SARS-CoV-2-related IgM/IgG detection with those of the standard RT-PCR laboratory test for identifying SARS-CoV-2 nucleic acid.

**Methods:** We simultaneously performed both serological and molecular tests with a consecutive series of 191 symptomatic patients. The results provided by a new rapid serological colorimetric test for analyzing IgM/IgG expression were compared with those of RT-PCR testing for SARS-CoV-2 detection.

**Results:** Of the 191 subjects, 70 (36.6%) tested positive for SARS-CoV-2 based on RT-PCR results, while 34 (17.3%) tested positive based on serological IgM/IgG expression. Additionally, 13 (6.8%) subjects tested positive based on serological test results, but also tested negative based on RT-PCR results. The rapid serological test had a sensitivity of 30% and a specificity of 89% compared to the standard RT-PCR assay. Interestingly, the performance of both assays improved 8 days after symptom appearance. After 10 days had passed since symptom appearance, the predictive value of the rapid serological test was higher than that of the standard molecular assay (proportion of positive results: 40% vs 20%). Multivariate analysis showed that age >58 years ( $P<.01$ ) and period of >15 days after symptom onset ( $P<.02$ ) were significant and independent factors associated with serological test positivity.

**Conclusions:** The rapid serological test analyzed in this study seems limited in terms of usefulness when diagnosing SARS-CoV-2 infection. However, it may be useful for providing relevant information on people's immunoreaction to COVID-19 exposure.



**KEYWORDS**

SARS-CoV-2; COVID-19; serological test; RT-PCR

## Introduction

Recently, a novel coronavirus, which was first reported in China, capable of interperson transmission has been causing lethal pneumonia in humans [1]. Subsequent molecular studies confirmed that the origin of this transmissible pneumonia was the novel SARS-CoV-2 virus, which causes the new COVID-19 disease [2].

As the COVID-19 disease rapidly spread to other Asian and European countries, the Italian Government had to take drastic measures to contain the outbreak, including establishing strict criteria to define patients from whom oropharyngeal swabs should be collected for the molecular polymerase chain reaction (PCR) diagnosis of COVID-19 and quarantining individuals who may have been in contact with SARS-CoV-2-infected people [3]. These measures were active for weeks, during which the number of new SARS-CoV-2 infection cases in Italy continued to increase, with more than 4000 new cases being reported daily [4]. Several attempts have been made to interpret the epidemiological trend of COVID-19 in Italy, and experts have focused on the limitations of early SARS-CoV-2 infection diagnosis [5] and the detection of SARS-CoV-2 infection in asymptomatic people [6].

The real-time PCR (RT-PCR) test for identifying viral nucleic acid is the current standard for the diagnosis of COVID-19. However, this assay has some practical limitations [3], such as the unpleasantness of obtaining biological material from the nasopharynx, the relatively long time required to generate results, and the need for certified laboratories and specific expertise. These limitations make RT-PCR unsuitable for quick and simple patient screening. Therefore, the search for a precise, rapid, simple, and large-scale screening test for quickly identifying SARS-CoV-2-infected patients has become urgent to prevent virus transmission and ensure timely treatment of patients.

The Saw Swee Hock School of Public Health at the National University of Singapore recently reviewed the diagnostic tests for COVID-19 infection currently undergoing clinical validation, including dozens of assays based on RT-PCR, next-generation sequencing, and microfluidics [7]. Additionally, 12 immunoassays based on evidence that COVID-19 is related to immunoglobulin G (IgG) and immunoglobulin M (IgM) expression were also listed. It has been argued that, based on previous experiences with viral SARS infection epidemics, specific IgM antibodies against SARS-CoV-2 could be detected in blood by performing immunoassays 3-6 days after symptom onset, while IgG detection could occur some days later [4]. It has also been speculated that, since SARS-CoV-2 belongs to the same large family of viruses that caused the Middle East Respiratory Syndrome and SARS epidemics, SARS-CoV-2 antibody seroconversion should be similar to that of other coronaviruses [5].

A report from the National University of Singapore has described the VivaDiag SARS-CoV-2 IgM/IgG Rapid Test kit as an immunoassay with available information regarding sensitivity and specificity [8] and a potential candidate for reliable and rapid (15 minutes) testing, according to the preliminary data available [6]. The test is based on the utilization of antihuman IgG and IgM against the recombinant antigen that represents the receptor-binding domain of the COVID-19 spike protein.

The aim of our study was to compare the results provided by the rapid serological VivaDiag test with those of standard RT-PCR testing for SARS-CoV-2 detection in swab specimens. The two tests were simultaneously performed on subjects with COVID-19 symptoms. We setup a prospective, mono-institutional, ad hoc, blinded, and independent study that enrolled a series of 191 subjects who were admitted to the Emergency Department of the Policlinico University Hospital in Bari, Italy.

## Methods

### Recruitment

Between March 23, 2020 and March 29, 2020, we enrolled a consecutive cohort of 191 patients who were admitted to the Emergency Department of the Policlinico University Hospital in Bari, Italy for COVID-19-related symptoms or because they were quarantined for previous exposure to COVID-19-positive individuals. Oropharyngeal swabs for standard SARS-CoV-2 RT-PCR analysis and venous blood samples for VivaDiag tests were simultaneously collected from each subject, and the tests were immediately performed in reference laboratories. Registries containing patients' main clinical data, including date of symptom onset (self-reported), were created. Informed written consent was obtained from all patients. Oropharyngeal swab samples were immediately analyzed for SARS-CoV-2 by RT-PCR at the Laboratory of Molecular Epidemiology and Public Health of the Hygiene Unit of the Policlinico University Hospital (Bari, Italy), the regional reference laboratory for SARS-CoV-2 identification. The venous blood samples were analyzed at the Clinical Pathology Laboratory (Certified ISO-9001/2015; Head E. Savino) and the Institutional BioBank (Certified ISO-9001/2015; Head A. Paradiso) of the IRCCS Istituto Tumori Giovanni Paolo II (Bari, Italy). This study was approved by the Ethical Committee of the IRCCS Istituto Tumori Giovanni Paolo II, Bari (Protocol number CE 870/2020).

### Molecular Detection of SARS-CoV-2

Nasopharyngeal/oropharyngeal swabs were subjected to nucleic acid extraction with the MagNA Pure System (Roche Diagnostics), in accordance with the manufacturer's instructions. The presence of the *E* gene, *RdRP* gene, and *N* gene of the SARS-CoV-2 virus were identified by a commercial RT-PCR assay (Allplex 2019-nCoV Assay; Seegene). Samples were considered positive at molecular screening if all three genes



were detected. The Centers for Disease Control and Prevention RT-PCR protocol was used to confirm the presence of SARS-CoV-2 [9]. To date [10,11], this methodology is considered the gold standard for the detection of SARS-CoV-2 infection.

### SARS-CoV-2 IgM/IgG Rapid Test

The SARS-CoV-2 IgM/IgG combined antibody rapid test kit, VivaDiag, (VivaChek Biotech) is a lateral flow qualitative immunoassay used for the rapid determination of the presence or absence of both anti-SARS-CoV-2-IgM and anti-SARS-CoV-2-IgG in human specimens (whole blood, serum, and plasma). A surface antigen from SARS-CoV-2, which can specifically bind to SARS-CoV-2 antibodies (including both IgM and IgG), is conjugated to colloidal gold nanoparticles and sprayed onto conjugation pads. The rapid SARS-CoV-2 IgG/IgM combined antibody test strip has two mouse antihuman monoclonal antibodies (anti-IgG and anti-IgM) on two separate test lines.

When testing, 10-15  $\mu$ L of a specimen was inserted into the sample port, and then the sample dilution buffer was added. As the specimen flowed through the device, anti-SARS-CoV-2 IgG and IgM antibodies, if present in the specimen, were bound by the SARS-CoV-2 antigens (ie, the gold colorimetric reagent fixed on the conjugate pad). As the conjugated sample continued to travel up the strip, the anti-SARS-CoV-2 IgM antibodies were bound on the M (IgM) line, and the anti-COVID-19 IgG antibodies were bound to the G (IgG) line. If the specimen did not contain SARS-CoV-2 antibodies, no labeled complexes were bound. The presence of SARS-CoV-2 IgG and IgM antibodies was indicated by a red/purple line on a specific region of the device. Each test was evaluated by two readers, and a picture was taken of the result. In case of disagreement, the picture was evaluated by a third party.

### Statistical Analysis

The performance of the VivaDiag tests was compared to that of the RT-PCR tests using the caret R package, which computed

all the parameters needed (sensitivity, specificity, accuracy, and Cohen  $\kappa$ ). The performance evaluation was carried out using RT-PCR as the gold standard. Both tests were performed on the same subjects. Univariate and multivariate logistic regression were performed. Age was dichotomized by using the median age as a cutoff, and the number of days after the onset of symptoms was used as a categorical variable (0-5 days, 6-8 days, 9-10 days, 11-15 days, >15 days). All analyses were carried out in R version 3.6 (The R Foundation), and results were considered to be significant when the  $P$  value was <.05.

## Results

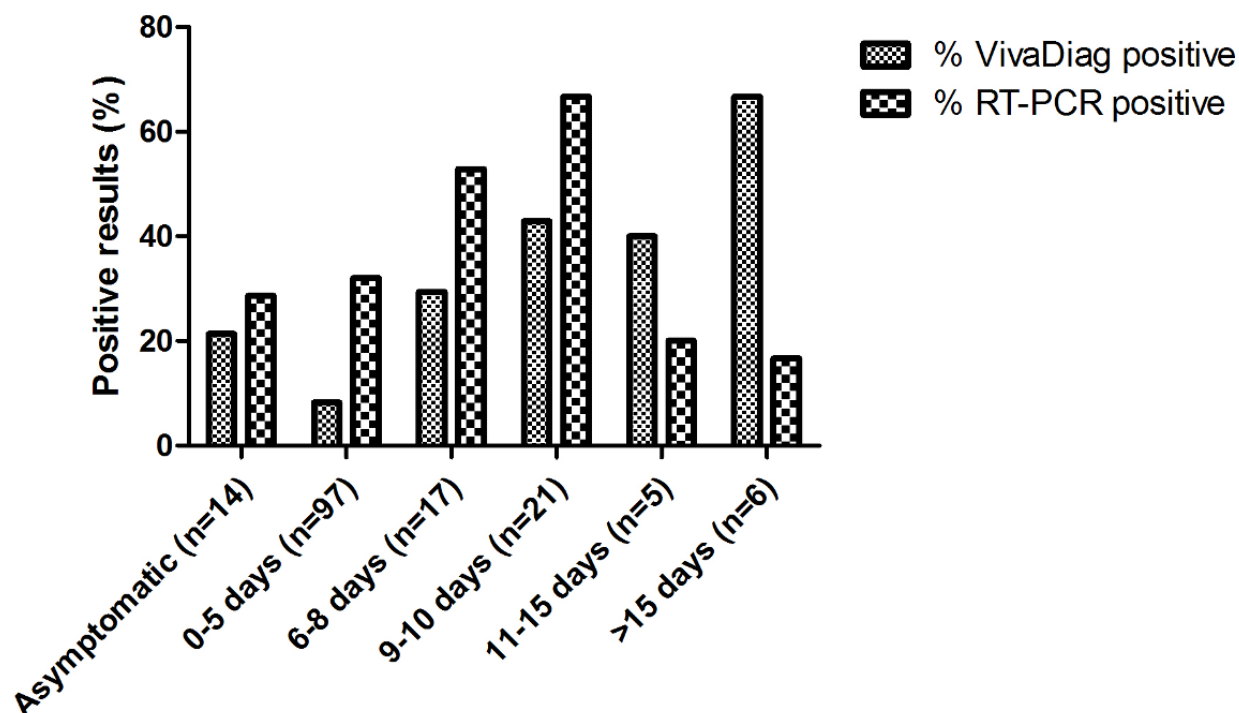
All 191 subjects enrolled in the study underwent a SARS-CoV-2 RT-PCR test and IgM/IgG rapid test. The cohort had a median age of 58.5 years, and 60.6% (116/191) were male. Subjects were admitted to the emergency room at different times after the onset of symptoms. A description of symptoms was available for 160 (83.8%) of the 191 patients. Of these 160 subjects, 14 (8.7%) were in quarantine and asymptomatic at the time they arrived to the emergency room.

Of the 191 patients, 70 (36.6%) tested positive for SARS-CoV-2 based on RT-PCR results, while 34 (17.8%) tested positive based on serum IgM/IgG rapid test results. Compared to the RT-PCR test, the serological test had an accuracy of 67% (95% CI 60-74), a sensitivity of 30%, and a specificity of 89%. The Cohen  $\kappa$  value was 0.21, meaning that the strength of agreement was, according to Altman [12], considered fair. Notably, 13 patients (6.8%) tested positive based on serological test results, but also tested negative based on RT-PCR results (Figure 1). Of these 13 subjects, 7 (54%) obtained positive IgG/IgM test results at different times after symptom onset (range 10-27 days), while 6 (46%) only obtained positive IgM results at various times after symptom appearance (range 4-25 days). The distribution of the percentage of positive results detected by both tests broken down by days from symptom onset is shown in Figure 2.

**Figure 1.** Comparison of RT-PCR and VivaDiag results from a series of 191 subjects ( $P=.001$ ). Compared to RT-PCR, VivaDiag had a sensitivity of 30%, specificity of 89%, accuracy of 67% (95% CI 60-74), and Cohen  $\kappa$  value of 0.21. RT-PCR: real-time polymerase chain reaction.

		RT-PCR	
		Negative, n (%)	Positive, n (%)
VivaDiag	Negative, n (%)	107 (56)	49 (29)
	Positive, n (%)	13 (7)	21 (11)

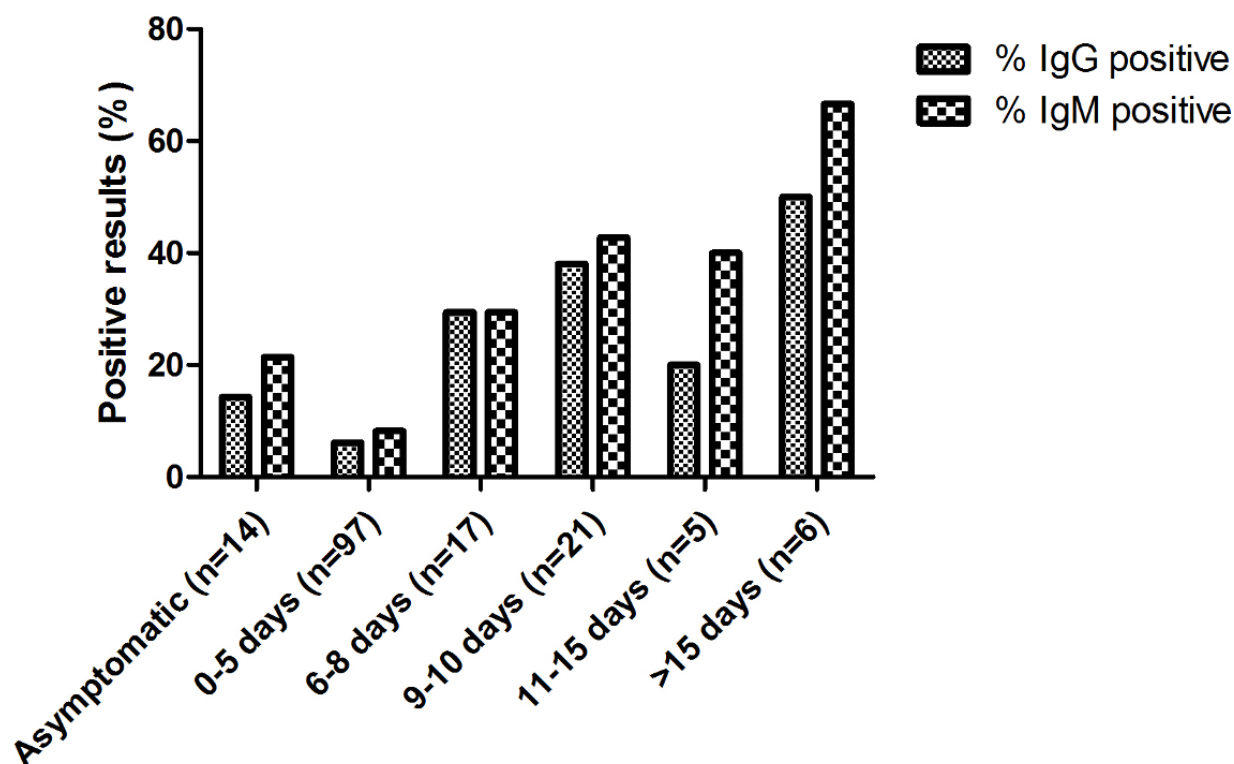
**Figure 2.** Bar plot depicting the distribution of the proportion of positive results from the VivaDiag serological test and those from RT-PCR testing for SARS-CoV-2 detection on oropharyngeal swab specimens according to time after symptom onset to test performance. RT-PCR: real-time polymerase chain reaction.



A clear increase in the number of positive serological tests was observed as more days elapsed from symptom appearance, reaching 66.7% at 15 days after symptom onset. Conversely, the highest likelihood for a positive RT-PCR test result was seen from 9 to 10 days after symptom onset, and it decreased rapidly afterwards over time. Of the 14 asymptomatic individuals, 4 (29%) had positive RT-PCR test results, while only 1 (7%) had a positive serological test result.

Further analysis regarding the behavior of IgM and IgG according to the time after symptom onset is described in [Figure 3](#). Only minimal differences in the behavior of the two immunoglobulins with respect to the time of symptom appearance became evident. However, all 13 patients with positive VivaDiag tests and negative RT-PCR results had positive IgM results, while only 7 (54%) of them also had positive IgG results.

**Figure 3.** Bar plot depicting the distribution of the proportion of positive IgG and IgM VivaDiag test results according to time after symptom onset to test performance. IgG: immunoglobulin G; IgM: immunoglobulin M.



Univariate and multivariate logistic regressions were performed to identify independent predictive variables for positive VivaDiag and RT-PCR test results (Tables 1-2). Both univariate and multivariate analyses showed that age >58.5 years and

period of >15 days from symptom onset were significantly associated with VivaDiag positivity, while 9-10 days after symptom onset was independently associated with a positive RT-PCR test result.

**Table 1.** Univariate and multivariate logistic regression results using VivaDiag positivity as a dependent variable.

Variable	OR (95% CI)	P value
<b>Univariate logistic regression analysis</b>		
<b>Days after symptom onset</b>		
Asymptomatic	Ref	Ref
0-5	0.32 (0.08-1.66)	.13
6-8	1.52 (0.3-8.9)	.61
9-10	2.74 (0.63-14.88)	.19
11-15	2.44 (0.23-23.39)	.42
>15	7.33 (0.96-77.28)	.06
<b>Age (years)</b>		
≤58.5	Ref	Ref
>58.5	2.99 (1.31-7.31)	.01
<b>Sex</b>		
Female	Ref	Ref
Male	1.22 (0.55-2.85)	.62
<b>Multivariate logistic regression analysis</b>		
<b>Days after symptom onset</b>		
>15	12.3 (1.44-148.14)	.02
<b>Age (years)</b>		
≤58.5	Ref	Ref
>58.5	3.59 (1.39-10.48)	.01

**Table 2.** Univariate and multivariate logistic regression results using real-time polymerase chain reaction SARS-CoV-2 positivity as a dependent variable.

Variable	OR (95% CI)	P value
<b>Univariate logistic regression analysis</b>		
<b>Days after symptom onset</b>		
Asymptomatic	Ref	Ref
0-5	1.17 (0.36-4.54)	.79
6-8	2.81(0.65-13.75)	.17
9-10	4.99 (1.21-24.11)	.03
11-15	0.62 (0.02-6.17)	.71
>15	0.53 (0.02-4.64)	.57
<b>Age (years)</b>		
≤58.5	Ref	Ref
>58.5	0.89 (0.47-1.7)	.74
<b>Sex</b>		
Female	Ref	Ref
Male	1.2 (0.62-2.33)	.58
<b>Multivariate logistic regression analysis</b>		
<b>Days after symptom onset</b>		
9-10	4.96 (1.2-24)	.03

## Discussion

### Principal Results

When we compared the performance of the rapid serological test to that of RT-PCR for the detection of SARS-CoV-2 infection, our findings showed that 17.8% (34/191) of the subjects tested positive based on serum IgM/IgG expression, whereas 36.6% (70/191) of the subjects tested positive based on SARS-CoV-2 RT-PCR test results, leading to a sensitivity of 30% and a specificity of 89% for the serological test.

The clinical relevance of so-called rapid serological testing is still an open issue, since the data currently available are still scarce [13]. For this reason, we compared its performance to that of standard RT-PCR testing and analyzed performance with respect to the time of COVID-19–related symptom onset. To this end, we set up a mono-institutional consecutive cohort of patients who were tested with both assays at a single qualified laboratory.

The design of our study allowed us to specifically analyze two aspects of the open issue: (1) the concordance of the rapid serological test results with those of standard molecular testing and (2) the relationship between IgG/IgM expression and the onset of clinical symptoms.

With regard to the degree of concordance between the two tests, the results reported in Figure 1 clearly show that the precision of the VivaDiag rapid test is unsatisfactory. Notably, only 11% (21/191) of the patients that tested positive for COVID-19 based on the molecular test results also tested positive based on the

serological test results. This percentage is impressively similar to the performance reported for serological tests in Spain [14] and Germany [15]. However, the first important finding from our study concerns the 6.8% (13/191) of subjects that tested negative based on RT-PCR results, but tested positive based on serological results. The two tests did not produce similar results, which is obvious for assays that are designed to analyze different aspects of COVID-19; the molecular test detects the presence of SARS-CoV-2 in samples based on specific anatomical parts of the respiratory system, while the serology test reveals the kinetics of immunoglobulins as the body reacts to viral infection. Negative serological test results in patients with a positive molecular test could mean that the patients are infected, but have not yet reached the stage of immunoglobulin reaction development. Conversely, subjects who have negative molecular test results, but also have serological test results showing the presence of specific IgG/IgM antibodies, may be recovering from COVID-19. The data shown in Figure 2 seem to confirm these assumptions, as the molecular test yielded more positive results during the early symptomatic phases of the disease in our subjects, while the serological test performed better later on (ie, 10 days after symptom appearance).

The second aspect we were able to analyze in our study was seroconversion and the kinetics of immunoglobulins with respect to the onset of COVID-19–related symptoms. Figure 3 shows the behavior of the two immunoglobulins according to symptom appearance. Interestingly, IgG and IgM did not seem to behave differently based on the number of days elapsing from symptom appearance, but they clearly and progressively increased along the course of the disease. This unexpected finding, in contrast



with common knowledge concerning the kinetics of the two immunoglobulins, is supported by the results presented by Zhang et al [16] and Lou et al [17]. Both authors reported that the detectable serology markers, IgG and IgM, had similar seroconversions in COVID-19 patients, with antibody levels increasing rapidly at 6 days after exposure, and this trend occurred with a concomitant decline in viral load. Such behavior in the 6-10-day time window after symptom appearance is typically accompanied by an improvement in serological test sensitivity compared to standard molecular testing.

Very recently, Whitman et al [18] evaluated the performance of 11 SARS-CoV-2 serological assays in a multicentric study that enrolled a cohort of subjects with positive RT-PCR test results. For the serological tests, the incidence of IgG/IgM positivity in the RT-PCR SARS-CoV-2-positive samples ranged from 26.9% to 0%. In particular, the incidence of positive results with the VivaDiag test was 8.2%, which is significantly lower than the value we found in our study. However, the enrollment criteria in the Whitman et al study were very different from ours. In our study, the subjects were consecutively recruited from a single institution, and the tests were performed simultaneously on fresh biospecimens. Interestingly, Whitman et al reported that among the SARS-CoV-2 RT-PCR-positive individuals, the percent seropositivity increased with time, peaking at 81.8%-100.0% in samples taken >20 days after symptom onset. The same trend of increased seropositivity with increased time from symptom onset was observed in our cohort of subjects.

### Limitations

Our study had some important limitations. First, the VivaDiag test was based on the colorimetric evaluation of the IgG and IgM bands performed by the operator, thus implying that all the limitations that a qualitative inter-intra-operator evaluation

produces in terms of variability are present in this study [19]. In our study, this was partially solved by resorting to double operator evaluation and taking pictures of all test results to be reanalyzed by a third-party reader in the case of first level evaluation disagreement. Therefore, our next step will be to use quantitative immunoassay methods to analyze SARS-CoV-2-specific immunoglobulins [20] to overcome these issues.

Second, the neutralizing antibodies used in the VivaDiag test might cross-react with other coronavirus antigens, such as those of the SARS-CoV. The recombinant antigen utilized in VivaDiag is the receptor-binding domain of the SARS-CoV-2 spike protein, for which information on possible cross-reactivity with other coronaviruses and flu viruses has not yet been studied [13]. Further studies are urgently needed to definitively clarify this point.

### Conclusions

Our study analyzed the clinical performance of the rapid serological test, VivaDiag, and confirmed the test's limited applicability for the diagnosis of SARS-CoV-2 infection by comparing its performance to that of standard molecular testing. However, this rapid serological test seems to provide important information concerning individuals' immunoreaction to the infection, and more importantly, it may detect previous exposure to the virus in currently healthy persons. A trial, recently registered in ClinicalTrials.gov (NCT04316728), will specifically address this issue by investigating the monitoring of seroconversion of COVID-19 IgG/IgM in healthy subjects who may develop COVID-19-related symptoms. In essence, our real-world results should be considered hypothesis-generating findings that warrant further examination in a controlled clinical trial in order to be confirmed [21].

### Acknowledgments

The authors wish to thank Clinical Pathology Laboratory Director, Dr. E Savino, and his coworkers for their cooperation.

### Conflicts of Interest

None declared.

### References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)] [Medline: [31986264](https://pubmed.ncbi.nlm.nih.gov/31986264/)]
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: [10.1056/NEJMoa2001017](https://doi.org/10.1056/NEJMoa2001017)] [Medline: [31978945](https://pubmed.ncbi.nlm.nih.gov/31978945/)]
3. Livingston E, Bucher K. Coronavirus Disease 2019 (COVID-19) in Italy. *JAMA* 2020 Apr 14;323(14):1335. [doi: [10.1001/jama.2020.4344](https://doi.org/10.1001/jama.2020.4344)] [Medline: [32181795](https://pubmed.ncbi.nlm.nih.gov/32181795/)]
4. Italy Coronavirus. Worldometer - real time world statistics. URL: <https://www.worldometers.info/coronavirus/country/italy/> [accessed 2020-10-20]
5. Jin YH, Cai L, Cheng ZS, Cheng H, Deng T, Fan YP, for the Zhongnan Hospital of Wuhan University Novel Coronavirus Management and Research Team, Evidence-Based Medicine Chapter of China International Exchange and Promotive Association for Medical and Health Care (CPAM). A rapid advice guideline for the diagnosis and treatment of 2019 novel

- coronavirus (2019-nCoV) infected pneumonia (standard version). *Mil Med Res* 2020 Feb 06;7(1):4 [FREE Full text] [doi: [10.1186/s40779-020-0233-6](https://doi.org/10.1186/s40779-020-0233-6)] [Medline: [32029004](https://pubmed.ncbi.nlm.nih.gov/32029004/)]
6. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *N Engl J Med* 2020 Mar 05;382(10):970-971 [FREE Full text] [doi: [10.1056/NEJMc2001468](https://doi.org/10.1056/NEJMc2001468)] [Medline: [32003551](https://pubmed.ncbi.nlm.nih.gov/32003551/)]
  7. Yap JCH, Ang IYH, Tan SHX, Chen JIP, Lewis RF, Yang Q, et al. COVID-19 Science Report: Diagnostics. National University of Singapore. 2020 Mar 19. URL: <https://cpb-ap-se2.wpmucdn.com/blogs.auckland.ac.nz/dist/f/688/files/2020/03/COVID-19-Testing-Landscape-Appendices.pdf> [accessed 2020-10-20]
  8. VivaDiag TM SARS-CoV-2 IgM/IgG Rapid Test (COVID-19 IgM/IgG Rapid Test). VivaChek. URL: <https://www.vivachek.com/en/prods/prod-rapidtest.html> [accessed 2020-10-20]
  9. Centers for Disease Control and Prevention. Real-time RT-PCR Panel for detection 2019-novel coronavirus : instructions for use. CDC Stacks Public Health Publications. 2020 Jan 20. URL: <https://stacks.cdc.gov/view/cdc/84526> [accessed 2020-10-20]
  10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977 Mar;33(1):159-174. [Medline: [843571](https://pubmed.ncbi.nlm.nih.gov/843571/)]
  11. Binnicker M. Emergence of a Novel Coronavirus Disease (COVID-19) and the Importance of Diagnostic Testing: Why Partnership between Clinical Laboratories, Public Health Agencies, and Industry Is Essential to Control the Outbreak. *Clin Chem* 2020 May 01;66(5):664-666 [FREE Full text] [doi: [10.1093/clinchem/hvaa071](https://doi.org/10.1093/clinchem/hvaa071)] [Medline: [32077933](https://pubmed.ncbi.nlm.nih.gov/32077933/)]
  12. Altman DG. Practical Statistics for Medical Research. London: CRC Press; 1990.
  13. Li Z, Yi Y, Luo X, Xiong N, Liu Y, Li S, et al. Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J Med Virol* 2020 Feb 27;10.1002/jmv.25727 [FREE Full text] [doi: [10.1002/jmv.25727](https://doi.org/10.1002/jmv.25727)] [Medline: [32104917](https://pubmed.ncbi.nlm.nih.gov/32104917/)]
  14. Elena S. Unreliability of new tests delays effort to slow coronavirus spread in Spain. *El Pais*. 2020 Mar 27. URL: <https://english.elpais.com/society/2020-03-27/unreliability-of-new-tests-delays-effort-to-slow-coronavirus-spread-in-spain.html> [accessed 2020-10-20]
  15. Schmitt PP. Franfurter Allgemeine. Wir haben neue Symptome entdeckt. 2020 Mar 16. URL: <https://www.faz.net/aktuell/gesellschaft/gesundheitscoronavirus/neue-corona-symptome-entdeckt-virologe-hendrik-streeck-zum-virus-16681450.html> [accessed 2020-10-27]
  16. Zhang J, Liu J, Li N, Liu Y, Ye R, Qin X, et al. Serological detection of 2019-nCoV respond to the epidemic: A useful complement to nucleic acid testing. *medRxiv Preprint* posted online on March 10, 2020. [doi: [10.1101/2020.03.04.20030916](https://doi.org/10.1101/2020.03.04.20030916)]
  17. Lou B, Li TD, Zheng SF, Su YY, Li ZY, Liu W, et al. Serology characteristics of SARS-CoV-2 infection after exposure and post-symptom onset. *Eur Respir J* 2020 Aug;56(2):2000763. [doi: [10.1183/13993003.00763-2020](https://doi.org/10.1183/13993003.00763-2020)] [Medline: [32430429](https://pubmed.ncbi.nlm.nih.gov/32430429/)]
  18. Whitman JD, Hiatt J, Mowery CT, Shy BR, Yu R, Yamamoto TN, et al. Test performance evaluation of SARS-CoV-2 serological assays. *medRxiv Preprint* posted online on May 17, 2020. [FREE Full text] [doi: [10.1101/2020.04.25.20074856](https://doi.org/10.1101/2020.04.25.20074856)] [Medline: [32511497](https://pubmed.ncbi.nlm.nih.gov/32511497/)]
  19. Paradiso A, Ellis IO, Zito FA, Marubini E, Pizzamiglio S, Verderio P. Short- and long-term effects of a training session on pathologists' performance: the INQAT experience for histological grading in breast cancer. *J Clin Pathol* 2009 Mar;62(3):279-281. [doi: [10.1136/jcp.2008.061036](https://doi.org/10.1136/jcp.2008.061036)] [Medline: [19251956](https://pubmed.ncbi.nlm.nih.gov/19251956/)]
  20. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling Early Humoral Response to Diagnose Novel Coronavirus Disease (COVID-19). *Clin Infect Dis* 2020 Jul 28;71(15):778-785 [FREE Full text] [doi: [10.1093/cid/ciaa310](https://doi.org/10.1093/cid/ciaa310)] [Medline: [32198501](https://pubmed.ncbi.nlm.nih.gov/32198501/)]
  21. Camm AJ, Fox KAA. Strengths and weaknesses of 'real-world' studies involving non-vitamin K antagonist oral anticoagulants. *Open Heart* 2018;5(1):e000788. [doi: [10.1136/openhrt-2018-000788](https://doi.org/10.1136/openhrt-2018-000788)] [Medline: [29713485](https://pubmed.ncbi.nlm.nih.gov/29713485/)]

## Abbreviations

**IgG:** immunoglobulin G

**IgM:** immunoglobulin M

**PCR:** polymerase chain reaction

**RT-PCR:** real-time polymerase chain reaction

*Edited by G Eysenbach; submitted 17.04.20; peer-reviewed by M Salman, T Rana, D Laborde; comments to author 02.06.20; revised version received 31.07.20; accepted 13.09.20; published 30.10.20.*

*Please cite as:*

*Paradiso AV, De Summa S, Loconsole D, Procacci V, Sallustio A, Centrone F, Silvestris N, Cafagna V, De Palma G, Tufaro A, Garrisi VM, Chironna M*

*Rapid Serological Assays and SARS-CoV-2 Real-Time Polymerase Chain Reaction Assays for the Detection of SARS-CoV-2: Comparative Study*

*J Med Internet Res 2020;22(10):e19152*

*URL: <http://www.jmir.org/2020/10/e19152/>*

*doi: [10.2196/19152](https://doi.org/10.2196/19152)*

*PMID: [33031048](https://pubmed.ncbi.nlm.nih.gov/33031048/)*

©Angelo Virgilio Paradiso, Simona De Summa, Daniela Loconsole, Vito Procacci, Anna Sallustio, Francesca Centrone, Nicola Silvestris, Vito Cafagna, Giuseppe De Palma, Antonio Tufaro, Vito Michele Garrisi, Maria Chironna. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

## Original Paper

# Depression, Anxiety, and Lifestyle Among Essential Workers: A Web Survey From Brazil and Spain During the COVID-19 Pandemic

Raquel Brandini De Boni<sup>1</sup>, MD, MSc, PhD; Vicent Balanzá-Martínez<sup>2</sup>, MD, PhD; Jurema Correa Mota<sup>1</sup>, PhD; Taiane De Azevedo Cardoso<sup>3</sup>, MSc, PhD; Pedro Ballester<sup>4</sup>, MSc; Beatriz Atienza-Carbonell<sup>5</sup>, MD; Francisco I Bastos<sup>1</sup>, MD, MSc, PhD; Flavio Kapczinski<sup>3,6,7</sup>, MD, MSc, PhD, FRCPC

<sup>1</sup>Institute of Scientific and Technological Communication and Information in Health, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

<sup>2</sup>Department of Medicine, University of Valencia, CIBERSAM, Valencia, Spain

<sup>3</sup>Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada

<sup>4</sup>Neuroscience Graduate Program, McMaster University, Hamilton, ON, Canada

<sup>5</sup>School of Medicine, University of Valencia, Valencia, Spain

<sup>6</sup>Bipolar Disorder Program, Laboratory of Molecular Psychiatry, Hospital de Clínicas de Porto Alegre, Instituto Nacional de Ciência e Tecnologia Translacional em Medicina, Porto Alegre, Brazil

<sup>7</sup>Department of Psychiatry, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

**Corresponding Author:**

Raquel Brandini De Boni, MD, MSc, PhD

Institute of Scientific and Technological Communication and Information in Health

Oswaldo Cruz Foundation

Av Brasil 4365 Pavilhao Haity Moussatche, Room 229

Rio de Janeiro, 21040-360

Brazil

Phone: 55 21 3865 3231

Email: [raqueldeboni@gmail.com](mailto:raqueldeboni@gmail.com)

## Abstract

**Background:** Essential workers have been shown to present a higher prevalence of positive screenings for anxiety and depression during the COVID-19 pandemic. Individuals from countries with socioeconomic inequalities may be at increased risk for mental health disorders.

**Objective:** We aimed to assess the prevalence and predictors of depression, anxiety, and their comorbidity among essential workers in Brazil and Spain during the COVID-19 pandemic.

**Methods:** A web survey was conducted between April and May 2020 in both countries. The main outcome was a positive screening for depression only, anxiety only, or both. Lifestyle was measured using a lifestyle multidimensional scale adapted for the COVID-19 pandemic (Short Multidimensional Inventory Lifestyle Evaluation–Confinement). A multinomial logistic regression model was performed to evaluate the factors associated with depression, anxiety, and the presence of both conditions.

**Results:** From the 22,786 individuals included in the web survey, 3745 self-reported to be essential workers. Overall, 8.3% (n=311), 11.6% (n=434), and 27.4% (n=1027) presented positive screenings for depression, anxiety, and both, respectively. After adjusting for confounding factors, the multinomial model showed that an unhealthy lifestyle increased the likelihood of depression (adjusted odds ratio [AOR] 4.00, 95% CI 2.72-5.87), anxiety (AOR 2.39, 95% CI 1.80-3.20), and both anxiety and depression (AOR 8.30, 95% CI 5.90-11.7). Living in Brazil was associated with increased odds of depression (AOR 2.89, 95% CI 2.07-4.06), anxiety (AOR 2.81, 95% CI 2.11-3.74), and both conditions (AOR 5.99, 95% CI 4.53-7.91).

**Conclusions:** Interventions addressing lifestyle may be useful in dealing with symptoms of common mental disorders during the strain imposed among essential workers by the COVID-19 pandemic. Essential workers who live in middle-income countries with higher rates of inequality may face additional challenges. Ensuring equitable treatment and support may be an important challenge ahead, considering the possible syndemic effect of the social determinants of health.

(*J Med Internet Res* 2020;22(10):e22835) doi:[10.2196/22835](https://doi.org/10.2196/22835)

**KEYWORDS**

COVID-19; depression; anxiety; lifestyle; Brazil; Spain

**Introduction**

The prevalence of COVID-19 continues to increase in Brazil [1], and mental health is recognized as an important challenge ahead worldwide [2-4]. Several studies have used web surveys to screen for common mental health disorders (mainly depressive and anxiety disorders) among the general population [5-8]. Some of these studies, but not all [9,10], have shown that essential workers, such as health care workers (HCWs), had an increased likelihood of anxiety and depression compared to other workers [11,12]. Essential workers may have increased workload and working hours during the pandemic [13], struggle with the lack of adequate personal protective equipment [14], and may be isolated from friends and families [15]. Burnout symptoms [16], emotional exhaustion [17], and fear of transmitting the virus [18] are commonly reported.

The first study investigating mental health problems among essential workers during the COVID-19 pandemic was conducted in China and found the prevalence of anxiety and depressive symptoms to be at 20.1% and 12.7%, respectively [19]. In two subsequent systematic reviews, HCWs presented increased depression/depressive symptoms, anxiety, psychological distress, and poor sleep quality [20,21]. The first meta-analysis evaluating the prevalence of mental health problems among HCWs (up to April 13, 2020) found 13 studies (N=33,062 participants), of which 12 were conducted in China. The pooled prevalence of a positive screening for anxiety was 23.2% (95% CI 17.8-29.1) while depression was estimated at 22.8% (95% CI 15.1-31.5) [22]. Evidence from outside of China is still scarce [23-30], and it is difficult to estimate the overall prevalence of common mental health disorders due to methodological issues, heterogeneity in study populations and sizes, and differences in criteria used to define a case (eg, different instruments and cut-offs). For instance, the prevalence of depression was as low as 10% among HCWs in Singapore and India [23] to as high as 64.7% among physicians in Turkey [31].

Four major groups of variables have been associated with an increased likelihood of having a positive screening for anxiety and/or depression among essential workers during the COVID-19 pandemic: demographic, professional/financial worries, COVID-19 exposure factors, and personal health factors. In terms of demographics, being female and younger were more frequently associated with depression [31,32] while a higher education level and residing in areas or provinces with a greater number of cases have been associated with anxiety [19,33,34]. Worrying about adequate training, knowledge, preparedness, and finances, as well as self-efficacy and career phase, were some of the professional/financial variables analyzed under this domain [34]. COVID-19 exposure variables

included being a frontline worker [31,33,34]; fear, suspicion, or diagnosis of COVID-19 for oneself and/or their significant other [19,32,33,35]; and having a deceased colleague [32]. Finally, regarding personal health factors, included variables were perceived stress [19], poor sleep [19,26], presence of a previous medical or psychiatric disorder [31,35], and a history of alcohol consumption [35]. Of note, good social support was frequently associated with a decreased likelihood of anxiety and depression among essential workers [19,36].

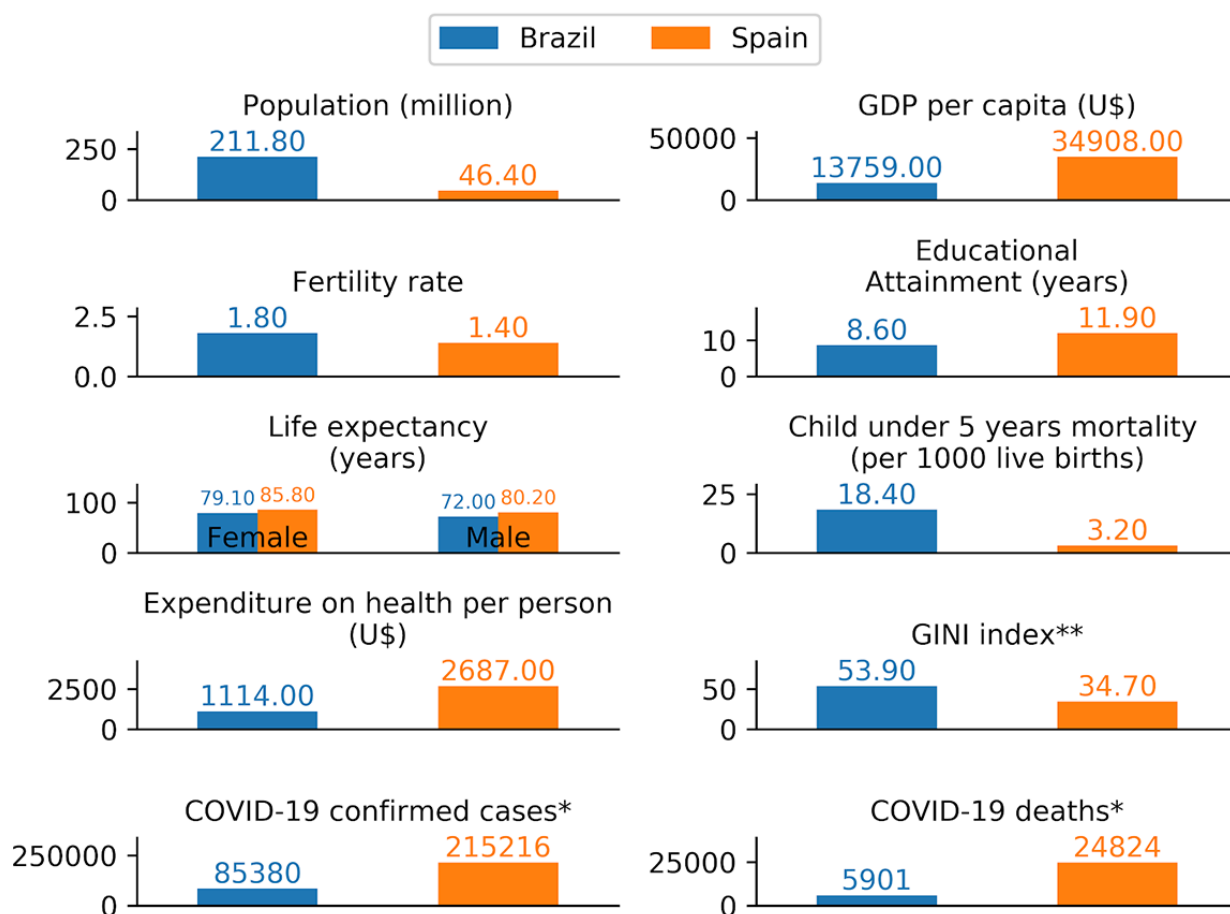
The fast growing field of lifestyle psychiatry [37] has been providing evidence on how health behaviors (eg, diet, physical activity, smoking, and sleep) relate to the prevalence, incidence, and adverse outcomes of mental health disorders [38-40]. For instance, consistent results point to the beneficial effects of physical activity in preventing the onset of depression as well as improving its symptoms [40]. As many people around the world are under confinement due to the COVID-19 pandemic, changes in lifestyle behaviors have attracted more research interest [41]. Behaviors, such as sleep quality [36,42], were assessed among HCWs instead of using a comprehensive, multidimensional approach to lifestyle. Multidimensional evaluations of lifestyle are still scarce [41], although it is possible that different health behaviors share a common pathway to improve mental health, such as anti-inflammatory effects [43,44]).

Considering the social determinants of health [45,46], it is possible that countries presenting poor social and health indicators may present a higher prevalence of unhealthy outcomes, which could include both COVID-19-related and mental health problems. Herein, we have included two countries with different social and health indicators, and at different stages of the COVID-19 pandemic, as summarized in Figure 1 [47-50]. Brazil has roughly 4.5 times the population of Spain, but 0.4 times the GDP (gross domestic product) and 0.4 times the expenditure on health. In addition, Brazil is considered one of the countries with the greatest inequalities in income/wealth globally, with a Gini index of 53.9. The first COVID-19 case was diagnosed on January 31, 2020, in Spain and on February 26, 2020, in Brazil, respectively. On May 2, 2020 (the midpoint of our data collection period), there were almost 25,000 deaths in Spain and 6000 deaths in Brazil. At that time, Spain was under a strict lockdown policy while the lockdown in Brazil was implemented partially and in select cities and counties.

So far, we are unaware of studies that have investigated mental health problems among essential workers from two countries presenting such different social and epidemic profiles. Thus, our major aim is to describe the prevalence of depression, anxiety, and the comorbidity of both, as well as their associated factors, among self-reported essential workers during the COVID-19 pandemic in Brazil and Spain.



**Figure 1.** Select social and health indicators (2017) and the COVID-19 situation (as of May 2, 2020) for Brazil and Spain. GDP: gross domestic product. Data sources: Institute from Health Metrics and Evaluation [47], \*World Health Organization [48,49], \*\*World Bank 2017-2018 [50].



## Methods

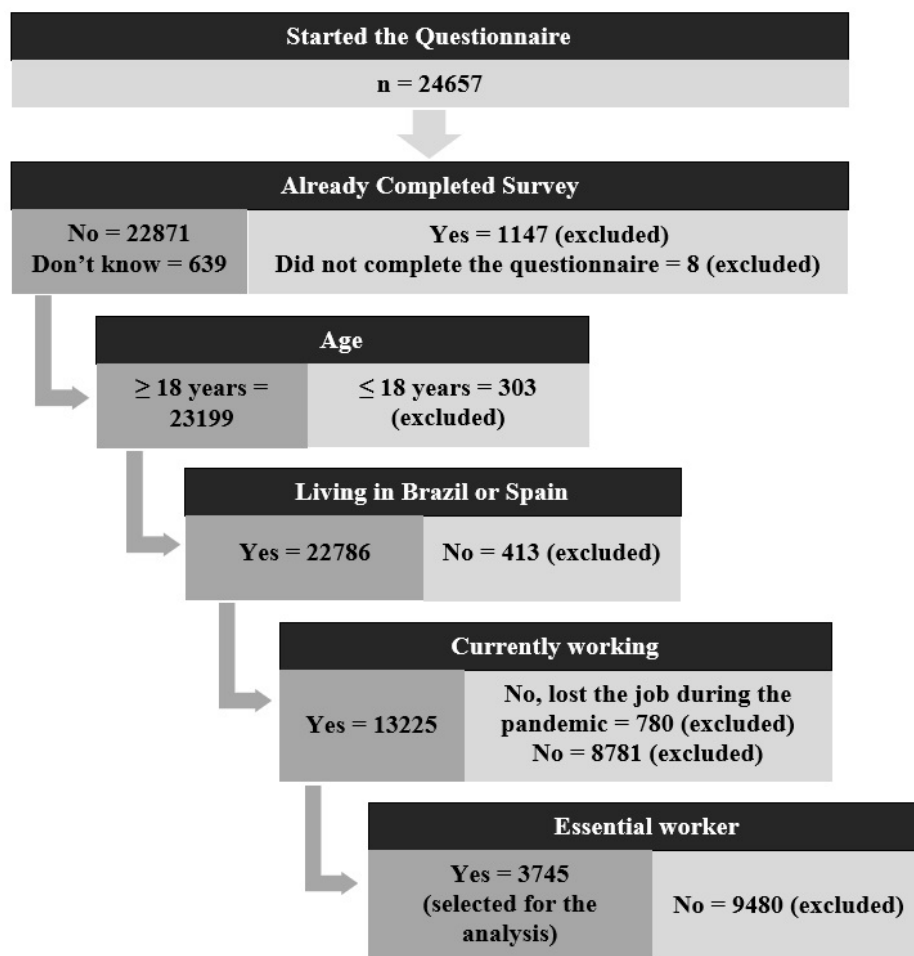
### Study Design

A web survey was conducted from April 15 to May 15, 2020, in Spain and April 20 to May 20, 2020, in Brazil. The online questionnaire was created using SurveyGizmo and included questions about demographics, COVID-19 experience, lifestyle behaviors, self-rated health, and previous diagnosed conditions. The questionnaire comprised 101 questions, and skips, when appropriate, were implemented to decrease the time of completion (Multimedia Appendix 1). The usability and technical functionality were tested before launching the survey in both countries. In addition, participants could read information regarding ways to maintain a healthy lifestyle during the pandemic while they were answering the questionnaire, and were provided with additional websites and telephone numbers to find reliable information regarding COVID-19. This information was compiled from the COVID-19-related

webpages of the Centers for Disease Control and Prevention, the National Institute of Health (United States), the Oswaldo Cruz Foundation, the Brazilian Ministry of Health, and the Spanish Ministry of Health.

### Study Population

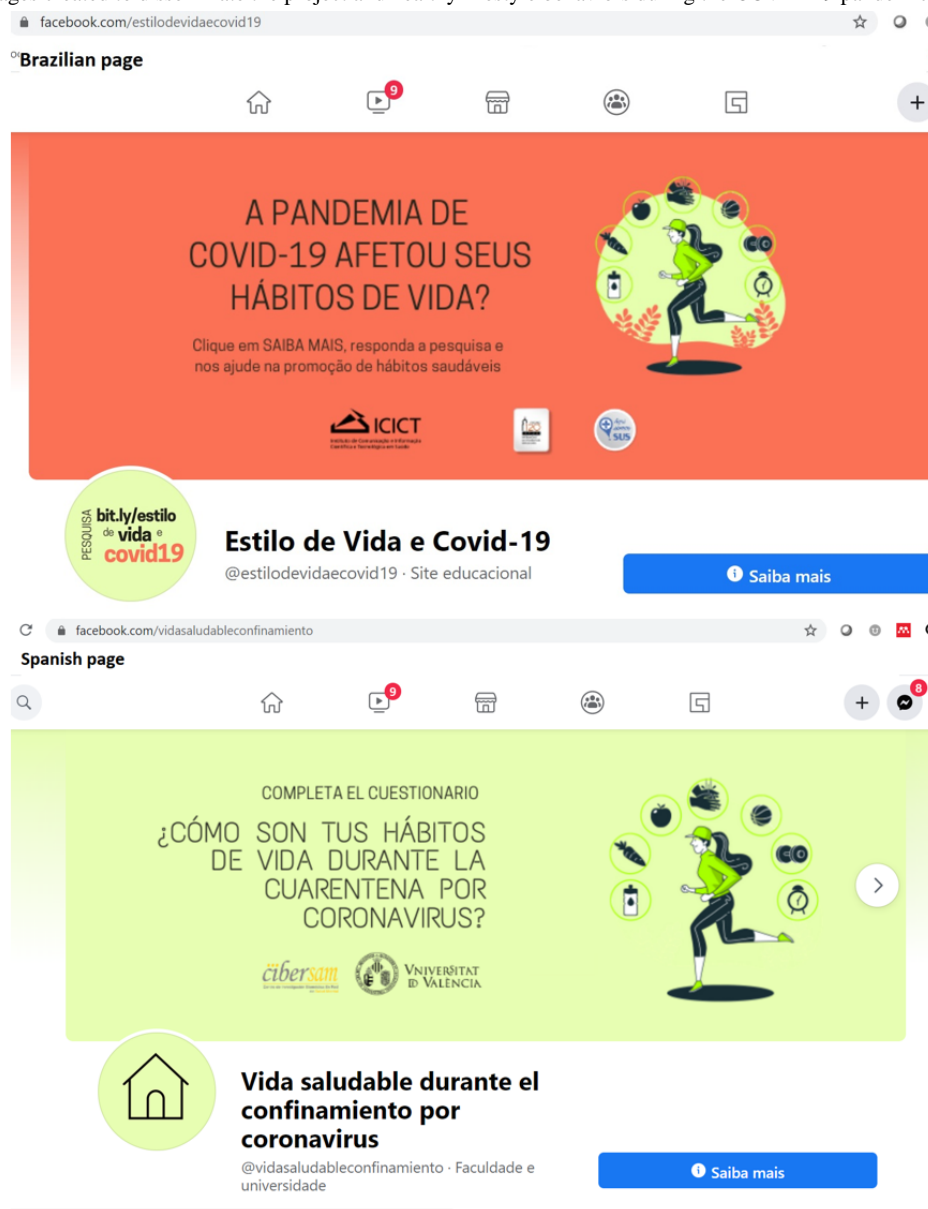
The study population included adults from both sexes living in Spain or in Brazil, who had access to the internet, and agreed to participate in the study after reading the informed consent form. Herein, we selected individuals who reported to be essential workers. Essential workers were considered all participants who answered “yes” to the following question: “Are you currently working as a health care worker or as a professional of other essential services (transportation, food, cleaning)?” To avoid duplicated responses, individuals who reported having previously completed the questionnaire were excluded (Figure 2) as no identification data (nor IP [Internet Protocol] address) were collected.

**Figure 2.** Inclusion flowchart.

### Sample and Recruitment

A convenience sample of participants was recruited via social networks (Facebook, WhatsApp, and Twitter) using a snowball technique and sponsored social network advertisements. A Facebook page for each country (Figure 3) was created and boosted using the following words: “healthy lifestyles,” “sad,” “happiness,” “fear,” “emotion,” “stress,” “well-being,” “self-esteem,” “quality of life,” “motivation,” “mind,” “boredom,” “panic,” “interpersonal relationship,” “life,”

“emotional intelligence,” “physical fitness,” and “physical exercise.” According to Facebook, the page would reach 87 million people in Brazil and 290,000 individuals in Spain. Fundamental parameters were unknown when the sample size was defined, such as the possible participation of superrecruiters [51] with the potential to skew the sampling process, or whether or not there would be structural bottlenecks. Therefore, the sample size was not defined a priori; instead, a 30-day period of data collection was specified.

**Figure 3.** Facebook pages created to disseminate the project and healthy lifestyle behaviors during the COVID-19 pandemic in Brazil and Spain.

## Response Rates

Response rates were not estimated in the study since the study denominator is unknown (ie, we were unable to estimate how many individuals had sent the survey link via different—and probably overlapping—social networks).

## Variables and Measurements

### Study Outcome (Dependent Variable)

The main outcome is a positive screening for depression and/or anxiety. Depression was screened using the Patient Health Questionnaire-2 (PHQ-2 [52-55]; cut-off  $\geq 3$ ), and anxiety was screened using the Generalized Anxiety Disorder 7-item scale (GAD-7 [56]; cut-off  $\geq 10$ ). Subsequently, two dichotomous variables were created: “Positive Depression” and “Positive Anxiety.” The outcome was a composite variable created using the aforementioned variables with the following categories: negative screening for both depression and anxiety, positive

screening for depression only, positive screening for anxiety only, and positive screening for both depression and anxiety.

### Independent Variables

Demographic information included sex, age, educational level (aggregated as primary/secondary education, a professional degree, a university degree, or a master’s/PhD degree), number of people living in the household, frontline worker (yes/no), and country of residence (Brazil/Spain). Social distancing/self-isolation was considered as a dichotomous variable (yes/no).

Questions related to COVID-19 were as follows: “Have you been diagnosed with COVID-19?”; “Have you been admitted to a hospital or hospitalized due to COVID-19?”; and “Have you lost a significant other?” Possible answers were yes or no.

Lifestyle habits were assessed using the Short Multidimensional Inventory Lifestyle Evaluation–Confinement (SMILE-C) [41]. This scale was developed specifically to allow a multidimensional measure of lifestyle during the COVID-19

pandemic. It comprises 27 items made up of 7 domains (diet and nutrition, substance abuse, physical activity, stress management, restorative sleep, social support, and environmental exposures), with response options measured using a 4-point Likert scale. The SMILE-C has an overall Cronbach  $\alpha=0.75$  and Kaiser-Meyer-Olkin Measure=0.77. The higher the score, the healthier the lifestyle pattern. In this study, the SMILE-C total score was dichotomized at the 75th percentile (up to 85%).

Self-rated health was measured using the question “How would you rate your health in general?” with possible answer choices of “very bad,” “bad,” “neither good nor bad,” “good,” and “very good” [57]. Response options were aggregated into very good/good and neither good nor bad/bad/very bad.

Previously diagnosed conditions were investigated using the question “In the last 12 months, have you been diagnosed by a medical doctor or health professional, or received treatment for any of the following conditions?” Conditions included diabetes, heart disease, hypertension, stroke, anemia, asthma, depression, anxiety, bipolar disorder, schizophrenia, anorexia/bulimia, HIV/AIDS, cancer, tuberculosis, cirrhosis, and renal disease [58]. The conditions were then aggregated as chronic diseases, mental health disorders, and infectious diseases.

Screening for alcohol abuse was performed using the Alcohol Use Disorder Identification Test (AUDIT-C; cut-off  $\geq 3$  [59]).

Changes in the SMILE-C domains during the pandemic were evaluated using questions like “Did you change your nutritional habits and diet during the COVID-19 pandemic?” Response options were measured using a 4-point Likert scale (completely, moderately, mildly, not at all) and aggregated into completely/moderately and mildly/not at all.

### Statistical Analysis

Nonresponse treatment is described in [Multimedia Appendix 2](#). Independent variables were described by outcome and proportions compared using chi-squared tests. Taking into consideration the complex, multiple associations of the different covariates with the outcomes under analysis (screening for depression and anxiety), preliminary analyses using Least Absolute Shrinkage and Selection Operator (LASSO [60, 61]), a simple machine learning procedure, were employed. LASSO is a penalized regression analysis method that helps to optimize variable selection and regularization in order to enhance the accuracy of the model to be implemented. The subset of factors (variables) that did not contribute to the hypothetical model

under assessment yield zero coefficients [62]. Such variables were excluded from subsequent multivariable analyses ([Multimedia Appendix 2](#)).

The second procedure to optimize variable selection was based on the simulation of different modeling strategies with the subsequent choice of the best subset of variables to be included in a parsimonious model based on the best  $R^2$  coefficient [63, 64]. The glmnet library and the regsubsets function from R 4.0.2 (The R Foundation) were used.

Using this subset of variables, a multinomial logistic regression was fitted, taking as the reference category “negative screening for both depression and anxiety.” Analyses used the backward strategy, with the progressive elimination of variables based on the results of the Wald test and maximum likelihood estimation of fitness, considering a significance level of 5.0%. The model fitness was evaluated using different diagnostic tools, such as the Hosmer-Lemeshow statistics, the Pearson chi-squared test, as well as the deviance information criterion. Additionally, an analysis of residuals was performed ([Multimedia Appendix 2](#)). The model yielded adjusted odds ratios (AORs) with their respective 95% CIs.

### Ethical Aspects

The study was approved by the Ethics Committee at the Hospital Universitari i Politècnic La Fe in Valencia, Spain, and by the Comissão Nacional de Ética em Pesquisa (CONEP, Brazil – 3.968.686). The survey was anonymous (no identifying data like name, zip code, or IP address were collected), and participants read the consent form and confirmed their interest in participating before starting the questionnaire. As mentioned before, as a direct benefit, participants were provided with tips on healthy lifestyles and reliable websites and telephone numbers for additional information regarding COVID-19.

## Results

Overall, 24,657 questionnaires were initiated, and 22,786 were eligible for the main study. Of those, 3745 reported to be working as an essential worker during the COVID-19 pandemic and comprised the study sample ([Figure 2](#)). In total, 2842 participants were from Brazil and 903 from Spain. Most were female (Brazil:  $n=2052$ , 72.2%; Spain:  $n=640$ , 70.9%;  $P=.44$ ), with a median age of 39 (IQR 32–51) years for Brazil versus 43 (IQR 32–52) years for Spain ( $P=.07$ ). Half of the sample ( $n=457$ , 50.6%) reported being a frontline personnel in Spain compared to 28.9% ( $n=822$ ) in Brazil ( $P<.001$ ) ([Table 1](#)).

**Table 1.** Demographics, COVID-19 experience, lifestyle, and self-reported health conditions by country among essential workers from Brazil and Spain (N=3745), April to May 2020.

Variable	Brazil (n=2842), n (%)	Spain (n=903), n (%)	Total (N=3745), n (%)	P value
<b>Screening for depression and/or anxiety</b>				<.001
Negative for both depression and anxiety	1280 (45.0)	693 (76.7)	1973 (52.7)	
Positive for depression only	262 (9.2)	49 (5.4)	311 (8.3)	
Positive for anxiety only	360 (12.7)	74 (8.2)	434 (11.6)	
Positive for both depression and anxiety	940 (33.1)	87 (9.6)	1027 (27.4)	
<b>Sex</b>				.44
Male	790 (27.8)	263 (29.1)	1053 (28.1)	
Female	2052 (72.2)	640 (70.9)	2692 (71.9)	
Age (years), median (IQR)	39 (32-51)	43 (32-52)	40 (32-51)	.07
<b>Educational level</b>				<.001
Primary/secondary education or professional degree	425 (15.0)	242 (26.8)	667 (17.8)	
University degree	1606 (56.5)	354 (39.2)	1960 (52.3)	
Master's/PhD degree	811 (28.5)	307 (34.0)	1118 (29.9)	
Frontline worker	822 (28.9)	457 (50.6)	1279 (34.2)	<.001
People living in the household, median (IQR) <sup>a</sup>	3 (2-4)	3 (2-4)	3 (2-4)	.56
Self-isolated <sup>b</sup>	1501 (53.3)	181 (20.2)	1682 (45.3)	<.001
Diagnosed with COVID-19 <sup>c</sup>	69 (2.4)	35 (3.9)	104 (2.8)	.02
Lost someone during the pandemic <sup>c</sup>	254 (9.0)	97 (10.8)	351 (9.4)	.10
SMILE-C <sup>d</sup> , median (IQR)	78 (71-84)	80 (75-85)	79 (72-85)	<.001
Self-reported health (neither good nor bad, bad or very bad) <sup>e</sup>	708 (24.9)	210 (23.3)	918 (24.5)	.32
<b>Diagnosed with or treated for...</b>				
Chronic diseases <sup>f</sup>	929 (32.9)	233 (26.1)	1162 (31.3)	<.001
Mental health disorders <sup>g</sup>	865 (30.9)	109 (12.2)	974 (26.4)	<.001
Infectious diseases <sup>h</sup>	108 (3.8)	3 (0.3)	111 (3.0)	<.001
Positive screening for alcohol abuse	1260 (44.3)	289 (32.0)	1549 (41.4)	<.001
<b>Changes in...</b>				
Dietary and nutritional habits <sup>i</sup>	1257 (44.3)	217 (24.0)	1474 (39.4)	<.001
Substance use habits <sup>j</sup>	459 (17.6)	83 (9.3)	542 (15.5)	<.001
Physical activity routine <sup>k</sup>	1656 (58.8)	584 (64.7)	2240 (60.2)	.002
Strategies to manage stress <sup>l</sup>	1530 (53.9)	281 (31.2)	1811 (48.5)	<.001
Sleep patterns <sup>i</sup>	1219 (42.9)	243 (26.9)	1462 (39.1)	<.001
Social support <sup>m</sup>	1543 (54.8)	336 (37.6)	1879 (50.6)	<.001
Time spent indoors/outdoors <sup>c</sup>	2461 (86.7)	823 (91.5)	3284 (87.9)	<.001

<sup>a</sup>n=2<sup>b</sup>n=30<sup>c</sup>n=7<sup>d</sup>SMILE-C: Short Multidimensional Inventory Lifestyle Evaluation-Confinement; the higher the score, the healthier the lifestyle.<sup>e</sup>n=5<sup>f</sup>n=31



<sup>g</sup><sub>n=56</sub><sup>h</sup><sub>n=6</sub><sup>i</sup><sub>n=3</sub><sup>j</sup><sub>n=242</sub><sup>k</sup><sub>n=25</sub><sup>l</sup><sub>n=8</sub><sup>m</sup><sub>n=33</sub>

The prevalence of positive screenings for depression, anxiety, and comorbidity of both was 8.3% (n=311), 11.6% (n=434), and 27.4% (n=1027), respectively. All were higher in Brazil compared to Spain (Table 1). Table 2 describes the

sociodemographic and clinical characteristics of the sample across the outcome categories (negative for depression and anxiety, positive for depression only, positive for anxiety only, and positive for both depression and anxiety).

**Table 2.** Demographics, COVID-19 experience, lifestyle, and self-reported health conditions by mental health outcomes among essential workers from Brazil and Spain (N=3745), April to May 2020.

Variable	Negative for both depression and anxiety (n=1973)	Positive for depression only (n=311)	Positive for anxiety only (n=434)	Positive for depression and anxiety (n=1027)	P value
<b>Country , n (%)</b>					<.001
Brazil	1280 (64.9)	262 (84.2)	360 (82.9)	940 (91.5)	
Spain	693 (35.1)	49 (15.8)	74 (17.1)	87 (8.5)	
<b>Sex , n (%)</b>					<.001
Male	619 (31.4)	91 (29.3)	112 (25.8)	231 (22.5)	
Female	1354 (68.6)	220 (70.7)	322 (74.2)	796 (77.5)	
Age (years) , mean (SD)	44.56 (12.33)	39.25 (11.78)	41.10 (10.63)	37.43 (10.96)	<.001
<b>Educational level , n (%)</b>					<.001
Primary/secondary education or professional degree	340 (17.2)	43 (13.8)	62 (14.3)	222 (21.6)	
University degree	984 (49.9)	177 (56.9)	228 (52.5)	571 (55.6)	
Master's/PhD degree	649 (32.9)	91 (29.3)	144 (33.2)	234 (22.8)	
Frontline worker, n (%)	660 (33.5)	83 (26.7)	172 (39.6)	364 (35.4)	.002
People living in the household, median (IQR) <sup>a</sup>	3 (2-4)	3 (2-4)	3 (2-4)	3 (2-4)	.008
Self-isolated <sup>b</sup> , n (%)	824 (41.9)	163 (53.1)	191 (44.5)	504 (49.7)	<.001
Diagnosed with COVID-19 <sup>c</sup> , n (%)	45 (2.3)	8 (2.6)	18 (4.2)	33 (3.2)	.13
Lost someone in the pandemic <sup>c</sup> , n (%)	169 (8.6)	24 (7.7)	48 (11.1)	110 (10.7)	.11
SMILE-C <sup>d</sup> , mean (SD)	82.23 (7.71)	76.16 (7.73)	77.67 (7.43)	71.41 (8.73)	<.001
Self-reported health (neither good nor bad, bad, or very bad) <sup>e</sup>	280 (14.2)	77 (24.8)	112 (25.8)	449 (43.8)	<.001
<b>Diagnosed with or treated for...</b>					
Chronic diseases <sup>f</sup>	568 (29.1)	75 (24.2)	140 (32.4)	379 (37.3)	<.001
Mental health disorders <sup>g</sup>	270 (13.8)	74 (24.0)	129 (30.3)	501 (50.4)	<.001
Infectious diseases <sup>h</sup>	50 (2.5)	8 (2.6)	14 (3.2)	39 (3.8)	.26
Positive screening for alcohol abuse	736 (37.3)	143 (46.0)	195 (44.9)	475 (46.3)	<.001
<b>Changes in...</b>					
Dietary and nutritional habits <sup>i</sup>	641 (32.5)	145 (46.6)	207 (47.7)	481 (46.9)	<.001
Substance use habits <sup>j</sup>	213 (11.4)	50 (17.4)	64 (16.2)	215 (22.5)	<.001
Physical activity routine <sup>k</sup>	1187 (60.4)	193 (62.9)	287 (66.4)	573 (56.4)	.003
Strategies to manage stress <sup>l</sup>	905 (45.9)	157 (50.6)	255 (58.9)	494 (48.2)	<.001
Sleep patterns <sup>i</sup>	540 (27.4)	133 (42.8)	203 (46.9)	586 (57.1)	<.001
Social support <sup>m</sup>	892 (45.6)	158 (51.6)	240 (55.6)	589 (58.0)	<.001
Time spent indoors/outdoors <sup>c</sup>	1756 (89.1)	269 (86.5)	381 (88.0)	878 (85.7)	.049

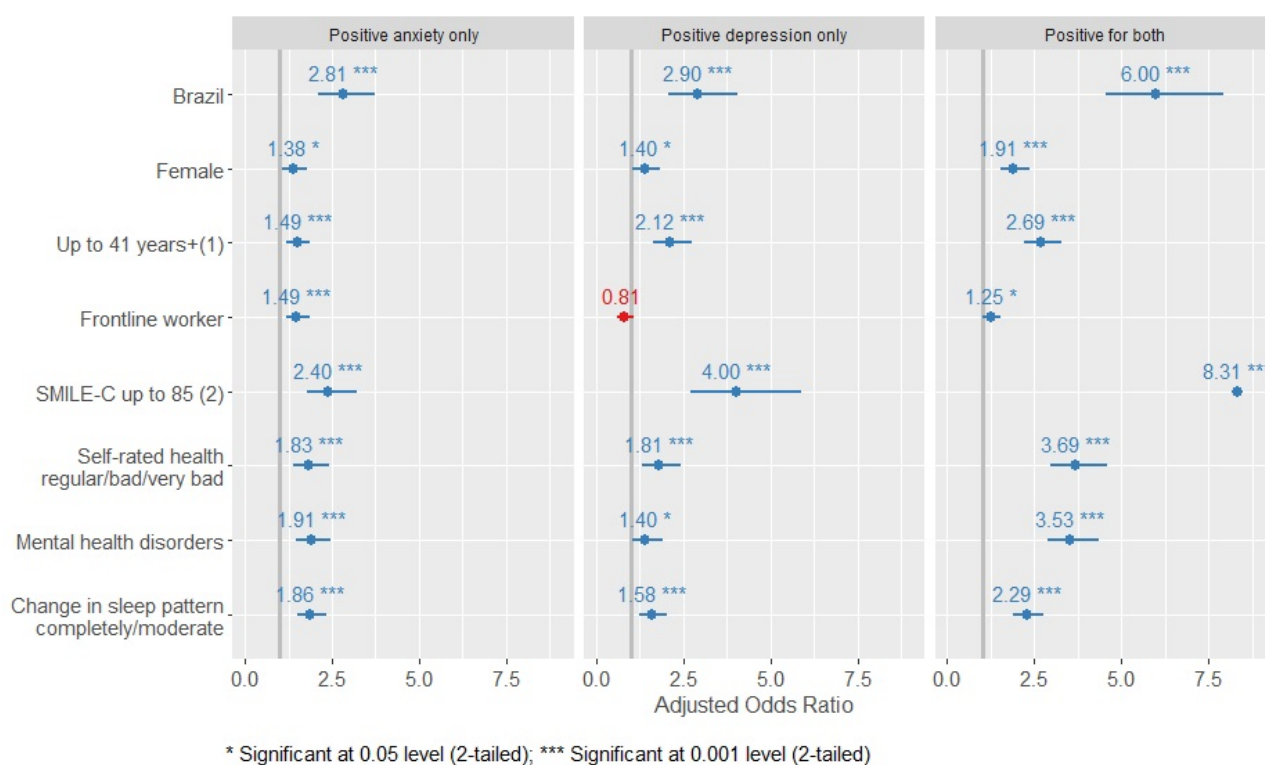
<sup>a</sup> n=2<sup>b</sup> n=30<sup>c</sup> n=7<sup>d</sup>SMILE-C: Short Multidimensional Inventory Lifestyle Evaluation-Confinement; the higher the score, the healthier the lifestyle.<sup>e</sup> n=5

f<sub>n</sub>=31g<sub>n</sub>=56h<sub>n</sub>=6i<sub>n</sub>=3j<sub>n</sub>=242k<sub>n</sub>=25l<sub>n</sub>=8m<sub>n</sub>=33

In the multinomial model, living in Brazil was associated with an AOR of 2.89 (95% CI 2.07-4.06) for a positive screening for depression, an AOR of 2.81 (95% CI 2.11-3.74) for anxiety, and an AOR of 5.99 (95% CI 4.53-7.91) for both conditions compared to living in Spain. An unhealthy lifestyle was associated with an AOR of 4.00 (95% CI 2.72-5.87) for depression, an AOR of 2.39 (95% CI 1.80-3.20) for anxiety, and an AOR of 8.30 (95% CI 5.90-11.7) for both conditions. Interestingly, being a frontline worker was not associated with depression only (AOR 0.81, 95% CI 0.60-1.08), although it

increased the likelihood of having anxiety (AOR 1.49, 95% CI 1.18-1.87) and both conditions (AOR 1.25, 95% CI 1.02-1.53). Additionally, we found that being female, being younger, presenting moderate or substantial changes in sleep patterns during the COVID-19 pandemic, being diagnosed with or treated for mental health disorders in the last year, and reporting a reduction in self-rated health were all associated with a higher likelihood of having depression, anxiety, or both conditions (Figure 4 and Multimedia Appendix 2).

**Figure 4.** Factors associated with depression or anxiety or both (multinomial model) among essential workers from Brazil and Spain (n=3745). April-May, 2020. Note: (1) Age was dichotomized by the sample median age. (2) The SMILE-C (Short Multidimensional Inventory Lifestyle Evaluation-Confinement) was dichotomized at the percentile 75%, the higher the score, the healthier the lifestyle).



## Discussion

In a web survey of 3745 essential workers from Brazil and Spain, we showed 8.3%, 11.6%, and 27.4% presenting positive screenings for depression, anxiety, and both conditions, respectively. As in previous studies conducted during the COVID-19 pandemic, we found that women, younger workers, frontline workers, those reporting a mental health diagnosis or treatment in the last 12 months, and those reporting changes in sleep patterns presented a higher likelihood of a positive

screening for anxiety and/or depression. Most importantly, higher odds ratios were observed in those living in Brazil and in those following an unhealthy lifestyle.

During data collection, Spain had 4 times the number of COVID-19 deaths than Brazil and had adopted a strict lockdown policy. It could be expected that essential workers under these conditions would be more prone to present anxiety and depressive symptoms. However, Brazil has additional social, structural, and political problems that may affect mental health. Recently, Baqui et al [65] reported on the higher mortality risk

among “pardo” and Black Brazilians admitted to hospital due to COVID-19; ethnicity was the second most important risk factor for death (after age). In addition, the authors highlighted that comorbid diseases and death were more common among Brazilians from the North region compared to the Central-South (except for Rio de Janeiro). Ribeiro et al [66] highlighted that the worst public health and socioeconomic scenarios were present in the northern regions of Brazil; higher proportions of individuals living in substandard housing (slums), with reduced schooling and a lack of sanitation and piped water, may interfere with adherence to hygiene recommendations. Although both papers discussed intracountry inequalities, social and health inequalities may be even higher between different countries [67] (Figure 1). In addition, the political instability and the government’s failure to acknowledge the seriousness of the pandemic (eg, official data on COVID-19 was not being published) [68] may be worsening the consequences of COVID-19 in Brazil, including repercussions concerning the mental health of essential workers. These workers may be seeing a high number of casualties, and working under intense fear and feelings of impotence, which may be related to the higher odds for anxiety and depression observed in our study.

Self-reported unhealthy lifestyle behaviors during confinement were associated with an increased likelihood of presenting a positive screening for both anxiety and depression in our study. Several studies have assessed psychological distress in HCWs during COVID-19, but its association with lifestyle remains underresearched. To our knowledge, the present study is the first web survey designed to assess a wide range of lifestyle changes and its relationship with anxiety and depression among essential workers during the pandemic. Among HCWs from New York, where almost half screened positive for depression, and one third for anxiety [30], physical activity/exercise was the most commonly endorsed solution to cope with COVID-19–related psychological distress, but its relationship with anxiety and depression was not explored. Our results are consistent with those pertaining to the general population and clinical studies. For instance, in Australia, adults who reported negative changes in physical activity, sleep, smoking, and alcohol intake after the onset of COVID-19 were more likely to have higher rates of depression, anxiety, and stress symptoms [69]. In the same way, individuals with anxiety and depression have shown higher ratios of unhealthy lifestyle habits, including poor diet quality, impaired sleep, reduced physical activity, smoking, and substance and/or alcohol misuse [39]. Based on the present results, the relationship between anxiety/depression and lifestyle as a multidimensional construct applies also to essential workers during the COVID-19 pandemic.

Our findings regarding sleep changes are also in accordance with a meta-analysis that showed that about 50% of HCWs have poor sleep quality in general (ie, during nonpandemic times) [36]. Subjective sleep quality, defined by the satisfaction with one’s overall sleep experience, may worsen among frontline HCWs treating patients with COVID-19 [42]. The present study expands the association between changes in sleep and anxiety/depression to a wider group of essential workers. This concurs with evidence supporting a bidirectional relationship between sleep disturbances and anxiety/depression [70].

Moreover, a reduced quality of sleep was associated with higher levels of depressive and anxiety symptoms during the COVID-19 lockdown in Italy [71].

Our results are in accordance with most of the literature: women and youth [17,31,32,72], frontline workers [31,33,34], a diagnosis of or treatment for mental health disorders in the past 12 months, and self-rated poor health [17] all increased the likelihood of depression and anxiety. Mental health disorders are associated with higher mortality rates and shorter life expectancies [73]. Our results, which support that of The Lancet Commission on Global Mental Health and Sustainable Development [74], showed the importance of social and environmental factors in mental health and highlighted the additional challenges experienced by populations living in countries with higher rates of inequality.

Consistent with other web surveys, where the population is recruited through social networks, our sample is not probabilistic and may therefore not represent the entire population of essential workers from Brazil and Spain [75]. Additionally, in 2018, 67% of Brazilian households had internet access (48% among the lower economic strata) [76] compared to 86.4% in Spain [77]. Although this difference may not have contributed to the different prevalence found between the countries (as we may have surveyed individuals with age, income, and schooling more similar to the Spanish population), we may be overrepresenting the highest socioeconomic strata in Brazil. Women were also overrepresented in both countries, as in many other web surveys conducted during the COVID-19 pandemic [18,23,32,34,72]. Due to the length of the original questionnaire, we did not include questions regarding gender and ethnicity/race, which may be associated with higher vulnerability to COVID-19 [65] and mental health outcomes. Additionally, we did not ask about the specific profession of essential workers and were not able to assess which professional groups were more vulnerable to mental health problems. In the United States [78], over 75% of Americans were estimated to work in occupations (including health care, manufacturing, retail, and food services) that are challenging to do from home. It was suggested that those workers may receive low wages and be subjected to stress due to the lower income and job insecurity, which could result in a large burden of mental health disorders.

One important strength of our survey was to disseminate reliable information on COVID-19 and strategies to maintain a healthy lifestyle. Considering the massive amount of information available, including fake news, and all the technology available for creating and disseminating online surveys, we believe that researchers can contribute to society by providing valuable information to respondents while obtaining data. Studies addressing participants’ opinions and effectiveness, when appropriate, on this information should be considered in the future.

Finally, our results provide additional support for The Lancet COVID-19 Commission [79], and are in accordance with Vigo et al [80], showing that countries with higher rates of inequality may be facing an important mental health burden in the forthcoming months.

## Acknowledgments

We would like to thank Gabriel Madeira and Ronaldo I Moreira for their assistance in creating and testing the online questionnaire; Thais Martini for social media management; Juliana Krapp and Daniele Souza for communications advice; all the institutions and individuals who promoted the survey; and all the individuals who answered it. RBDB is grateful for long-term funding from the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (E-26/203.154/2017) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (310541/2017-4). FIB acknowledges funding from FAPERJ (Health Networks; E-26/010.002428/2019). FK acknowledges funding from the Fundação de Amparo à Pesquisa do Estado de São Paulo (INCT 2014, 2014/50891-1), CNPq (INCT 465458/2014-9), and the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (47177.584.16785.16042020).

Funding sources had no role in the study design, data collection, analysis, interpretation of data, writing of the report, or in the decision to submit the paper for publication.

## Authors' Contributions

RBDB, VBM, and FK conceived and designed the study. JCM, TDAC, and PB analyzed the data and created the figures. RBDB and FIB supervised data analysis. RBDB, VBM, and BC revised the literature. RBDB wrote the first draft, and all authors revised and provided significant intellectual contribution. All the authors approved the submitted version.

## Conflicts of Interest

VBM has been a consultant, advisor, or Continuing Medical Education speaker over the last 3 years for the following companies: Angelini, Ferrer, Lundbeck, Nutrición Médica, and Otsuka. The remaining authors have no conflicts to declare.

### Multimedia Appendix 1

Data collection.

[PDF File (Adobe PDF File), 428 KB - [jmir\\_v22i10e22835\\_app1.pdf](#)]

### Multimedia Appendix 2

Statistical analyses.

[DOCX File, 224 KB - [jmir\\_v22i10e22835\\_app2.docx](#)]

## References

1. Silveira M, Barros A, Horta B, Pellanda L, Victora G, Dellagostin O, et al. Population-based surveys of antibodies against SARS-CoV-2 in Southern Brazil. medRxiv Preprint posted online on May 10, 2020. [doi: [10.1101/2020.05.01.20087205](#)]
2. Amsalem D, Dixon LB, Neria Y. The Coronavirus Disease 2019 (COVID-19) Outbreak and Mental Health: Current Risks and Recommended Actions. JAMA Psychiatry 2020 Jun 24. [doi: [10.1001/jamapsychiatry.2020.1730](#)] [Medline: [32579160](#)]
3. Greenberg N, Brooks SK, Wessely S, Tracy DK. How might the NHS protect the mental health of health-care workers after the COVID-19 crisis? Lancet Psychiatry 2020 May 28 [FREE Full text] [doi: [10.1016/S2215-0366\(20\)30224-8](#)] [Medline: [32473664](#)]
4. Balanzá-Martínez V, Atienza-Carbonell B, Kapczinski F, De Boni RB. Lifestyle behaviours during the COVID-19 - time to connect. Acta Psychiatr Scand 2020 May;141(5):399-400 [FREE Full text] [doi: [10.1111/acps.13177](#)] [Medline: [32324252](#)]
5. Verma S, Mishra A. Depression, anxiety, and stress and socio-demographic correlates among general Indian public during COVID-19. Int J Soc Psychiatry 2020 Dec 20;66(8):756-762. [doi: [10.1177/0020764020934508](#)] [Medline: [32567466](#)]
6. Hou T, Zhang T, Cai W, Song X, Chen A, Deng G, et al. Social support and mental health among health care workers during Coronavirus Disease 2019 outbreak: A moderated mediation model. PLoS One 2020;15(5):e0233831 [FREE Full text] [doi: [10.1371/journal.pone.0233831](#)] [Medline: [32470007](#)]
7. Moghanibashi-Mansourieh A. Assessing the anxiety level of Iranian general population during COVID-19 outbreak. Asian J Psychiatr 2020 Jun;51:102076 [FREE Full text] [doi: [10.1016/j.ajp.2020.102076](#)] [Medline: [32334409](#)]
8. Tian F, Li H, Tian S, Yang J, Shao J, Tian C. Psychological symptoms of ordinary Chinese citizens based on SCL-90 during the level I emergency response to COVID-19. Psychiatry Res 2020 Jun;288:112992 [FREE Full text] [doi: [10.1016/j.psychres.2020.112992](#)] [Medline: [32302816](#)]
9. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: [10.1016/j.psychres.2020.112954](#)] [Medline: [32325383](#)]
10. Ni MY, Yang L, Leung CMC, Li N, Yao XI, Wang Y, et al. Mental Health, Risk Factors, and Social Media Use During the COVID-19 Epidemic and Cordon Sanitaire Among the Community and Health Professionals in Wuhan, China:



- Cross-Sectional Survey. *JMIR Ment Health* 2020 May 12;7(5):e19009 [[FREE Full text](#)] [doi: [10.2196/19009](https://doi.org/10.2196/19009)] [Medline: [32365044](#)]
11. Lu W, Wang H, Lin Y, Li L. Psychological status of medical workforce during the COVID-19 pandemic: A cross-sectional study. *Psychiatry Res* 2020 Jun;288:112936 [[FREE Full text](#)] [doi: [10.1016/j.psychres.2020.112936](https://doi.org/10.1016/j.psychres.2020.112936)] [Medline: [32276196](#)]
  12. García-Fernández L, Romero-Ferreiro V, López-Roldán PD, Padilla S, Calero-Sierra I, Monzó-García M, et al. Mental health impact of COVID-19 pandemic on Spanish healthcare workers. *Psychol Med* 2020 May 27:1-3 [[FREE Full text](#)] [doi: [10.1017/S0033291720002019](https://doi.org/10.1017/S0033291720002019)] [Medline: [32456735](#)]
  13. Mo Y, Deng L, Zhang L, Lang Q, Liao C, Wang N, et al. Work stress among Chinese nurses to support Wuhan in fighting against COVID-19 epidemic. *J Nurs Manag* 2020 Jul;28(5):1002-1009 [[FREE Full text](#)] [doi: [10.1111/jonm.13014](https://doi.org/10.1111/jonm.13014)] [Medline: [32255222](#)]
  14. Tabah A, Ramanan M, Laupland KB, Buetti N, Cortegiani A, Mellinshoff J, PPE-SAFE contributors. Personal protective equipment and intensive care unit healthcare worker safety in the COVID-19 era (PPE-SAFE): An international survey. *J Crit Care* 2020 Oct;59:70-75 [[FREE Full text](#)] [doi: [10.1016/j.jcrc.2020.06.005](https://doi.org/10.1016/j.jcrc.2020.06.005)] [Medline: [32570052](#)]
  15. El-Hage W, Hingray C, Lemogne C, Yroni A, Brunault P, Bienvenu T, et al. [Health professionals facing the coronavirus disease 2019 (COVID-19) pandemic: What are the mental health risks?]. *Encephale* 2020 Jun;46(3S):S73-S80 [[FREE Full text](#)] [doi: [10.1016/j.encep.2020.04.008](https://doi.org/10.1016/j.encep.2020.04.008)] [Medline: [32370984](#)]
  16. Morgantini LA, Naha U, Wang H, Francavilla S, Acar O, Flores JM, et al. Factors Contributing to Healthcare Professional Burnout During the COVID-19 Pandemic: A Rapid Turnaround Global Survey. *medRxiv* 2020 May 22 [[FREE Full text](#)] [doi: [10.1101/2020.05.17.20101915](https://doi.org/10.1101/2020.05.17.20101915)] [Medline: [32511501](#)]
  17. Barello S, Palamenghi L, Graffigna G. Burnout and somatic symptoms among frontline healthcare professionals at the peak of the Italian COVID-19 pandemic. *Psychiatry Res* 2020 May 27;290:113129 [[FREE Full text](#)] [doi: [10.1016/j.psychres.2020.113129](https://doi.org/10.1016/j.psychres.2020.113129)] [Medline: [32485487](#)]
  18. Temsah M, Al-Sohime F, Alamro N, Al-Eyadhy A, Al-Hasan K, Jamal A, et al. The psychological impact of COVID-19 pandemic on health care workers in a MERS-CoV endemic country. *J Infect Public Health* 2020 Jun;13(6):877-882 [[FREE Full text](#)] [doi: [10.1016/j.jiph.2020.05.021](https://doi.org/10.1016/j.jiph.2020.05.021)] [Medline: [32505461](#)]
  19. Du J, Dong L, Wang T, Yuan C, Fu R, Zhang L, et al. Psychological symptoms among frontline healthcare workers during COVID-19 outbreak in Wuhan. *Gen Hosp Psychiatry* 2020 Apr 03 [[FREE Full text](#)] [doi: [10.1016/j.genhosppsych.2020.03.011](https://doi.org/10.1016/j.genhosppsych.2020.03.011)] [Medline: [32381270](#)]
  20. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. *Brain Behav Immun* 2020 Oct 30;89:531-542 [[FREE Full text](#)] [doi: [10.1016/j.bbi.2020.05.048](https://doi.org/10.1016/j.bbi.2020.05.048)] [Medline: [32485289](#)]
  21. Spoorthy MS, Pratapa SK, Mahant S. Mental health problems faced by healthcare workers due to the COVID-19 pandemic-A review. *Asian J Psychiatr* 2020 Jun;51:102119 [[FREE Full text](#)] [doi: [10.1016/j.ajp.2020.102119](https://doi.org/10.1016/j.ajp.2020.102119)] [Medline: [32339895](#)]
  22. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsis E, Katsounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. *Brain Behav Immun* 2020 Aug;88:901-907 [[FREE Full text](#)] [doi: [10.1016/j.bbi.2020.05.026](https://doi.org/10.1016/j.bbi.2020.05.026)] [Medline: [32437915](#)]
  23. Chew NW, Lee GK, Tan BY, Jing M, Goh Y, Ngiam NJ, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. *Brain Behav Immun* 2020 Aug;88:559-565 [[FREE Full text](#)] [doi: [10.1016/j.bbi.2020.04.049](https://doi.org/10.1016/j.bbi.2020.04.049)] [Medline: [32330593](#)]
  24. Sharif S, Amin F, Hafiz M, Benzel E, Peev N, Dahlan RH, World Spinal Column Society Executive Board. COVID 19-Depression and Neurosurgeons. *World Neurosurg* 2020 Jun 06 [[FREE Full text](#)] [doi: [10.1016/j.wneu.2020.06.007](https://doi.org/10.1016/j.wneu.2020.06.007)] [Medline: [32512242](#)]
  25. Vanni G, Materazzo M, Santori F, Pellicciaro M, Costesta M, Orsaria P, et al. The Effect of Coronavirus (COVID-19) on Breast Cancer Teamwork: A Multicentric Survey. *In Vivo* 2020 Jun;34(3 Suppl):1685-1694. [doi: [10.21873/invivo.11962](https://doi.org/10.21873/invivo.11962)] [Medline: [32503830](#)]
  26. Wang S, Xie L, Xu Y, Yu S, Yao B, Xiang D. Sleep disturbances among medical workers during the outbreak of COVID-2019. *Occup Med (Lond)* 2020 Jul 17;70(5):364-369 [[FREE Full text](#)] [doi: [10.1093/occmed/kqaa074](https://doi.org/10.1093/occmed/kqaa074)] [Medline: [32372077](#)]
  27. Zhang SX, Liu J, Afshar Jahanshahi A, Nawaser K, Yousefi A, Li J, et al. At the height of the storm: Healthcare staff's health conditions and job satisfaction and their associated predictors during the epidemic peak of COVID-19. *Brain Behav Immun* 2020 Jul;87:144-146 [[FREE Full text](#)] [doi: [10.1016/j.bbi.2020.05.010](https://doi.org/10.1016/j.bbi.2020.05.010)] [Medline: [32387345](#)]
  28. Tan BYQ, Chew NWS, Lee GKH, Jing M, Goh Y, Yeo LLL, et al. Psychological Impact of the COVID-19 Pandemic on Health Care Workers in Singapore. *Ann Intern Med* 2020 Aug 18;173(4):317-320. [doi: [10.7326/M20-1083](https://doi.org/10.7326/M20-1083)] [Medline: [32251513](#)]
  29. Yang S, Kwak SG, Ko EJ, Chang MC. The Mental Health Burden of the COVID-19 Pandemic on Physical Therapists. *Int J Environ Res Public Health* 2020 May 25;17(10) [[FREE Full text](#)] [doi: [10.3390/ijerph17103723](https://doi.org/10.3390/ijerph17103723)] [Medline: [32466164](#)]
  30. Shechter A, Diaz F, Moise N, Anstey DE, Ye S, Agarwal S, et al. Psychological distress, coping behaviors, and preferences for support among New York healthcare workers during the COVID-19 pandemic. *Gen Hosp Psychiatry* 2020 Sep;66:1-8 [[FREE Full text](#)] [doi: [10.1016/j.genhosppsych.2020.06.007](https://doi.org/10.1016/j.genhosppsych.2020.06.007)] [Medline: [32590254](#)]

31. Elbay RY, Kurtulmuş A, Arpacioğlu S, Karadere E. Depression, anxiety, stress levels of physicians and associated factors in Covid-19 pandemics. *Psychiatry Res* 2020 Aug;290:113130 [FREE Full text] [doi: [10.1016/j.psychres.2020.113130](https://doi.org/10.1016/j.psychres.2020.113130)] [Medline: [32497969](https://pubmed.ncbi.nlm.nih.gov/32497969/)]
32. Rossi R, Soggi V, Pacitti F, Di Lorenzo G, Di Marco A, Siracusano A, et al. Mental Health Outcomes Among Frontline and Second-Line Health Care Workers During the Coronavirus Disease 2019 (COVID-19) Pandemic in Italy. *JAMA Netw Open* 2020 May 01;3(5):e2010185 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.10185](https://doi.org/10.1001/jamanetworkopen.2020.10185)] [Medline: [32463467](https://pubmed.ncbi.nlm.nih.gov/32463467/)]
33. Liu C, Yang Y, Zhang X, Xu X, Dou Q, Zhang W, et al. The prevalence and influencing factors in anxiety in medical workers fighting COVID-19 in China: a cross-sectional survey. *Epidemiol Infect* 2020 May 20;148:e98 [FREE Full text] [doi: [10.1017/S0950268820001107](https://doi.org/10.1017/S0950268820001107)] [Medline: [32430088](https://pubmed.ncbi.nlm.nih.gov/32430088/)]
34. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. *JAMA Netw Open* 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.3976](https://doi.org/10.1001/jamanetworkopen.2020.3976)] [Medline: [32202646](https://pubmed.ncbi.nlm.nih.gov/32202646/)]
35. Li G, Miao J, Wang H, Xu S, Sun W, Fan Y, et al. Psychological impact on women health workers involved in COVID-19 outbreak in Wuhan: a cross-sectional study. *J Neurol Neurosurg Psychiatry* 2020 Aug;91(8):895-897. [doi: [10.1136/jnnp-2020-323134](https://doi.org/10.1136/jnnp-2020-323134)] [Medline: [32366684](https://pubmed.ncbi.nlm.nih.gov/32366684/)]
36. Qiu D, Yu Y, Li R, Li Y, Xiao S. Prevalence of sleep disturbances in Chinese healthcare professionals: a systematic review and meta-analysis. *Sleep Med* 2020 Mar;67:258-266. [doi: [10.1016/j.sleep.2019.01.047](https://doi.org/10.1016/j.sleep.2019.01.047)] [Medline: [31040078](https://pubmed.ncbi.nlm.nih.gov/31040078/)]
37. Noordsy DL. *Lifestyle Psychiatry*. Washington DC: American Psychiatric Association; 2019.
38. Firth J, Ward PB, Stubbs B. Editorial: Lifestyle Psychiatry. *Front Psychiatry* 2019 Aug 26;10:597 [FREE Full text] [doi: [10.3389/fpsy.2019.00597](https://doi.org/10.3389/fpsy.2019.00597)] [Medline: [31507466](https://pubmed.ncbi.nlm.nih.gov/31507466/)]
39. Firth J, Rosenbaum S, Galletly C, Siddiqi N, Stubbs B, Killackey E, et al. Protecting physical health in people with mental illness – Authors' reply. *The Lancet Psychiatry* 2019 Nov;6(11):890-891. [doi: [10.1016/s2215-0366\(19\)30387-6](https://doi.org/10.1016/s2215-0366(19)30387-6)]
40. Firth J, Solmi M, Wootton RE, Vancampfort D, Schuch FB, Hoare E, et al. A meta-review of "lifestyle psychiatry": the role of exercise, smoking, diet and sleep in the prevention and treatment of mental disorders. *World Psychiatry* 2020 Oct 15;19(3):360-380 [FREE Full text] [doi: [10.1002/wps.20773](https://doi.org/10.1002/wps.20773)] [Medline: [32931092](https://pubmed.ncbi.nlm.nih.gov/32931092/)]
41. Balanzá-Martínez V, Kapczinski F, de Azevedo Cardoso T, Atienza-Carbonell B, Rosa AR, Mota JC, et al. The assessment of lifestyle changes during the COVID-19 pandemic using a multidimensional scale. *Rev Psiquiatr Salud Ment* 2020 Aug 29 [FREE Full text] [doi: [10.1016/j.rpsm.2020.07.003](https://doi.org/10.1016/j.rpsm.2020.07.003)] [Medline: [32962948](https://pubmed.ncbi.nlm.nih.gov/32962948/)]
42. Xiao H, Zhang Y, Kong D, Li S, Yang N. The Effects of Social Support on Sleep Quality of Medical Staff Treating Patients with Coronavirus Disease 2019 (COVID-19) in January and February 2020 in China. *Med Sci Monit* 2020 Mar 05;26:e923549 [FREE Full text] [doi: [10.12659/MSM.923549](https://doi.org/10.12659/MSM.923549)] [Medline: [32132521](https://pubmed.ncbi.nlm.nih.gov/32132521/)]
43. Kastorini C, Milionis HJ, Esposito K, Giugliano D, Goudevenos JA, Panagiotakos DB. The effect of Mediterranean diet on metabolic syndrome and its components: a meta-analysis of 50 studies and 534,906 individuals. *J Am Coll Cardiol* 2011 Mar 15;57(11):1299-1313 [FREE Full text] [doi: [10.1016/j.jacc.2010.09.073](https://doi.org/10.1016/j.jacc.2010.09.073)] [Medline: [21392646](https://pubmed.ncbi.nlm.nih.gov/21392646/)]
44. Gleeson M, Bishop NC, Stensel DJ, Lindley MR, Mastana SS, Nimmo MA. The anti-inflammatory effects of exercise: mechanisms and implications for the prevention and treatment of disease. *Nat Rev Immunol* 2011 Aug 05;11(9):607-615. [doi: [10.1038/nri3041](https://doi.org/10.1038/nri3041)] [Medline: [21818123](https://pubmed.ncbi.nlm.nih.gov/21818123/)]
45. Donkin A, Goldblatt P, Allen J, Nathanson V, Marmot M. Global action on the social determinants of health. *BMJ Glob Health* 2018 Dec 01;3(Suppl 1):e000603. [doi: [10.1136/bmjgh-2017-000603](https://doi.org/10.1136/bmjgh-2017-000603)] [Medline: [29379648](https://pubmed.ncbi.nlm.nih.gov/29379648/)]
46. Holmes SM, Hansen H, Jenks A, Stonington SD, Morse M, Greene JA, et al. Misdiagnosis, Mistreatment, and Harm - When Medical Care Ignores Social Forces. *N Engl J Med* 2020 Mar 19;382(12):1083-1086. [doi: [10.1056/NEJMp1916269](https://doi.org/10.1056/NEJMp1916269)] [Medline: [32187466](https://pubmed.ncbi.nlm.nih.gov/32187466/)]
47. Country profiles-Brazil, Spain. Seattle, WA: IHME, University of Washington. 2018. URL: <http://www.healthdata.org/results/country-profiles> [accessed 2020-07-03]
48. Coronavirus Disease (COVID-19) Dashboard - Brazil. World Health Organization. 2020. URL: <https://covid19.who.int/region/amro/country/br> [accessed 2020-07-03]
49. Coronavirus Disease (COVID-19) Dashboard - Spain. World Health Organization. 2020. URL: <https://covid19.who.int/region/euro/country/es> [accessed 2020-07-03]
50. World Development Indicators. World Bank. URL: <http://datatopics.worldbank.org/world-development-indicators/themes/poverty-and-inequality.html> [accessed 2020-07-03]
51. Tiffany J. Respondent-driven sampling in participatory research contexts: participant-driven recruitment. *J Urban Health* 2006 Nov;83(6 Suppl):i113-i124 [FREE Full text] [doi: [10.1007/s11524-006-9107-9](https://doi.org/10.1007/s11524-006-9107-9)] [Medline: [16933100](https://pubmed.ncbi.nlm.nih.gov/16933100/)]
52. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003 Nov;41(11):1284-1292. [doi: [10.1097/01.MLR.0000093487.78664.3C](https://doi.org/10.1097/01.MLR.0000093487.78664.3C)] [Medline: [14583691](https://pubmed.ncbi.nlm.nih.gov/14583691/)]
53. de Lima Osório F, Vilela Mendes A, Crippa JA, Loureiro SR. Study of the discriminative validity of the PHQ-9 and PHQ-2 in a sample of Brazilian women in the context of primary health care. *Perspect Psychiatr Care* 2009 Jul;45(3):216-227. [doi: [10.1111/j.1744-6163.2009.00224.x](https://doi.org/10.1111/j.1744-6163.2009.00224.x)] [Medline: [19566694](https://pubmed.ncbi.nlm.nih.gov/19566694/)]

54. Rodríguez-Muñoz MDLF, Castelao Legazpi PC, Olivares Crespo ME, Soto Balbuena C, Izquierdo Méndez N, Ferrer Barrientos FJ, et al. [PHQ-2 as First Screening Instrument of Prenatal Depression in Primary Health Care, Spain]. *Rev Esp Salud Publica* 2017 Jan 30;91 [FREE Full text] [Medline: [28134236](#)]
55. Cano-Vindel A, Muñoz-Navarro R, Medrano LA, Ruiz-Rodríguez P, González-Blanch C, Gómez-Castillo MD, PsicAP Research Group. A computerized version of the Patient Health Questionnaire-4 as an ultra-brief screening tool to detect emotional disorders in primary care. *J Affect Disord* 2018 Jul;234:247-255. [doi: [10.1016/j.jad.2018.01.030](#)] [Medline: [29549826](#)]
56. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006 May 22;166(10):1092-1097. [doi: [10.1001/archinte.166.10.1092](#)] [Medline: [16717171](#)]
57. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Soc Sci Med* 2009 Aug;69(3):307-316. [doi: [10.1016/j.socscimed.2009.05.013](#)] [Medline: [19520474](#)]
58. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional de Saúde 2013: percepção do estado de saúde, estilos de vida e doenças crônicas: Brasil, Grandes Regiões e Unidades da Federação. Rio de Janeiro: IBGE; 2014.
59. Bush K, Kivlahan DR, McDonnell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Arch Intern Med* 1998 Sep 14;158(16):1789-1795. [doi: [10.1001/archinte.158.16.1789](#)] [Medline: [9738608](#)]
60. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Statistical Soc B* 2005 Apr;67(2):301-320. [doi: [10.1111/j.1467-9868.2005.00503.x](#)]
61. Musoro JZ, Zwiderman AH, Puhon MA, ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. *BMC Med Res Methodol* 2014 Oct 16;14:116 [FREE Full text] [doi: [10.1186/1471-2288-14-116](#)] [Medline: [25323009](#)]
62. Tutz G, Pöbnecker W, Uhlmann L. Variable selection in general multinomial logit models. 2015 Feb;82:207-222. [doi: [10.1016/j.csda.2014.09.009](#)]
63. de Jong VMT, Eijkemans MJC, van Calster B, Timmerman D, Moons KGM, Steyerberg EW, et al. Sample size considerations and predictive performance of multinomial logistic prediction models. *Stat Med* 2019 Apr 30;38(9):1601-1619 [FREE Full text] [doi: [10.1002/sim.8063](#)] [Medline: [30614028](#)]
64. Smith GCS, Seaman SR, Wood AM, Royston P, White IR. Correcting for optimistic prediction in small data sets. *Am J Epidemiol* 2014 Aug 01;180(3):318-324 [FREE Full text] [doi: [10.1093/aje/kwu140](#)] [Medline: [24966219](#)]
65. Baqui P, Bica I, Marra V, Ercole A, van DSM. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. medRxiv Preprint posted online on May 10, 2020. [doi: [10.1101/2020.05.19.20107094](#)]
66. Ribeiro H, Lima VM, Waldman EA. In the COVID-19 pandemic in Brazil, do brown lives matter? *Lancet Glob Health* 2020 Aug;8(8):e976-e977 [FREE Full text] [doi: [10.1016/S2214-109X\(20\)30314-4](#)] [Medline: [32622403](#)]
67. Nepomuceno MR, Acosta E, Albrez-Gutierrez D, Aburto JM, Gagnon A, Turra CM. Besides population age structure, health and other demographic factors can contribute to understanding the COVID-19 burden. *Proc Natl Acad Sci U S A* 2020 Jun 23;117(25):13881-13883. [doi: [10.1073/pnas.2008760117](#)] [Medline: [32576710](#)]
68. The Lancet Infectious Diseases. Political casualties of the COVID-19 pandemic. *Lancet Infect Dis* 2020 Jul;20(7):755 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30496-5](#)] [Medline: [32534606](#)]
69. Stanton R, To QG, Khalesi S, Williams SL, Alley SJ, Thwaite TL, et al. Depression, Anxiety and Stress during COVID-19: Associations with Changes in Physical Activity, Sleep, Tobacco and Alcohol Use in Australian Adults. *Int J Environ Res Public Health* 2020 Jun 07;17(11):4065 [FREE Full text] [doi: [10.3390/ijerph17114065](#)] [Medline: [32517294](#)]
70. Alvaro PK, Roberts RM, Harris JK. A Systematic Review Assessing Bidirectionality between Sleep Disturbances, Anxiety, and Depression. *Sleep* 2013 Jul 01;36(7):1059-1068 [FREE Full text] [doi: [10.5665/sleep.2810](#)] [Medline: [23814343](#)]
71. Cellini N, Canale N, Mioni G, Costa S. Changes in sleep pattern, sense of time, and digital media use during COVID-19 lockdown in Italy. *PsyArXiv Preprint* posted online on May 15, 2020. [doi: [10.31234/osf.io/284mr](#)]
72. Kang L, Ma S, Chen M, Yang J, Wang Y, Li R, et al. Impact on mental health and perceptions of psychological care among medical and nursing staff in Wuhan during the 2019 novel coronavirus disease outbreak: A cross-sectional study. *Brain Behav Immun* 2020 Jul;87:11-17 [FREE Full text] [doi: [10.1016/j.bbi.2020.03.028](#)] [Medline: [32240764](#)]
73. Plana-Ripoll O, Pedersen CB, Agerbo E, Holtz Y, Erlangsen A, Canudas-Romo V, et al. A comprehensive analysis of mortality-related health metrics associated with mental disorders: a nationwide, register-based cohort study. *The Lancet* 2019 Nov;394(10211):1827-1835. [doi: [10.1016/s0140-6736\(19\)32316-5](#)]
74. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. *The Lancet* 2018 Oct;392(10157):1553-1598. [doi: [10.1016/S0140-6736\(18\)31612-X](#)]
75. De Boni RB. Web surveys in the time of COVID-19. *Cad Saude Publica* 2020;36(7):e00155820. [doi: [10.1590/0102-311x00155820](#)] [Medline: [32638874](#)]
76. TIC Domicílios - 2019. Núcleo de Informação e Coordenação do Ponto BR. 2020. URL: <https://cetic.br/pesquisa/domicilios/indicadores/> [accessed 2020-05-28]

77. Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares. INE. 2019. URL: [https://www.ine.es/prensa/tich\\_2019.pdf](https://www.ine.es/prensa/tich_2019.pdf) [accessed 2020-05-28]
78. Baker MG. Nonrelocatable Occupations at Increased Risk During Pandemics: United States, 2018. *Am J Public Health* 2020 Aug;110(8):1126-1132. [doi: [10.2105/ajph.2020.305738](https://doi.org/10.2105/ajph.2020.305738)]
79. Sachs JD, Horton R, Bagenal J, Ben Amor Y, Karadag Caman O, Lafortune G. The Lancet COVID-19 Commission. *The Lancet* 2020 Aug;396(10249):454-455. [doi: [10.1016/s0140-6736\(20\)31494-x](https://doi.org/10.1016/s0140-6736(20)31494-x)]
80. Vigo D, Thornicroft G, Gureje O. The Differential Outcomes of Coronavirus Disease 2019 in Low- and Middle-Income Countries vs High-Income Countries. *JAMA Psychiatry* 2020 Jun 11. [doi: [10.1001/jamapsychiatry.2020.2174](https://doi.org/10.1001/jamapsychiatry.2020.2174)] [Medline: [32525529](https://pubmed.ncbi.nlm.nih.gov/32525529/)]

## Abbreviations

**AOR:** adjusted odds ratio

**AUDIT-C:** Alcohol Use Disorder Identification Test

**CNPq:** Conselho Nacional de Desenvolvimento Científico e Tecnológico

**CONEP:** Comissão Nacional de Ética em Pesquisa

**FAPERJ:** Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro

**GAD-7:** Generalized Anxiety Disorder 7-item

**GDP:** gross domestic product

**HCW:** health care worker

**IP:** Internet Protocol

**LASSO:** Least Absolute Shrinkage and Selection Operator

**PHQ-2:** Patient Health Questionnaire-2

**SMILE-C:** Short Multidimensional Inventory Lifestyle Evaluation–Confinement

*Edited by G Eysenbach; submitted 24.07.20; peer-reviewed by A Bergeron, CH Wu; comments to author 22.09.20; revised version received 25.09.20; accepted 09.10.20; published 30.10.20.*

*Please cite as:*

De Boni RB, Balanzá-Martínez V, Mota JC, Cardoso TDA, Ballester P, Atienza-Carbonell B, Bastos FI, Kapczinski F. Depression, Anxiety, and Lifestyle Among Essential Workers: A Web Survey From Brazil and Spain During the COVID-19 Pandemic. *J Med Internet Res* 2020;22(10):e22835

URL: <http://www.jmir.org/2020/10/e22835/>

doi: [10.2196/22835](https://doi.org/10.2196/22835)

PMID: [33038075](https://pubmed.ncbi.nlm.nih.gov/33038075/)

©Raquel Brandini De Boni, Vicent Balanzá-Martínez, Jurema Correa Mota, Taiane De Azevedo Cardoso, Pedro Ballester, Beatriz Atienza-Carbonell, Francisco I Bastos, Flavio Kapczinski. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org/>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.



Original Paper

# Real-Life Gait Performance as a Digital Biomarker for Motor Fluctuations: The Parkinson@Home Validation Study

Luc JW Evers<sup>1,2</sup>, BSc; Yordan P Raykov<sup>3</sup>, PhD; Jesse H Krijthe<sup>2</sup>, PhD; Ana Lgia Silva de Lima<sup>1</sup>, PhD; Reham Badawy<sup>4</sup>, BSc; Kasper Claes<sup>5</sup>, PhD; Tom M Heskes<sup>2</sup>, PhD; Max A Little<sup>4</sup>, PhD; Marjan J Meinders<sup>6</sup>, PhD; Bastiaan R Bloem<sup>1</sup>, MD, PhD, FRCPE

<sup>1</sup>Center of Expertise for Parkinson and Movement Disorders, department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, Netherlands

<sup>2</sup>Institute for Computing and Information Sciences, Radboud University, Nijmegen, Netherlands

<sup>3</sup>Department of Mathematics, School of Engineering and Applied Sciences, Aston University, Birmingham, United Kingdom

<sup>4</sup>School of Computer Science, University of Birmingham, Birmingham, United Kingdom

<sup>5</sup>UCB Pharma, Brussels, Belgium

<sup>6</sup>Scientific Center for Quality of Healthcare (IQ healthcare), Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

**Corresponding Author:**

Luc JW Evers, BSc

Center of Expertise for Parkinson and Movement Disorders, department of Neurology

Donders Institute for Brain, Cognition and Behaviour

Radboud University Medical Center

Reinier Postlaan 4

Nijmegen

Netherlands

Phone: 31 3616600

Email: [luc.evers@radboudumc.nl](mailto:luc.evers@radboudumc.nl)

## Abstract

**Background:** Wearable sensors have been used successfully to characterize bradykinetic gait in patients with Parkinson disease (PD), but most studies to date have been conducted in highly controlled laboratory environments.

**Objective:** This paper aims to assess whether sensor-based analysis of real-life gait can be used to objectively and remotely monitor motor fluctuations in PD.

**Methods:** The Parkinson@Home validation study provides a new reference data set for the development of digital biomarkers to monitor persons with PD in daily life. Specifically, a group of 25 patients with PD with motor fluctuations and 25 age-matched controls performed unscripted daily activities in and around their homes for at least one hour while being recorded on video. Patients with PD did this twice: once after overnight withdrawal of dopaminergic medication and again 1 hour after medication intake. Participants wore sensors on both wrists and ankles, on the lower back, and in the front pants pocket, capturing movement and contextual data. Gait segments of 25 seconds were extracted from accelerometer signals based on manual video annotations. The power spectral density of each segment and device was estimated using Welch's method, from which the total power in the 0.5- to 10-Hz band, width of the dominant frequency, and cadence were derived. The ability to discriminate between before and after medication intake and between patients with PD and controls was evaluated using leave-one-subject-out nested cross-validation.

**Results:** From 18 patients with PD (11 men; median age 65 years) and 24 controls (13 men; median age 68 years),  $\geq 10$  gait segments were available. Using logistic LASSO (least absolute shrinkage and selection operator) regression, we classified whether the unscripted gait segments occurred before or after medication intake, with mean area under the receiver operator curves (AUCs) varying between 0.70 (ankle of least affected side, 95% CI 0.60-0.81) and 0.82 (ankle of most affected side, 95% CI 0.72-0.92) across sensor locations. Combining all sensor locations did not significantly improve classification (AUC 0.84, 95% CI 0.75-0.93). Of all signal properties, the total power in the 0.5- to 10-Hz band was most responsive to dopaminergic medication. Discriminating between patients with PD and controls was generally more difficult (AUC of all sensor locations combined: 0.76, 95% CI



0.62-0.90). The video recordings revealed that the positioning of the hands during real-life gait had a substantial impact on the power spectral density of both the wrist and pants pocket sensor.

**Conclusions:** We present a new video-referenced data set that includes unscripted activities in and around the participants' homes. Using this data set, we show the feasibility of using sensor-based analysis of real-life gait to monitor motor fluctuations with a single sensor location. Future work may assess the value of contextual sensors to control for real-world confounders.

(*J Med Internet Res* 2020;22(10):e19068) doi:[10.2196/19068](https://doi.org/10.2196/19068)

## KEYWORDS

digital biomarkers; remote patient monitoring; wearable sensors; real-life gait; Parkinson disease; biomarker; patient monitoring; wearables; gait

## Introduction

### Background

The core treatment of patients with Parkinson disease (PD) is symptomatic and consists of dopamine replacement therapy. Although motor symptoms such as bradykinesia and rigidity can be well controlled in early disease stages, most patients experience motor fluctuations in symptom severity after a few years of treatment with levodopa [1]. The pattern of these motor fluctuations varies between patients and may include wearing off, unpredictable off phases, peak-dose dyskinesias, and diphasic dyskinesias. Therefore, optimal management of motor fluctuations requires a highly personalized approach. The current evaluation of motor fluctuations involves paper diaries, such as the Hauser diary, which are burdensome to complete and which demonstrate reduced compliance when used for more than three days [2]. Moreover, patients are not always able to accurately recognize their own symptoms (eg, dyskinesia is often confused with tremor and vice versa). A more objective and unobtrusive way to assess motor fluctuations over longer periods of time in daily life could shed light on the real-world presentation of PD and benefit both individual patient care and follow-up in clinical trials.

Wearable movement sensors could potentially address this need. It is important to distinguish between active monitoring (ie, the analysis of specific tasks, such as the timed up and go [TUG] test [3]) and passive monitoring (ie, the analysis of natural behavior of daily living). Although active monitoring has received by far the most attention [4,5], this approach is susceptible to attrition in patient compliance [6]. By contrast, excellent long-term compliance can be achieved when patients are only asked to wear an unobtrusive sensor, such as a smartwatch, as they go about their daily activities [7]. Passive monitoring can also provide more continuous insights into symptom fluctuations throughout the day. However, these benefits come at a cost: it is challenging to develop algorithms that can cope with the large variation in signals encountered in daily life, of which only a small proportion may be explained by PD-related impairments. Since obtaining accurate labels in daily life environments is difficult and costly, most of the currently available reference data sets have been collected in the lab and involved patients performing a standardized set of tasks and activities. Algorithms trained on these data sets are unlikely to perform well in real-life settings [8]. Some studies have simulated a home environment in the lab and included unscripted behavior [9,10]. Although this is more realistic, all

data are still collected in the same environment, so the expected variation in activities and symptom severity is much reduced compared with real life. Consequently, the real-life performance of new sensor-based methods to assess motor fluctuations remains largely unknown.

Another challenge is finding the best strategy to derive meaningful and interpretable outcome measures from daily life sensor data. One approach is to detect and quantify a specific behavior that probes the presence of PD symptoms. Gait is a promising candidate in this regard for three reasons. First, gait is highly stereotypic, allowing for accurate detection in daily life signals. Second, gait becomes abnormal even in very early stages of the disease, likely already in the prodromal phase [11], and gait impairments worsen as the disease progresses [12,13]. Recording and analyzing gait is therefore useful across a wide spectrum of disease severities. Third, even among patients with PD without freezing of gait, the gait pattern changes in response to dopaminergic medication. Specifically, the step length and arm swing are reduced in the off state, which is known as bradykinetic gait [14,15]. However, these findings are based on data collected during standardized tasks (typically the TUG test) in highly controlled laboratory environments. It is not self-evident that these effects directly translate to daily life [16]. Various behavioral and environmental factors may influence gait in daily life. In addition, patients with PD show different gait patterns when they are aware that they are being evaluated (partially a Hawthorne effect and partially the phenomenon of *kinesia paradoxa*, the sudden transient ability of patients with PD to perform a task due to increased alertness or arousal) [17,18]. Indeed, gait patterns as measured in the lab and in real life (using accelerometry) are only weakly correlated [19,20].

Empirical evidence that underlines the value of analyzing real-life gait in PD is emerging. This approach may discriminate better than lab-based gait analysis between patients with PD and controls [19]. Similarly, real-life gait variability, derived from an accelerometer worn on the lower back, may predict the time to first fall better than various in-lab measurements, including gait speed in the off state [21]. In addition, a pilot study showed that at-home gait speed, measured using a radio wave-based monitor, correlates with PD severity [22]. However, studies that have examined how on-off changes are reflected in real-life gait patterns are scarce. Moore et al [23] suggested that step length derived from an ankle sensor could be used to monitor on-off changes in real life but illustrated this in only one patient. Sama et al [24] proposed that the energy in the 0- to 10-Hz frequency band during gait, obtained using a

waist-worn sensor, could predict the patient's motor state. When tested in daily life, the algorithm's predictions demonstrated high agreement with on-off diaries completed by the patients [25]. We conclude that analyzing real-life gait to monitor motor fluctuations has thus far shown promising results, but developments remain in the early stages. This is reflected by the lack of consensus on the best sensor location [26]. The field would benefit from an objective comparison of commonly used sensor locations on labelled data collected in the patients' home environments.

## Study Objective

In the Parkinson@Home validation study, we collected data from multiple wearable sensors during unscripted activities in and around the participants' homes and recorded the activities on video. The objective of this publication is twofold. First, we aim to describe this new reference data set, which will be made available to the research community in collaboration with the Michael J Fox Foundation. Second, using this data set, we aim to assess whether real-life gait analysis can be used to obtain a digital biomarker for motor fluctuations in PD. Specifically, we will assess how well various sensor locations can discriminate between gait performance before and after intake of dopaminergic medication and explore which factors complicate gait analysis in daily life.

## Methods

### Study Sample

We included a group of 25 patients with PD and 25 age-matched participants without PD (controls). Patients were recruited using various strategies, including advertisements in the Dutch Parkinson Patient Association's magazine and on social media, visits to support groups, and through physiotherapists specialized in the treatment of PD. Controls were recruited from partners and acquaintances of the participating patients and by advertisements on social media. The inclusion criteria were (1) aged  $\geq 30$  years, (2) in possession of a smartphone running on Android 4.4 or higher, and (3) living within travelling distance from the study center. Additional inclusion criteria for the PD group were (1) Parkinson disease diagnosed by a neurologist, (2) currently using levodopa and/or a dopamine agonist, (3) experiencing at least slight motor fluctuations (Movement Disorders Society Unified Parkinson's Disease Rating Scale [MDS-UPDRS] part IV item 4.3  $\geq 1$ ), and (4) experiencing at least some Parkinson-related gait impairments (MDS-UPDRS part II item 2.12  $\geq 1$  and/or item 2.13  $\geq 1$ ). The exclusion criteria were (1) any type of advanced treatment (deep brain stimulation or intestinal levodopa or apomorphine infusion) and (2) psychiatric or cognitive impairments that may hinder successful completion of the study protocol (based on judgement of the assessor running the recruitment). We did not exclude patients with PD or controls who used assistive devices or reported other medical problems affecting their movements. We refer to

[Multimedia Appendix 1](#) for a flow diagram of the inclusion procedure.

### Data Collection Procedure

Data were collected during single visits to the participants' homes from July 2017 to July 2018. A visit consisted of 2 parts: an unscripted free-living part and a standardized clinical assessment. In the PD group, both parts were conducted twice during the same visit, once in the morning after overnight withdrawal of dopaminergic medication (premedication) and once after intake of the patient's prescribed dopaminergic medication (postmedication).

During the free-living part, the assessors encouraged the participants to perform habitual activities in and around their house for at least one hour. In order to capture natural behavior, there was no script for this part of the visit. Instead, the assessors used a checklist to ensure that essential behaviors were captured, such as doing normal morning routines, preparing and having breakfast, walking indoors, walking in the neighborhood, walking up and down the stairs, sitting down, standing up, and doing some household chores. It should be noted that the participants' normal routines were leading; for example, if participants hardly did any household chores themselves, we did not ask them to do so during the visit.

The clinical assessments were also conducted in the participants' homes and included the TUG test, the Abnormal Involuntary Movement Scale, and the complete MDS-UPDRS, except for the self-reported items of part I and II, which were completed through an online survey after the visit. In the PD group, the clinical assessments were performed before and after medication intake, except for part I (assessor-rated items) and part IV of the MDS-UPDRS, which were only performed after medication intake. In addition, both the patient and the assessor scored the disease state (as off, on without dyskinesia, or on with dyskinesia) for every 30-minute epoch after medication intake. All assessments were conducted by a single assessor who had received appropriate prior training. [Multimedia Appendix 1](#) provides an overview of the events during the home visits.

During the full duration of the visit, participants wore lightweight sensors on both wrists, both ankles, the lower back, and in the front pants pocket, containing movement, contextual, and physiological sensors ([Table 1](#)). Except for the smartphone, all devices were attached using adjustable straps. In addition, the entire visits were recorded on video using a handheld high-definition video recorder by one of the assessors following the participant. To put the patient at ease and make the setting more naturalistic, the other assessor helped to comfort the patient and shift the focus from being recorded to performing habitual activities together. To allow for time synchronization between all sensors and the video recordings, all devices were triggered (hit simultaneously against the table 10 times) in front of the camera at the beginning and end of the recordings.

**Table 1.** Overview of the wearable sensors used during the study visits.

Device	Locations	Collected sensor data
Gait Up Physilog 4 <sup>a</sup>	Both ankles, both wrists, lower back (strap around waist)	Accelerometer, gyroscope, magnetometer, barometer
Android Wear smartwatch <sup>b</sup>	Wrist (PD <sup>c</sup> group: self-reported most affected side; control group: most comfortable side)	Accelerometer, gyroscope, barometer, light
Android smartphone <sup>d</sup>	Pants pocket (same side as Android Wear smartwatch)	Accelerometer, magnetometer, light, proximity, GPS, Wi-Fi, and cellular networks
Empatica E4 <sup>e</sup>	Wrist (opposite wrist to Android Wear smartwatch)	GSR <sup>f</sup> , PPG <sup>g</sup> , skin temperature, accelerometer

<sup>a</sup>Gait Up SA.<sup>b</sup>Motorola Moto 360 Sport (Motorola Inc) with custom application collecting raw sensor data.<sup>c</sup>PD: Parkinson disease.<sup>d</sup>Various models with the HopkinsPD app collecting raw sensor data.<sup>e</sup>Empatica Inc.<sup>f</sup>GSR: galvanic skin response.<sup>g</sup>PPG: photoplethysmogram.

After the home visits, all participants continued to use a subset of the sensors (smartphone and smartwatch) for 2 weeks and completed symptom diaries as reference. As this part of the data set is not used in the current analyses, we refer to [Multimedia Appendix 1](#) for a detailed description of the corresponding protocol.

The study protocol was approved by the local medical ethics committee (Commissie Mensgebonden Onderzoek, region Arnhem-Nijmegen, file number 2016-1776). All participants received verbal and written information about the study protocol and signed a consent form prior to participation, in line with the Declaration of Helsinki.

## Data Availability

The full data set as described in the “Methods” section, with the exception of the raw video recordings and absolute GPS coordinates, will be made available to the worldwide research community in collaboration with the Michael J Fox Foundation. To protect participant privacy, manual annotations are available from the video recordings, and the GPS data will be deidentified before sharing. A subset of the data is being used in the international Biomarker and Endpoint Assessment to Track Parkinson’s Disease (BEAT-PD) Challenge [27]. The full curated and deidentified data set will be released when this data challenge is completed. The specific data and analysis scripts that support the findings of this study are available from the corresponding author upon reasonable request.

## Data Processing

### Video Annotations

To provide ground truth labels for algorithm development and validation, the video recordings were annotated by trained research assistants for (1) the protocol structure (ie, when the clinical assessments and free-living parts were performed), (2) the occurrence of general behaviors during the free-living parts (such as standing, walking, and sitting), and (3) the presence and severity of tremor and the presence and manifestations of freezing of gait during the free-living parts. Annotations from the last category were checked by a physician with experience

in movement disorders. In addition, those items of the MDS-UPDRS part III that can be evaluated from video were assessed by a second, independent rater (physician with experience in movement disorders). The annotations were created using ELAN (The Language Archive), an open source program for creating annotations in video recordings [28]. For a detailed description of the video annotation protocol, we refer to [Multimedia Appendix 1](#).

### Sensor Data Preprocessing

In the current analysis, we used the triaxial accelerometer data (in m/s<sup>2</sup>) from all Physilog devices and the smartphone worn in the pants pocket. First, data were interpolated to a uniform sample rate of 120 Hz using piecewise cubic interpolation. Next, the effects of gravity were removed by applying an  $\ell_1$  trend filter to each of the 3 axes separately (MATLAB implementation by Kim et al [29]). For each device, we used the 3 resulting dynamic acceleration signals,  $a_x$ ,  $a_y$ , and  $a_z$ , to compute the magnitude of dynamic acceleration (ie, the square root of [ $a_x^2 + a_y^2 + a_z^2$ ]).

### Frequency Analysis of Gait Segments

Various approaches have been used to quantify the gait pattern in patients with PD using accelerometer data. Some studies rely on the detection of the initial and final contact of the feet, from which temporal gait features such as the step and swing time can be derived [30]. Based on the exact sensor positioning and some assumptions derived from the biomechanics of gait, location-specific algorithms can be used to estimate spatial gait features. For example, having identified the initial and final contact, one can use the inverted pendulum model to estimate the step length using a sensor on the lower back [31]. Although this approach produces outcome measures that directly relate to the way gait is evaluated by clinicians, its location dependency complicates the comparison of different sensor locations. Additionally, detecting the initial and final contact is more challenging in real-life circumstances [32]. Other studies analyze the periodicity of the accelerometer signal during gait based on the power spectral density (PSD) or autocorrelation



[21,33]. Since it does not rely on location-specific assumptions, we used this approach in our analysis.

The video annotations were used to locate periods of gait (defined as five or more consecutive steps) during the free-living parts. From these, we extracted nonoverlapping gait segments of equal length (3000 samples, corresponding to 25 seconds). This length was selected because prior research showed that using shorter free-living gait segments discriminated less well between patients with PD and controls [19] and in order to achieve sufficient resolution in the frequency domain. To be included in the analyses, participants needed to have at least 10 gait segments of 25 seconds. In addition, patients were required to have at least 5 segments before and 5 segments after medication intake. The PSD of each gait segment and sensor location was estimated using Welch's method [34] (with, per segment, Hamming windows of 1024 samples with approximately 50% overlap). From each of the 6 sensor locations, we extracted 4 different signal properties, resulting in 24 features. We computed the total power in the 0.5- to 10-Hz interval, which captures practically all contributions from human gait [35]. In addition, we extracted the frequency, height, and width (at half the height) of the dominant peak in the PSD. The frequency of the dominant peak was used to derive the participants' cadence (the method is described in [Multimedia Appendix 1](#)).

## Evaluation

For the main objective of this study, we compared real-life gait segments before and after medication intake. Since we expected that the relative differences within subjects are most relevant in this context, we normalized each feature using  $z$  scores based on each patient's mean and standard deviation. In addition, we compared gait segments between patients with PD and controls. For these analyses, we rescaled each feature using its between-subject standard deviation in order to aid the comparison of effect sizes between features and to obtain a common scale for regularization (see "Classification").

## Individual Features

First, we examined the effect of medication intake on the individual features. To deal with the varying number of gait segments between participants, we used linear mixed effect models with the normalized features as dependent variables. For each feature, we estimated a fixed effect of the timing of the gait segment (premedication/postmedication) and random intercepts and slopes per patient. We used separate linear mixed effect models to estimate fixed effects of the group (patients with PD premedication, postmedication, or controls) and random intercepts per participant. Because we aim to show the magnitude and spread of the individual fixed effects rather than testing overarching hypotheses based on multiple comparisons, we report the unadjusted 95% confidence intervals of the estimated effects, as recommended by Gelman et al [36].

## Classification

Next, we evaluated whether combinations of features could be used to predict whether a gait segment occurred before or after medication intake. For this, we used logistic LASSO (least absolute shrinkage and selection operator) regression with

uniform prior class probabilities. To account for the varying number of gait segments per patient, we weighted each gait segment by the inverse of the number of gait segments per patient. We evaluated the performance using leave-one-subject-out nested cross-validation (CV), with the LASSO regularization hyperparameter being selected in the inner CV loops. The main performance measure was the mean area under the receiver operator curve (AUC). In addition, we evaluated the performance of logistic LASSO regression to predict whether a gait segment was from a patient with PD (premedication) or control. We used nested cross-validation for this as well, leaving 1 patient and 1 control out in each fold. To avoid information leakage, normalization was performed separately for each fold using only training data.

For both problems, we trained one classifier for each sensor location and one for all sensor locations combined. To prevent the patient with PD or control classifiers from learning the differences between the dominant and nondominant side, we ensured that the proportion of participants with measurements from the dominant side was always equal in the PD and control group. We tested whether the AUCs of the individual sensor locations were lower than the AUC of all sensor locations combined using the Wilcoxon matched pairs signed rank test ( $P < .05$  considered statistically significant). The results of the individual comparisons were then used to test the overarching null hypothesis that using all sensor locations combined is superior to using any of the single sensor locations (we rejected this hypothesis if three or more individual comparisons were nonsignificant, which corresponds to a significance level  $\alpha$  of approximately .01; for the calculation, see [Multimedia Appendix 1](#)).

Using the trained classifiers of the sensor location with the highest AUC, we constructed continuous scores that could serve as digital biomarkers for the response to medication intake (premedication/postmedication classifier) and PD gait impairment severity (PD/control classifier). For this, we used the classifiers' decision values (linear combinations of feature values), which, in the case of logistic regression, correspond to the logarithm of the odds (logit) of the posterior class probabilities. We then examined their correlation with relevant clinical measures (Spearman  $\rho$ ). Specifically, using the "best" classifier for predicting premedication and postmedication in the PD group, the patients' mean changes in decision values after medication intake (corresponding to the patients' mean log odds ratio) were correlated to the changes in the TUG score and the subtotal of the items related to mobility of the MDS-UPDRS part III (items 3.9, 3.10, 3.11, 3.12, and 3.13). Using the same sensor location's classifier for predicting patient with PD or control, the patients' mean decision values were correlated to the absolute TUG score, time since diagnosis, and the subtotal of the MDS-UPDRS part III mobility items. All analyses were performed in MATLAB 2018a (MathWorks).

## Qualitative Evaluation

To explore which factors unrelated to PD may disturb gait analysis in real-world settings, we inspected individual patients and sensor locations in which the classifier performed worse than random classification in predicting premedication and

postmedication (AUC <0.5). This was done by visually identifying change points in the PSDs and evaluating the corresponding video recordings for potential explanations. If an identified factor was also present in the video recordings of another patient, we evaluated whether it had a similar impact on the PSD (regardless of the patient's AUC).

## Results

### Participant Characteristics

A minimum of 10 gait segments of 25 seconds were available in 18 patients with PD (median 46.5 segments, range 14-95) and 24 controls (median 31 segments, range 11-61). These participants' demographic and clinical characteristics are presented in [Table 2](#). Reasons for collecting an insufficient number of gait segments included rainy weather (n=4), a desire not to be filmed in the neighborhood (n=1), use of a wheelchair

for longer distances (n=1), fatigue (n=1), and technical problems with the video recordings (n=1). The included patients did not differ substantially from the excluded patients in terms of disease severity ([Multimedia Appendix 1](#)). In 9 participants, there were technical problems with the sensor worn on the lower back. Therefore, we excluded this sensor location from the analyses combining multiple sensor locations (we refer to the [Multimedia Appendix 1](#) for the results of the 33 patients for whom the lower back sensor data were available). In addition, technical problems caused data loss for 1 ankle sensor in 1 participant; this participant (control) was excluded from the analyses combining multiple sensor locations. The distribution of the length of the analyzed gait segments was similar among patients with PD (premedication and postmedication) and controls ([Multimedia Appendix 1](#)). The experiences of participants during the home visit collected by the online exit survey are included in [Multimedia Appendix 1](#).



**Table 2.** Demographics and clinical characteristics of patients included in the analyses.

Characteristic	Patients with PD <sup>a</sup> (n=18)	Controls (n=24)
Age (years), median (IQR)	65.0 (60.5-69.0)	67.5 (55.0-70.0)
Gender (men), n (%)	11 (61)	13 (54)
<b>Most affected side<sup>b</sup> and hand dominance, n (%)</b>		
Most affected=dominant	8 (44)	N/A <sup>c</sup>
Most affected=nondominant	8 (44)	N/A
Mixed handedness	2 (11)	N/A
Time since diagnosis of PD (years), median (IQR)	6.5 (4.8-10.3)	N/A
<b>Use of dopaminergic medication, n (%)</b>		
Levodopa only	6 (33)	N/A
Levodopa and dopamine agonist	10 (55)	N/A
Levodopa and MAO-B <sup>d</sup> inhibitor	1 (6)	N/A
Levodopa, dopamine agonist, and MAO-B inhibitor	1 (6)	N/A
<b>Hoehn &amp; Yahr stage, n (%)</b>		
Stage 1	1 (6)	N/A
Stage 2	13 (72)	N/A
Stage 3	4 (22)	N/A
<b>MDS-UPDRS<sup>e</sup> (scores on subscales), median (IQR)</b>		
Part I (scale range: 0 to 52)	9.5 (7.8-15.0)	3.0 (0.3-4.0)
Part II (scale range: 0 to 52)	11.0 (8.5-15.3)	0.0 (0.0-0.0) <sup>f</sup>
Part III (off state) (scale range: 0 to 132)	41.5 (31.5-57.8)	6.5 (4.3-11.0)
Part III (on state) (scale range: 0 to 132)	28.0 (18.5-38.0)	N/A
Part IV <sup>g</sup> (scale range: 0 to 24)	6.0 (4.5-9.3)	N/A
<b>AIMS<sup>h</sup> (scale range: 0 to 40), n (%)</b>		
0	13 (72)	20 (83)
1-3	2 (11)	3 (13)
>3	3 (17)	1 (4) <sup>i</sup>
<b>TUG<sup>j</sup> (median of 4 trials, in seconds), median (IQR)</b>		
Off state	12.0 (11.3-13.7)	10.0 (9.3-10.8)
On state	11.4 (9.7-12.4)	N/A
<b>Falls in last 12 months, n (%)</b>		
0	10 (56)	20 (83)
1-2	8 (44)	4 (17)
<b>Freezing episodes in last month, n (%)</b>		
0	13 (72)	N/A
1 or more	5 (28)	N/A

<sup>a</sup>PD: Parkinson disease.<sup>b</sup>Most affected side: side where the PD symptoms are most severe, as reported by the patients.<sup>c</sup>N/A: not applicable.<sup>d</sup>MAO-B: monamine oxidase B.<sup>e</sup>MDS-UPDRS: Movement Disorders Society Unified Parkinson's Disease Rating Scale.<sup>f</sup>1 missing value.

<sup>g</sup>Specific to PD: side effects of dopaminergic medication.

<sup>h</sup>AIMS: Abnormal Involuntary Movement Scale.

<sup>i</sup>This participant demonstrated facial synkinesis during the assessment.

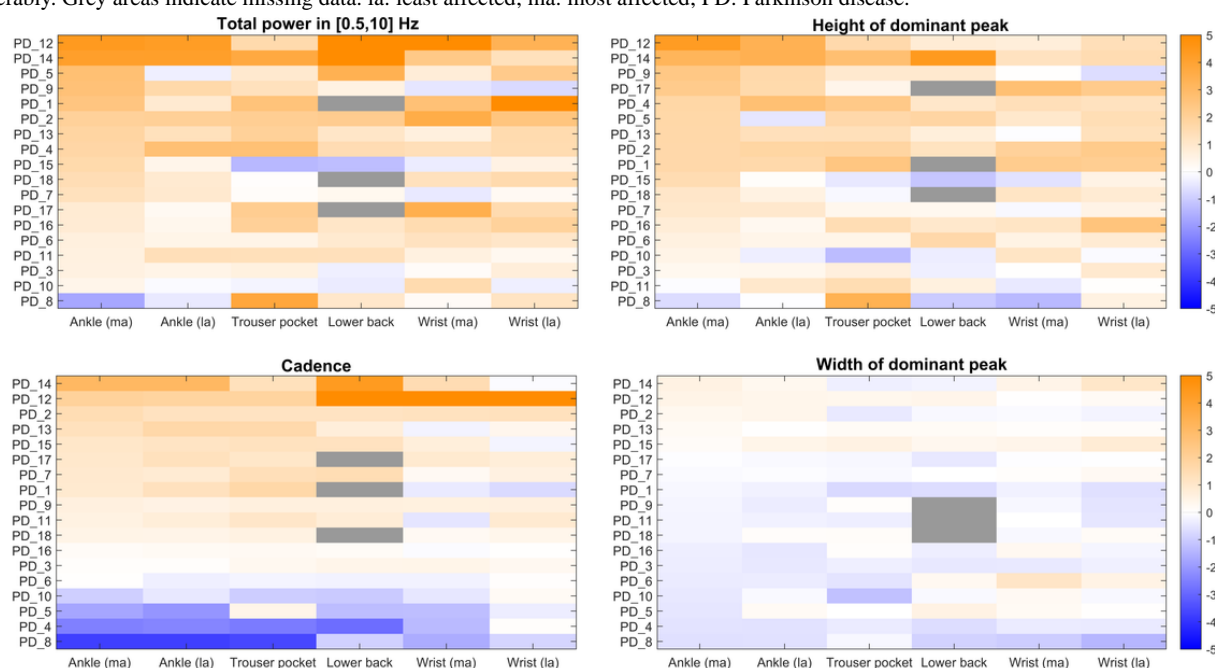
<sup>j</sup>TUG: timed up and go test.

## Comparison of Before and After Medication Intake

Figure 1 shows all individual changes in the signal properties after medication intake. For illustration, Figure 2 displays the power spectral density of one of the patients who demonstrated a clear response to medication intake (PD\_14). At the group level, the total power and height of the dominant peak increased in all sensor locations after medication intake, whereas the width

of the dominant peak did not change significantly in any sensor location (Figure 3). Cadence increased in some patients but markedly decreased in others, resulting in a nonsignificant change at the group level for all sensor locations (Figure 3). Because of the high correlation between changes in the height of the dominant peak and total power (mean Pearson  $r$  of 0.80) and the clearer response of the latter across the sensor locations, the height was not included in subsequent analyses.

**Figure 1.** Changes in the 4 signal properties after medication intake, expressed in z scores (color bar). The x-axis displays the various sensor locations; the y-axis shows all individual patients (sorted on the values of the ankle sensor of the most affected side). The figure highlights that the total power and height of the dominant peak increase after medication intake in most patients (with considerable variation between sensor locations), the cadence increases in some but decreases in others (with high agreement between sensor locations), and the width of the dominant peak does not change considerably. Grey areas indicate missing data. la: least affected; ma: most affected; PD: Parkinson disease.



**Figure 2.** Visualization of the PSD of one of the patients who shows a clear response after medication intake (PD\_14). x-axis: nth gait segment of 25 seconds (the white line indicates intake of dopaminergic medication). y-axis: frequency (Hertz). Color bar: PSD in  $(\text{m/s}^2)^2/\text{Hz}$ . PD: Parkinson disease; PSD: power spectral density.

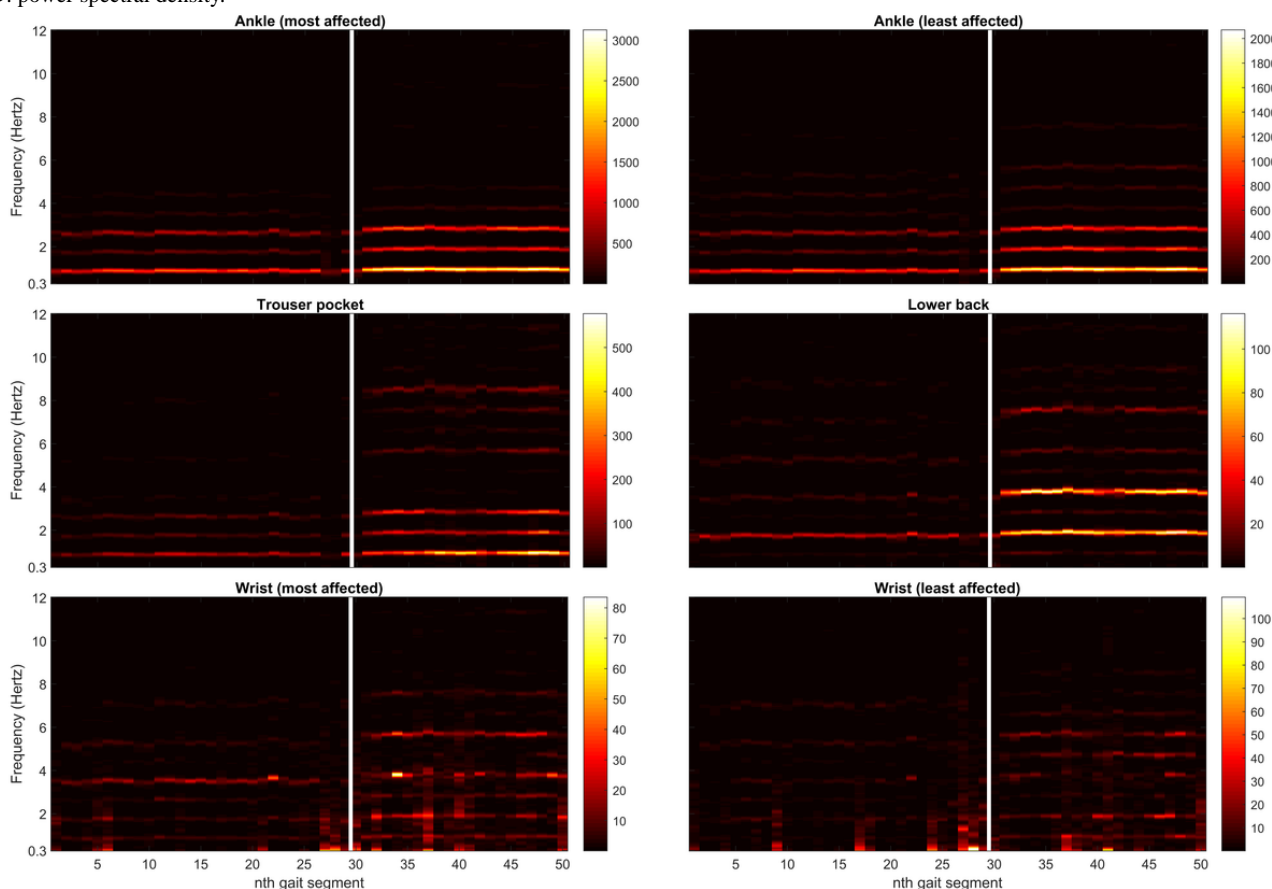


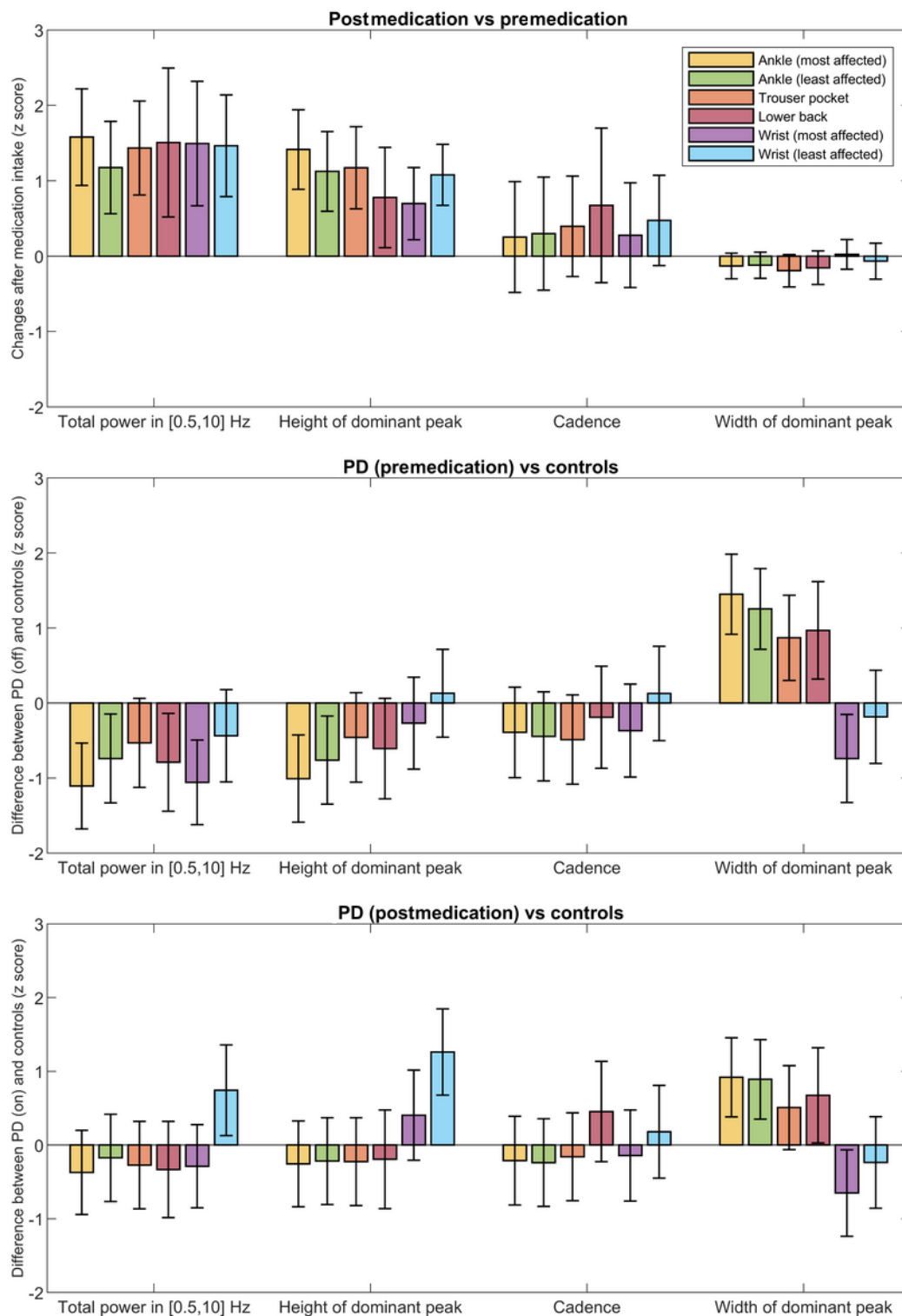
Figure 4 displays the receiver operator curves (ROCs) of the logistic classifiers trained using the various sensor locations and signal properties. The mean AUC for all sensor locations and signal properties combined was 0.84 (95% CI 0.75-0.93). The mean AUCs of the ankle of the most affected side (0.82, 95% CI 0.72-0.92), the wrist of the most affected side (0.76, 95% CI 0.66-0.87), and the wrist of the least affected side (0.79, 95% CI 0.69-0.88) were not significantly worse than using all sensor locations combined. Based on this, we reject the overarching hypothesis that using all sensor locations combined is superior to using any of the single sensor locations. Of the signal properties, the total power was most informative to discriminate between before and after medication intake, with a mean AUC of 0.80 (95% CI 0.69-0.91). The mean AUC of the width of the dominant peak (0.51, 95% CI 0.45-0.58) and cadence (0.55, 95% CI 0.43-0.66) were by themselves not significantly different from 0.5 (random classifier), although the width of the dominant peak resulted in a modest (but statistically significant) improvement in AUC when combined with the total power (Figure 4). The AUC and balanced accuracy

(based on an equal class prior) of all locations and signal properties are presented in Table 3. Multimedia Appendix 1 includes a sensitivity analysis on the effects of the choice of per-subject normalization.

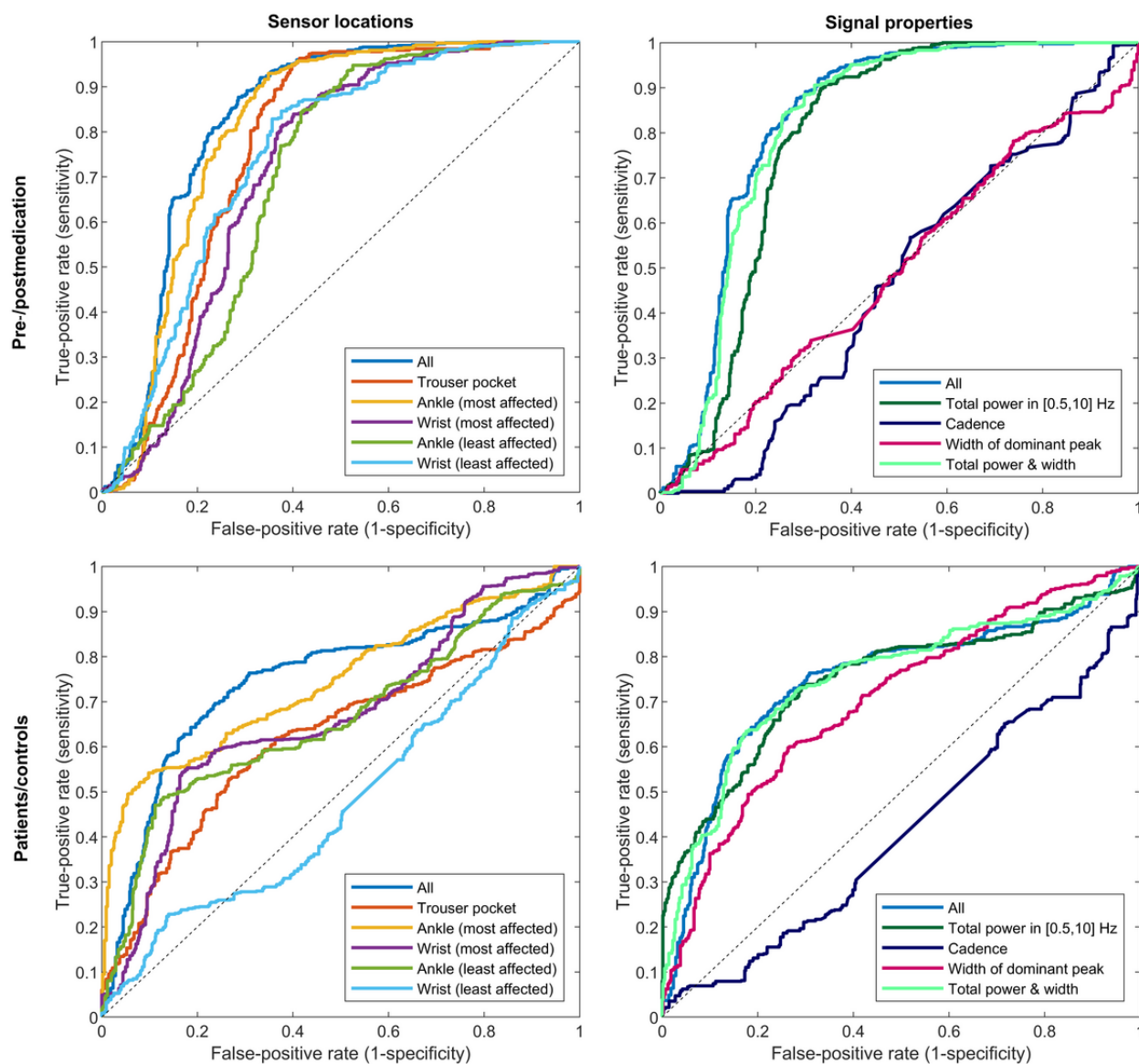
Both the TUG (−1.51, 95% CI −2.29 to −0.73) and the subtotal of the MDS-UPDRS part III mobility items (−1.31, 95% CI −2.13 to −0.48) decreased after medication intake. However, the patients' mean changes in decision values, obtained from the classifier based on the ankle sensor of the most affected side, did not correlate with changes in the TUG ( $p=0.02$ , 95% CI −0.49 to 0.45), nor with changes in the subtotal of the MDS-UPDRS part III mobility items ( $p=0.11$ , 95% CI −0.38 to 0.45).

Based on the video recordings, freezing of gait (FOG) was observed in 2 patients (PD\_1 and PD\_4). However, no FOG episodes occurred during any of the gait segments of at least 25 seconds, so FOG is not expected to have influenced the observed changes.

**Figure 3.** Top: changes after medication intake on a group level (mean and 95% CI). Middle: differences between patients with PD (premedication) and controls (mean and 95% CI, positive means higher in patients). Bottom: differences between patients with PD (postmedication) and controls (mean and 95% CI, positive means higher in patients). All estimates are based on linear mixed effects models for each sensor location and signal property. PD: Parkinson disease.



**Figure 4.** Receiver operating characteristic curves of the logistic classifiers, averaged over the cross-validation folds. Top half: premedication and postmedication classification. Bottom half: patients and controls classification. Left half: comparison between different sensor locations. Right half: comparison between different signal properties, based on all sensor locations combined.





**Table 3.** Performance of the logistic classifiers (mean, SE 1.96 over the cross-validation folds). Accuracies are based on the optimal classifier for each fold with equal misclassification costs and equal class prior (also referred to as balanced accuracy).

Feature set	Premedication/postmedication, mean (95% CI)		Patients with PD <sup>a</sup> (premedication) versus controls, mean (95% CI)	
	AUC <sup>b</sup>	Accuracy	AUC	Accuracy
All	0.84 (0.75-0.93)	0.79 (0.71-0.87)	0.76 (0.62-0.90)	0.72 (0.61-0.83)
<b>Sensor locations</b>				
Pants pocket	0.78 (0.67-0.89)	0.75 (0.67-0.83)	0.62 (0.46-0.78)	0.63 (0.51-0.74)
Ankle (most affected)	0.82 (0.72-0.92)	0.77 (0.71-0.84)	0.74 (0.58-0.90)	0.66 (0.54-0.79)
Wrist (most affected)	0.76 (0.66-0.87)	0.72 (0.64-0.81)	0.75 (0.62-0.88)	0.57 (0.50-0.65)
Ankle (least affected)	0.70 (0.60-0.81)	0.68 (0.60-0.76)	0.70 (0.53-0.86)	0.62 (0.50-0.74)
Wrist (least affected)	0.79 (0.69-0.88)	0.74 (0.67-0.81)	0.49 (0.38-0.61)	0.48 (0.40-0.55)
<b>Signal properties</b>				
Total power in 0.5-10 Hz	0.80 (0.69-0.91)	0.77 (0.70-0.85)	0.74 (0.60-0.89)	0.72 (0.61-0.82)
Cadence	0.55 (0.43-0.66)	0.52 (0.44-0.60)	0.41 (0.32-0.50)	0.45 (0.38-0.52)
Width of dominant peak	0.51 (0.45-0.58)	0.50 (0.45-0.55)	0.71 (0.63-0.78)	0.66 (0.60-0.73)
Total power and width	0.83 (0.73-0.93)	0.78 (0.71-0.85)	0.76 (0.62-0.90)	0.72 (0.61-0.82)

<sup>a</sup>PD: Parkinson disease.<sup>b</sup>AUC: area under the receiver operator curve.

## Comparison of Patients and Controls

The width of the dominant peak differed between patients and controls in most sensor locations, both premedication and postmedication ([Figure 3](#)). Interestingly, the PD patients' width was elevated in the ankle, pants, and lower back sensors but lowered in the wrist of the most affected side. The estimated cadence did not differ between patients (pre- and postmedication) and controls, regardless of the sensor location used. The total power and height of the dominant peak were lower in patients with PD (premedication) compared with controls in most sensor locations, but this difference was not present after medication intake, with the exception of the wrist of the least affected side, which showed elevated values after medication intake and no differences before medication intake. The ROCs for the various sensor locations and signal properties ([Figure 4](#)) show that it was generally more difficult to discriminate between the gait of patients with PD and controls than between before and after medication intake within patients with PD. The AUC and balanced accuracy of all locations and signal properties are presented in [Table 3](#).

The patients' decision values obtained by the classifier based on the ankle of the most affected side demonstrated low to moderate correlations with the time since diagnosis ( $p=0.55$ , 95% CI 0.11 to 0.78), the absolute subtotal of the MDS-UPDRS part III mobility items ( $p=0.41$ , 95% CI -0.07 to 0.81), and the absolute TUG score ( $p=0.33$ , 95% CI -0.16 to 0.69).

## Impact of Real-world Factors

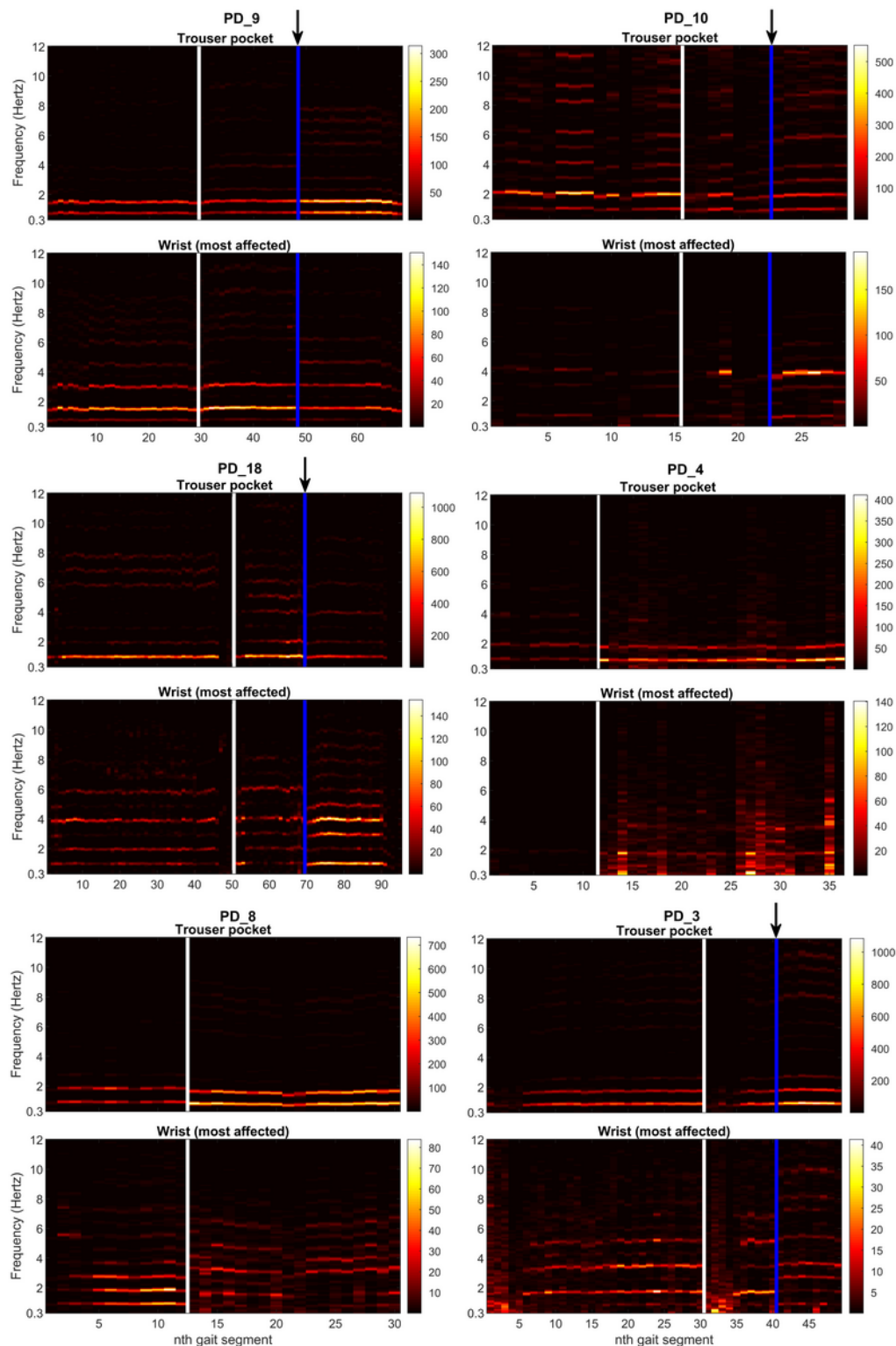
We observed considerable variation in the premedication and postmedication classification performance (AUC) between

individual patients ([Multimedia Appendix 1](#)). This may be partly explained by variation in the strength of the response to dopaminergic medication. In addition, real-world factors unrelated to PD may negatively impact the classification performance in uncontrolled settings. By inspecting the video recordings and PSDs, we identified various factors that may have influenced the PSD of devices worn on the wrist and in the pants pocket.

## Wrist-Worn Devices

For the wrist-worn devices, the positioning of the hands (eg, in or out of the pocket of the pants or jacket) appeared to have a significant impact on the PSD during gait. To explore this, we evaluated the video recordings and PSDs of all patients ( $n=5$ ) who changed the position of their hands within either the premedication or postmedication free-living part (to be able to rule out that the observed difference was caused by the effects of medication). In all 5 patients, we found an association in time between changes in the position of the hands and changes in the PSD (examples are shown in [Figure 5](#)). One patient (PD\_9) changed the position of his hands from the pants pocket to the pocket of his jacket, which was associated with changes in the PSD. In the other 4 patients, the nature of the changes in the PSD was associated with the amplitude of the arm swing; in one patient with a pronounced arm swing (based on the video recordings), the total power clearly increased when removing the hand from the pocket (PD\_10). In 3 patients with a reduced arm swing, the contribution from the harmonic frequency components changed without an increase or even with a decrease in the total power (PD\_3, PD\_11, and PD\_18).

**Figure 5.** Visualization of the PSD of the devices worn on the wrist (most affected side) and in the pants pocket of 6 illustrative patients. x-axis: nth gait segment of 25 seconds (the white line indicates intake of dopaminergic medication). y-axis: frequency (hertz). Color bar: PSD in  $(\text{m/s}^2)^2/\text{Hz}$ . PD\_9: blue line with arrow indicates when patient moved his hand from his pants pocket to the pocket of his jacket (during the premedication part, his hand was in his pants pocket). PD\_10: blue line with arrow indicates when patient removed her hand from the pocket of her jacket (during the premedication part, her hand was outside the pocket). PD\_18: blue line with arrow indicates when patient put his hand in his pants pocket (during the premedication part, his hand was outside the pocket). PD\_4: after medication intake, the patient presented with choreic dyskinesias in both arms and legs, which were most severe on his most affected side. PD\_8: before medication intake, the arm swing on his most affected side was practically absent. After medication intake, he showed an arm swing with a small amplitude and started to make occasional gestures. PD\_3: before medication intake, the arm swing on his most affected side was practically absent (hand was outside the pocket). After medication intake, first he had his hand in the pocket of his jacket. Blue line with arrow: patient removed his hand from the pocket and demonstrated an arm swing with a small amplitude, and he started to make occasional gestures.



The effect of an (almost) absent arm swing in patients with PD may closely resemble the effect of putting the hand in the pants pocket. This is illustrated by patients PD\_3 and PD\_8 (Figure 5). Before medication intake, their arm swing was practically absent, while after medication intake, these patients showed an arm swing with a small amplitude and started to make occasional gestures (eg, point at things). Similar to the patients who removed their hands from their pockets and demonstrated a modest arm swing, the contribution from the harmonic frequency components changed without an increase in the total power (Figure 5).

The video recordings provided no explanation for the low performance of the wrist-worn sensor (most affected side) in PD\_7 and PD\_15.

### **Pants Pocket**

The positioning of the hands also appeared to be relevant when using the pants pocket as a sensor location. Two patients removed their hand from the pants pocket containing the smartphone during one of the free-living parts (PD\_9 and PD\_18). In both patients, this was not only reflected in the PSD of the wrist-worn device, but also in the PSD of the smartphone (Figure 5). Given the observed changes, we hypothesize that the presence of the hands in the pants pocket reduces the total power measured by the smartphone, which might be explained by reduced freedom of movement of the smartphone. There were other patients with low classification performance in the smartphone (PD\_10 and PD\_15), but the video recordings provided no explanation.

## **Discussion**

### **Principal Findings**

This work presents a new reference data set for the development of wearable sensor algorithms that can passively monitor persons with Parkinson disease in daily life. The novelty of this data set consists of (1) the inclusion of unscripted daily activities in and around the participants' own homes in combination with a simultaneous video reference that was later used to annotate specific activities and symptoms, (2) the combination of multiple movement sensors positioned on 6 different body locations and 8 different contextual and physiological sensors (eg, GPS, light, photoplethysmogram), (3) objective monitoring of these everyday activities during both a practically defined off state and a subjective optimal on state, allowing for testing of responsiveness, (4) the addition of a subsequent longitudinal follow-up with a more limited set of sensors on 2 body positions for 2 weeks, and (5) the principle of data sharing, such that this rich and versatile data set will be made openly available to the scientific community.

We have also explored this data set ourselves, with the main aim of assessing whether analysis of free-living gait can be used to obtain a digital biomarker for motor fluctuations in PD. We demonstrated that, despite the natural variation that is inevitably present in unscripted gait, it is feasible to measure the response to dopaminergic medication using simple signal properties derived from the PSD of accelerometer signals. Moreover, we present the first simultaneous evaluation of multiple sensor

locations in this context. This is a highly relevant issue in the field of wearable sensors, where much uncertainty exists about the ideal sensor placement to obtain the most sensitive and reliable markers for any specific sign of PD [26]. Our results indicate that the effects of dopaminergic medication on gait can be detected in all tested sensor locations (ie, both ankles, both wrists, the lower back, and the pants pocket). Combining multiple sensor locations did not improve classification significantly, suggesting that 1 individual sensor is a reasonable setup, from both an algorithmic and a usability perspective, to monitor on-off changes in free-living gait. Based on the used signal properties, it was generally more difficult to distinguish between patients with PD and age-matched controls than between premedication and postmedication among patients with PD. This may be explained by the fact that, in addition to the presence of PD, other factors also influence a person's gait pattern (eg, idiosyncrasies, age, comorbidities). When studying within-subject changes, many of these factors remain stable and hence do not introduce variability that is unrelated to PD. Finally, the qualitative analysis using the video recordings and PSDs revealed that the positioning of the hands has a significant impact on the PSD of both the wrist and the pants pocket in real-world settings, highlighting potential points of improvement for future research. Note that we illustrate this approach here for gait, but the data set lends itself well to future analyses by other groups who may want to study the optimal composition of sensor types and positions for a wide range of other symptoms or unscripted daily activities.

### **Limitations**

Before sensor-based gait analysis can be recommended to clinicians and researchers as a tool to assess therapy responses in real life, it is important to consider this study's limitations. First, we only included patients with at least some motor fluctuations and PD-related gait impairments. Although motor fluctuators may also be the main target group for such an objective evaluation tool, the generalizability to fluctuators who do not report any gait impairments remains to be evaluated. Second, differences in gait pattern between on and off periods might be less pronounced in daily life because most patients do not withhold their medication for as long as they did in this study. We purposely included a practically defined off assessment (ie, after overnight withdrawal of all dopaminergic medication) to assess which sensor locations and signal properties would be most responsive. Nevertheless, our findings remain to be confirmed on truly real-life data, where fluctuations can be more subtle. We should also note that, for practical reasons, we could not randomize the order of the premedication and postmedication parts (the premedication session was always performed first), so we cannot rule out that fatigue influenced the measured postmedication performance (although we minimized its effects by including a break between both parts). Third, the presented assessment of motor fluctuations depends on the occurrence of gait segments of at least 25 seconds. This limits the potential target group to patients who are able to walk during both on and off periods, excluding patients with Hoehn & Yahr stage 5. However, most fluctuators with a lower Hoehn & Yahr stage can still walk when the medication effects have worn off [37]. Finally, the presented analysis depends on

accurate localization of gait segments in time. In this study, we used the available video annotations for this, which enabled us to focus on how well the different sensor locations capture on-off changes in the gait patterns without introducing additional variability in how well the different sensor locations can detect gait. Several methods have been proposed for the detection of gait based on accelerometer data, with accuracies varying across sensor locations and techniques [38,39]. We are currently working on the implementation and evaluation of a novel probabilistic gait segmentation framework, which localizes gait segments in time with stationary (periodic) behavior [40].

### Comparison With Other Strategies

In contrast to the analysis of a specific behavior, such as gait, others have proposed activity-independent algorithms to monitor motor fluctuations in daily life. These algorithms typically assume that whenever the patient is wearing the sensor, it is possible to estimate the disease state. For example, Hammerla et al [8] used deep learning to predict the patient's motor state (ie, asleep, off, on without dyskinesias, or on with dyskinesias) every 5 minutes based on the accelerometer signal from 2 wrist-worn devices. The resulting performance was suboptimal (sensitivity for off detection: 0.50; for on detection: 0.52). This suggests that it may be too ambitious to predict the patients' motor state regardless of the activities performed.

Another consideration is the interpretability of such activity-independent algorithms. Because it is unclear what the algorithms' predictions are based on, it is challenging to relate the outcomes to clinical impairments. This might form a barrier to use in clinical practice and makes it difficult to form hypotheses about potential covariates and confounders in daily life. In our approach, it is transparent that the outcomes relate to (changes in) the gait pattern. This means that we can predict circumstances in which the algorithm will fail (eg, when a patient only uses a wheelchair to travel longer distances) and that we can use our knowledge about gait to interpret the results.

### Clinical Interpretation

Our results show that the total power in the 0.5- to 10-Hz interval of the PSD during gait increases in response to dopaminergic medication, whereas the cadence increases in some but decreases in other patients. The width of the dominant peak does not change. The finding related to cadence contrasts with earlier lab-based findings showing an increase in cadence after intake of levodopa [14]. We demonstrated that, because of its marked variability between patients in real life, cadence is not a useful signal property for patient-independent algorithms. A likely explanation for this variability is that the ability to modulate cadence is intact in persons with PD, and some patients use this as compensation for reduced step length [41].

Whereas the clinical meaning of cadence is clear, the clinical meaning of the total power and width of the dominant peak are less straightforward. Sama et al [24] refer to the total power in the 0- to 10-Hz interval as the "fluidity of patients' movements during gait." However, the most pronounced clinical changes after medication intake are an increase in step length and arm swing [14,15]. Since it is reasonable to assume that an increased

step length and arm swing are reflected in an increased power of acceleration, the total power in the 0.5- to 10-Hz interval is most likely sensitive to these changes in gait pattern. The finding that the total power of the least affected arm is elevated in patients with PD after medication intake in comparison with controls may be explained by excessive movements (including dyskinesias) caused by dopaminergic medication. The width of the dominant peak in the PSD has been referred to as a measure for the inconsistency or variability of gait [21,42]. Weiss et al [21] showed that it predicts the fall risk in patients with PD. Interestingly, although the width of the dominant peak did not change after medication intake in this study, it was elevated in patients with PD compared with controls for the sensors worn on both ankles, the pants pocket, and the lower back. This might be explained by the insensitivity of balance impairments to levodopa [15]. The finding that the width of the dominant peak is lowered in the wrist of the most affected side, pointing to less variability, can possibly be attributed to reduced or absent arm swing and gestures, resulting in a less variable PSD (which matches with our observations, described in "Impact of Real-world Factors").

We observed absent or weak correlations between the sensor-based predictions and currently used clinical assessments (TUG and MDS-UPDRS part III mobility items). There are multiple explanations for this. The measurement modality is different (for the TUG, it is time versus acceleration-based measures), and the measured construct is related but not identical (for the TUG, postural transitions and gait versus only gait). Perhaps the most important difference is the measurement setting. There is growing evidence that following instructions in a standardized track results in different gait patterns compared with walking freely in a natural environment [16,18-20]. It can be argued that measurements in daily life are more valid, if the aim is to know how the patient performs in the presence of real-life facilitators and challenges. Since free-living assessments are inherently different from clinical assessments, alternative methods to establish validity may be more appropriate. The predictive power (eg, the ability to detect the effect of dopaminergic medication) and, in the end, the clinical utility (eg, whether the information helps clinicians to provide better treatments) might serve as reasonable alternatives.

### Future Directions

The combined inspection of the video recordings and PSDs highlighted some factors other than PD that may impact the (within-subject changes in the) gait pattern in real life, such as putting the hands in the pocket and making gestures. This part of the study was exploratory, and the effects of these factors need to be examined quantitatively in independent experiments. Other potential time-varying factors include voluntary changes in the walking pace [12], the type of shoes worn, carrying objects, the ground type, the location (eg, crowded city versus countryside), and other factors in the physical environment [43]. This area of research has not received much attention so far, largely due to a lack of available data sets. However, controlling for important covariates and confounders could be a significant step forward in the analysis of free-living gait. The inclusion of multiple sensor types that measure context in this study (eg, GPS, Wi-Fi, barometer) and the emergence of smart home



sensors [22,44] provide an opportunity to further examine the added value of a multimodal gait pattern analysis. Finally, real-life gait analysis in patients with PD may not only be useful to monitor motor fluctuations but could also yield much-needed digital biomarkers to quantify disease progression in clinical trials testing new disease-modifying therapies. Future studies

that include long-term passive monitoring with wearable sensors (such as the Personalized Parkinson Project [45]) will reveal whether the promising results of lab-based gait analysis in this context [12,13] can be further improved by using highly frequent measurements obtained in the patients' natural environment.

## Acknowledgments

The authors wish to thank all participants for their enthusiasm to contribute to this study and for welcoming us in their own homes. We are grateful to the Michael J Fox Foundation for Parkinson's Research, UCB Biopharma, Stichting ParkinsonFonds, UCB, the Dutch Ministry of Economic Affairs (PPP Allowance made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships), and ZonMw (91215076 "Big Data for Personalised Medicine") for financially supporting both the data collection and analyses efforts of the Parkinson@Home validation study. We also thank Tim Hahn for his help with setting up and executing the study, as well as all involved research assistants and physicians for creating the detailed annotations of the video recordings.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Additional information and analyses.

[PDF File (Adobe PDF File), 4370 KB - [jmir\\_v22i10e19068\\_app1.pdf](#)]

## References

1. Jankovic J. Motor fluctuations and dyskinesias in Parkinson's disease: clinical manifestations. *Mov Disord* 2005;20 Suppl 11:S11-S16. [doi: [10.1002/mds.20458](#)] [Medline: [15822109](#)]
2. Hauser RA, Deckers F, Leher P. Parkinson's disease home diary: further validation and implications for clinical trials. *Mov Disord* 2004 Dec;19(12):1409-1413. [doi: [10.1002/mds.20248](#)] [Medline: [15390057](#)]
3. Palmerini L, Mellone S, Avanzolini G, Valzania F, Chiari L. Quantification of motor impairment in Parkinson's disease using an instrumented timed up and go test. *IEEE Trans Neural Syst Rehabil Eng* 2013 Jul;21(4):664-673. [doi: [10.1109/TNSRE.2012.2236577](#)] [Medline: [23292821](#)]
4. Sánchez-Ferro Á, Elshehabi M, Godinho C, Salkovic D, Hobert MA, Domingos J, et al. New methods for the assessment of Parkinson's disease (2005 to 2015): A systematic review. *Mov Disord* 2016 Sep;31(9):1283-1292. [doi: [10.1002/mds.26723](#)] [Medline: [27430969](#)]
5. Zhan A, Mohan S, Tarolli C, Schneider RB, Adams JL, Sharma S, et al. Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity: The Mobile Parkinson Disease Score. *JAMA Neurol* 2018 Jul 01;75(7):876-880 [FREE Full text] [doi: [10.1001/jamaneurol.2018.0809](#)] [Medline: [29582075](#)]
6. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. *NPJ Digit Med* 2020;3:21 [FREE Full text] [doi: [10.1038/s41746-020-0224-8](#)] [Medline: [32128451](#)]
7. Silva de Lima AL, Hahn T, Evers LJW, de Vries NM, Cohen E, Afek M, et al. Feasibility of large-scale deployment of multiple wearable sensors in Parkinson's disease. *PLoS One* 2017;12(12):e0189161 [FREE Full text] [doi: [10.1371/journal.pone.0189161](#)] [Medline: [29261709](#)]
8. Hammerla NY, Fisher JM, Andras P, Rochester L, Walker R, Plotz T. PD disease state assessment in naturalistic environments using deep learning. 2015 Presented at: Twenty-Ninth AAAI Conference on Artificial Intelligence; Jan 25-30, 2015; Austin, TX.
9. Roy SH, Cole BT, Gilmore LD, De Luca CJ, Thomas CA, Saint-Hilaire MM, et al. High-resolution tracking of motor disorders in Parkinson's disease during unconstrained activity. *Mov Disord* 2013 Jul;28(8):1080-1087 [FREE Full text] [doi: [10.1002/mds.25391](#)] [Medline: [23520058](#)]
10. Pulliam CL, Heldman DA, Brokaw EB, Mera TO, Mari ZK, Burack MA. Continuous Assessment of Levodopa Response in Parkinson's Disease Using Wearable Motion Sensors. *IEEE Trans Biomed Eng* 2018 Jan;65(1):159-164 [FREE Full text] [doi: [10.1109/TBME.2017.2697764](#)] [Medline: [28459677](#)]
11. Mirelman A, Gurevich T, Giladi N, Bar-Shira A, Orr-Urtreger A, Hausdorff JM. Gait alterations in healthy carriers of the LRRK2 G2019S mutation. *Ann Neurol* 2011 Jan;69(1):193-197. [doi: [10.1002/ana.22165](#)] [Medline: [21280089](#)]



12. Hobert MA, Nussbaum S, Heger T, Berg D, Maetzler W, Heinzel S. Progressive Gait Deficits in Parkinson's Disease: A Wearable-Based Biannual 5-Year Prospective Study. *Front Aging Neurosci* 2019;11:22 [FREE Full text] [doi: [10.3389/fnagi.2019.00022](https://doi.org/10.3389/fnagi.2019.00022)] [Medline: [30814947](https://pubmed.ncbi.nlm.nih.gov/30814947/)]
13. Schlachetzki JCM, Barth J, Marxreiter F, Gossler J, Kohl Z, Reinfelder S, et al. Wearable sensors objectively measure gait parameters in Parkinson's disease. *PLoS One* 2017;12(10):e0183989 [FREE Full text] [doi: [10.1371/journal.pone.0183989](https://doi.org/10.1371/journal.pone.0183989)] [Medline: [29020012](https://pubmed.ncbi.nlm.nih.gov/29020012/)]
14. Curtze C, Nutt JG, Carlson-Kuhta P, Mancini M, Horak FB. Levodopa Is a Double-Edged Sword for Balance and Gait in People With Parkinson's Disease. *Mov Disord* 2015 Sep;30(10):1361-1370 [FREE Full text] [doi: [10.1002/mds.26269](https://doi.org/10.1002/mds.26269)] [Medline: [26095928](https://pubmed.ncbi.nlm.nih.gov/26095928/)]
15. Smulders K, Dale ML, Carlson-Kuhta P, Nutt JG, Horak FB. Pharmacological treatment in Parkinson's disease: Effects on gait. *Parkinsonism Relat Disord* 2016 Oct;31:3-13 [FREE Full text] [doi: [10.1016/j.parkreldis.2016.07.006](https://doi.org/10.1016/j.parkreldis.2016.07.006)] [Medline: [27461783](https://pubmed.ncbi.nlm.nih.gov/27461783/)]
16. Warmerdam E, Hausdorff JM, Atrsaei A, Zhou Y, Mirelman A, Aminian K, et al. Long-term unsupervised mobility assessment in movement disorders. *Lancet Neurol* 2020 May;19(5):462-470. [doi: [10.1016/S1474-4422\(19\)30397-7](https://doi.org/10.1016/S1474-4422(19)30397-7)] [Medline: [32059811](https://pubmed.ncbi.nlm.nih.gov/32059811/)]
17. Nonnekes J, Ružicka E, Nieuwboer A, Hallett M, Fasano A, Bloem BR. Compensation Strategies for Gait Impairments in Parkinson Disease: A Review. *JAMA Neurol* 2019 Jun 01;76(6):718-725. [doi: [10.1001/jamaneurol.2019.0033](https://doi.org/10.1001/jamaneurol.2019.0033)] [Medline: [30907948](https://pubmed.ncbi.nlm.nih.gov/30907948/)]
18. Robles-García V, Corral-Bergantiños Y, Espinosa N, Jácome MA, García-Sancho C, Cudeiro J, et al. Spatiotemporal Gait Patterns During Overt and Covert Evaluation in Patients With Parkinson's Disease and Healthy Subjects: Is There a Hawthorne Effect? *J Appl Biomech* 2015 Jun;31(3):189-194. [doi: [10.1123/jab.2013-0319](https://doi.org/10.1123/jab.2013-0319)] [Medline: [25536440](https://pubmed.ncbi.nlm.nih.gov/25536440/)]
19. Del Din S, Godfrey A, Galna B, Lord S, Rochester L. Free-living gait characteristics in ageing and Parkinson's disease: impact of environment and ambulatory bout length. *J Neuroeng Rehabil* 2016 May 12;13(1):46 [FREE Full text] [doi: [10.1186/s12984-016-0154-5](https://doi.org/10.1186/s12984-016-0154-5)] [Medline: [27175731](https://pubmed.ncbi.nlm.nih.gov/27175731/)]
20. Brodie MAD, Coppens MJM, Lord SR, Lovell NH, Gschwind YJ, Redmond SJ, et al. Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different. *Med Biol Eng Comput* 2015 Aug 6;54(4):663-674. [doi: [10.1007/s11517-015-1357-9](https://doi.org/10.1007/s11517-015-1357-9)]
21. Weiss A, Herman T, Giladi N, Hausdorff JM. Objective assessment of fall risk in Parkinson's disease using a body-fixed sensor worn for 3 days. *PLoS One* 2014;9(5):e96675 [FREE Full text] [doi: [10.1371/journal.pone.0096675](https://doi.org/10.1371/journal.pone.0096675)] [Medline: [24801889](https://pubmed.ncbi.nlm.nih.gov/24801889/)]
22. Kabelac Z, Tarolli CG, Snyder C, Feldman B, Glidden A, Hsu CY, et al. Passive Monitoring at Home: A Pilot Study in Parkinson Disease. *Digit Biomark* 2019;3(1):22-30 [FREE Full text] [doi: [10.1159/000498922](https://doi.org/10.1159/000498922)] [Medline: [32095766](https://pubmed.ncbi.nlm.nih.gov/32095766/)]
23. Moore ST, MacDougall HG, Gracies J, Cohen HS, Ondo WG. Long-term monitoring of gait in Parkinson's disease. *Gait Posture* 2007 Jul;26(2):200-207. [doi: [10.1016/j.gaitpost.2006.09.011](https://doi.org/10.1016/j.gaitpost.2006.09.011)] [Medline: [17046261](https://pubmed.ncbi.nlm.nih.gov/17046261/)]
24. Samà A, Pérez-López C, Rodríguez-Martín D, Català A, Moreno-Aróstegui JM, Cabestany J, et al. Estimating bradykinesia severity in Parkinson's disease by analysing gait through a waist-worn sensor. *Comput Biol Med* 2017 Dec 01;84:114-123. [doi: [10.1016/j.compbiomed.2017.03.020](https://doi.org/10.1016/j.compbiomed.2017.03.020)] [Medline: [28351715](https://pubmed.ncbi.nlm.nih.gov/28351715/)]
25. Rodríguez-Molinero A, Pérez-López C, Samà A, de Mingo E, Rodríguez-Martín D, Hernández-Vara J, et al. A Kinematic Sensor and Algorithm to Detect Motor Fluctuations in Parkinson Disease: Validation Study Under Real Conditions of Use. *JMIR Rehabil Assist Technol* 2018 Apr 25;5(1):e8 [FREE Full text] [doi: [10.2196/rehab.8335](https://doi.org/10.2196/rehab.8335)] [Medline: [29695377](https://pubmed.ncbi.nlm.nih.gov/29695377/)]
26. Brognara L, Palumbo P, Grimm B, Palmerini L. Assessing Gait in Parkinson's Disease Using Wearable Motion Sensors: A Systematic Review. *Diseases* 2019 Feb 05;7(1):18 [FREE Full text] [doi: [10.3390/diseases7010018](https://doi.org/10.3390/diseases7010018)] [Medline: [30764502](https://pubmed.ncbi.nlm.nih.gov/30764502/)]
27. BEAT-PD data release. Brain Commons. URL: <https://www.braincommons.org/beat-pd-data-release/> [accessed 2020-09-16]
28. ELAN. The Language Archive. URL: <https://tla.mpi.nl/tools/tla-tools/elan/> [accessed 2020-09-16]
29. Kim SJ, Koh K, Boyd S, Gorinevsky D. I(1) Trend Filtering. *SIAM Rev* 2009 May;51(2):339-360. [doi: [10.1137/070690274](https://doi.org/10.1137/070690274)]
30. McCamley J, Donati M, Grimpampi E, Mazzà C. An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data. *Gait Posture* 2012 Jun;36(2):316-318. [doi: [10.1016/j.gaitpost.2012.02.019](https://doi.org/10.1016/j.gaitpost.2012.02.019)] [Medline: [22465705](https://pubmed.ncbi.nlm.nih.gov/22465705/)]
31. Zijlstra W, Hof AL. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. *Gait Posture* 2003 Oct;18(2):1-10. [doi: [10.1016/s0966-6362\(02\)00190-x](https://doi.org/10.1016/s0966-6362(02)00190-x)] [Medline: [14654202](https://pubmed.ncbi.nlm.nih.gov/14654202/)]
32. Khandelwal S, Wickström N. Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database. *Gait Posture* 2017 Jan;51:84-90. [doi: [10.1016/j.gaitpost.2016.09.023](https://doi.org/10.1016/j.gaitpost.2016.09.023)] [Medline: [27736735](https://pubmed.ncbi.nlm.nih.gov/27736735/)]
33. Moe-Nilssen R, Helbostad JL. Estimation of gait cycle characteristics by trunk accelerometry. *J Biomech* 2004 Jan;37(1):121-126. [doi: [10.1016/s0021-9290\(03\)00233-1](https://doi.org/10.1016/s0021-9290(03)00233-1)] [Medline: [14672575](https://pubmed.ncbi.nlm.nih.gov/14672575/)]
34. Welch P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoust* 1967 Jun;15(2):70-73. [doi: [10.1109/tau.1967.1161901](https://doi.org/10.1109/tau.1967.1161901)]
35. Antonsson EK, Mann RW. The frequency content of gait. *J Biomech* 1985;18(1):39-47. [doi: [10.1016/0021-9290\(85\)90043-0](https://doi.org/10.1016/0021-9290(85)90043-0)] [Medline: [3980487](https://pubmed.ncbi.nlm.nih.gov/3980487/)]

36. Gelman A, Hill J, Yajima M. Why We (Usually) Don't Have to Worry About Multiple Comparisons. *J Res Educ Effectiveness* 2012 Apr;5(2):189-211. [doi: [10.1080/19345747.2011.618213](https://doi.org/10.1080/19345747.2011.618213)]
37. Silva de Lima AL, Evers LJW, Hahn T, de Vries NM, Daeschler M, Boroojerdi B, et al. Impact of motor fluctuations on real-life gait in Parkinson's patients. *Gait Posture* 2018 May;62:388-394. [doi: [10.1016/j.gaitpost.2018.03.045](https://doi.org/10.1016/j.gaitpost.2018.03.045)] [Medline: [29627498](https://pubmed.ncbi.nlm.nih.gov/29627498/)]
38. Haji Ghassemi N, Hannink J, Martindale CF, Gaßner H, Müller M, Klucken J, et al. Segmentation of Gait Sequences in Sensor-Based Movement Analysis: A Comparison of Methods in Parkinson's Disease. *Sensors (Basel)* 2018 Jan 06;18(1):145 [FREE Full text] [doi: [10.3390/s18010145](https://doi.org/10.3390/s18010145)] [Medline: [29316636](https://pubmed.ncbi.nlm.nih.gov/29316636/)]
39. Mannini A, Rosenberger M, Haskell WL, Sabatini AM, Intille SS. Activity Recognition in Youth Using Single Accelerometer Placed at Wrist or Ankle. *Med Sci Sports Exerc* 2017 Apr;49(4):801-812 [FREE Full text] [doi: [10.1249/MSS.0000000000001144](https://doi.org/10.1249/MSS.0000000000001144)] [Medline: [27820724](https://pubmed.ncbi.nlm.nih.gov/27820724/)]
40. Raykov YP, Evers LJW, Badawy R, Faber MJ, Bloem BR, Claes K, et al. Probabilistic modelling of gait for remote passive monitoring applications. 2018 Presented at: Machine Learning for Health (ML4H) Workshop at NeurIPS; Dec 8, 2018; Montreal, QC, Canada.
41. Morris ME, Iansek R, Matyas TA, Summers JJ. Stride length regulation in Parkinson's disease. Normalization strategies and underlying mechanisms. *Brain* 1996 Apr;119 (Pt 2):551-568. [doi: [10.1093/brain/119.2.551](https://doi.org/10.1093/brain/119.2.551)] [Medline: [8800948](https://pubmed.ncbi.nlm.nih.gov/8800948/)]
42. Rispen SM, van Schooten KS, Pijnappels M, Daffertshofer A, Beek PJ, van Dieën JH. Identification of fall risk predictors in daily life measurements: gait characteristics' reliability and association with self-reported fall history. *Neurorehabil Neural Repair* 2015 Jan;29(1):54-61. [doi: [10.1177/1545968314532031](https://doi.org/10.1177/1545968314532031)] [Medline: [24759809](https://pubmed.ncbi.nlm.nih.gov/24759809/)]
43. Bouchrika I, Nixon MS. Exploratory factor analysis of gait recognition. 2008 Presented at: 8th IEEE International Conference on Automatic Face & Gesture Recognition; Sep 17-19, 2008; Amsterdam, Netherlands. [doi: [10.1109/afgr.2008.4813395](https://doi.org/10.1109/afgr.2008.4813395)]
44. Piau A, Mattek N, Crissey R, Beattie Z, Dodge H, Kaye J. When Will My Patient Fall? Sensor-Based In-Home Walking Speed Identifies Future Falls in Older Adults. *J Gerontol A Biol Sci Med Sci* 2020 Apr 17;75(5):968-973 [FREE Full text] [doi: [10.1093/gerona/glz128](https://doi.org/10.1093/gerona/glz128)] [Medline: [31095283](https://pubmed.ncbi.nlm.nih.gov/31095283/)]
45. Bloem BR, Marks WJ, Silva de Lima AL, Kuijff ML, van Laar T, Jacobs BPF, et al. The Personalized Parkinson Project: examining disease progression through broad biomarkers in early Parkinson's disease. *BMC Neurol* 2019 Jul 17;19(1):160 [FREE Full text] [doi: [10.1186/s12883-019-1394-3](https://doi.org/10.1186/s12883-019-1394-3)] [Medline: [31315608](https://pubmed.ncbi.nlm.nih.gov/31315608/)]

## Abbreviations

**AUC:** area under the receiver operator curve

**BEAT-PD:** Biomarker and Endpoint Assessment to Track Parkinson's Disease

**CV:** cross-validation

**FOG:** freezing of gait

**LASSO:** least absolute shrinkage and selection operator

**MDS-UPDRS:** Movement Disorders Society Unified Parkinson's Disease Rating Scale

**PD:** Parkinson disease

**PSD:** power spectral density

**ROC:** receiver operator curve

**TUG:** timed up and go test

*Edited by G Eysenbach; submitted 02.04.20; peer-reviewed by J van Uem, P Beek, L Omberg, R Schneider; comments to author 22.06.20; revised version received 10.08.20; accepted 21.08.20; published 09.10.20.*

*Please cite as:*

Evers LJW, Raykov YP, Krijthe JH, Silva de Lima AL, Badawy R, Claes K, Heskes TM, Little MA, Meinders MJ, Bloem BR

Real-Life Gait Performance as a Digital Biomarker for Motor Fluctuations: The Parkinson@Home Validation Study

*J Med Internet Res* 2020;22(10):e19068

URL: <https://www.jmir.org/2020/10/e19068>

doi: [10.2196/19068](https://doi.org/10.2196/19068)

PMID: [33034562](https://pubmed.ncbi.nlm.nih.gov/33034562/)

©Luc JW Evers, Yordan P Raykov, Jesse H Krijthe, Ana Lúcia Silva de Lima, Reham Badawy, Kasper Claes, Tom M Heskes, Max A Little, Marjan J Meinders, Bastiaan R Bloem. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 09.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete

bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Digital Pain Mapping and Tracking in Patients With Chronic Pain: Longitudinal Study

Maria Galve Villa<sup>1</sup>, MSc; Thorvaldur S Palsson<sup>2</sup>, PhD; Albert Cid Royo<sup>1</sup>, MSc; Carsten R Bjarkam<sup>3</sup>, PhD; Shellie A Boudreau<sup>1</sup>, PhD

<sup>1</sup>Center for Neuroplasticity and Pain, Center for Sensory Motor Interaction, Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Aalborg, Denmark

<sup>2</sup>Center for Sensory Motor Interaction, Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Aalborg, Denmark

<sup>3</sup>Department of Neurosurgery, Institute of Clinical Medicine, Aalborg University Hospital, Aalborg, Denmark

**Corresponding Author:**

Shellie A Boudreau, PhD

Center for Neuroplasticity and Pain, Center for Sensory Motor Interaction, Department of Health Science and Technology

Faculty of Medicine

Aalborg University

Frederik Bajers Vej 7D

Aalborg, 9200

Denmark

Phone: 45 99409829

Email: [sboudreau@hst.aau.dk](mailto:sboudreau@hst.aau.dk)

## Abstract

**Background:** Digital pain mapping allows for remote and ecological momentary assessment in patients over multiple time points spanning days to months. Frequent ecological assessments may reveal tendencies and fluctuations more clearly and provide insights into the trajectory of a patient's pain.

**Objective:** The primary aim of this study is to remotely map and track the intensity and distribution of pain and discomfort (eg, burning, aching, and tingling) in patients with nonmalignant spinal referred pain over 12 weeks using a web-based app for digital pain mapping. The secondary aim is to explore the barriers of use by determining the differences in clinical and user characteristics between patients with good (regular users) and poor (nonregular users) reporting compliance.

**Methods:** Patients (N=91; n=53 women) with spinal referred pain were recruited using web-based and traditional in-house strategies. Patients were asked to submit weekly digital pain reports for 12 weeks. Each pain report consisted of digital pain drawings on a pseudo-three-dimensional body chart and pain intensity ratings. The pain drawings captured the distribution of pain and discomfort (pain quality descriptors) expressed as the total extent and location. Differences in weekly pain reports were explored using the total extent (pixels), current and usual pain intensity ratings, frequency of quality descriptor selection, and Jaccard similarity index. Validated e-questionnaires were completed at baseline to determine the patients' characteristics (adapted Danish National Spine Register), disability (Oswestry Disability Index and Neck Disability Index), and pain catastrophizing (Pain Catastrophizing Scale) profiles. Barriers of use were assessed at 6 weeks using a health care-related usability and acceptance e-questionnaire and a self-developed technology-specific e-questionnaire to assess the accessibility and ease of access of the pain mapping app. Associations between total extent, pain intensity, disability, and catastrophizing were explored to further understand pain. Differences between regular and nonregular users were assessed to understand the pain mapping app reporting compliance.

**Results:** Fluctuations were identified in pain reports for total extent and pain intensity ratings ( $P<.001$ ). However, quality descriptor selection ( $P=.99$ ) and pain drawing ( $P=.49$ ), compared using the Jaccard index, were similar over time. Interestingly, current pain intensity was greater than usual pain intensity ( $P<.001$ ), suggesting that the timing of pain reporting coincided with a more intense pain experience than usual. Usability and acceptance were similar between regular and nonregular users. Regular users were younger ( $P<.001$ ) and reported a larger total extent of pain than nonregular users ( $P<.001$ ).

**Conclusions:** This is the first study to examine digital reports of pain intensity and distribution in patients with nonmalignant spinal referred pain remotely for a sustained period and barriers of use and compliance using a digital pain mapping app. Differences in age, pain distribution, and current pain intensity may influence reporting behavior and compliance.

(*J Med Internet Res* 2020;22(10):e21475) doi:[10.2196/21475](https://doi.org/10.2196/21475)

**KEYWORDS**

eHealth; medical illustrations; pain perception; mHealth; pain measurement; disease progression; patient-reported outcome measures; musculoskeletal pain; mobile phone; surveys and questionnaires; pain management

## Introduction

### Background

The term *spinal pain* refers to pain in the cervical, thoracic, and low back areas of the spine [1,2]. Spinal pain can remain localized to the spine and refer to other areas. Cervical and thoracic spinal pain may refer to one or both the upper limbs [2], whereas low back spinal pain may refer to one or both of the lower limbs [2]. Chronic spinal pain is a common reason for clinical consultation [3,4] with an increasing prevalence [5,6] and associated high levels of disability [7].

Pain is a subjective sensory and emotional experience [8], and patients' self-reported measures of pain, such as pain intensity and quality descriptors, are common during pain assessment [9]. In addition, self-reports of pain distribution can improve the understanding of pain mechanisms [10-12]. Therefore, more efficient patient-clinician communication is paramount to optimize pain management. Traditionally, numerical rating scales capture pain intensity, whereas pain drawings capture pain distribution that is expressed by area (extent) and location [13].

Among the self-reported methodologies, ecological (from the patients' environment) momentary (in real time) assessment (EMA) methods can help to repeatedly collect information about the patient's pain condition [14,15]. In the clinical pain field, EMA methods can be used to minimize pain recall bias and assess pain within the patients' real-world context [14,16] to provide detailed spatiotemporal information and support pain management [12,17-21]. For example, repeated pain intensity reports of patients with musculoskeletal spinal pain may present a stable or fluctuating temporal pattern [22-27]. Similarly, spatiotemporal patterns of pain may present a stable (localized or widespread) or a variable course [28,29]. However, there is little knowledge about the relationship between the course of pain intensity and extent.

Digital biomarkers are defined as objective and quantifiable physiological and behavioral data acquired and measured using digital devices, such as smartphones, tablets, or computers. Thus, digital pain mapping can acquire time-stamped digital pain biomarkers (pain intensity, distribution, and quality). When acquired over multiple time points, a more detailed overview of a patient's pain experience can be achieved, thereby improving the clinician's understanding of the patient's pain [12].

The feasibility of health apps is mostly dependent on its usability and reporting compliance [30,31]. Usability is a broad term that defines the appropriateness of a technology to fit its purpose [32]. A good compliance rate ensures that the requested information is collected [33-38]. However, little is known about the influence of user characteristics and symptom severity on compliance.

### Objectives

This feasibility study aims to map and track pain intensity (current and usual), distribution (extent and location), and quality of pain and discomfort in patients with nonmalignant spinal referred pain for a duration of 12 weeks using a digital pain mapping app. The secondary aim is to determine the barriers to use and individual and clinical factors influencing a patient's pain reporting compliance using the digital mapping app.

The study aims to address 3 main hypotheses: (1) fluctuations in pain intensity, distribution, and quality over 12 weeks would be identified; (2) pain extent would be associated with current pain intensity and levels of disability; and (3) better reporting compliance would be more likely in patients with more severe pain symptoms.

## Methods

### Overview

This web-based prospective observational cohort study recruited people with nonmalignant spinal referred pain (somatic and radicular) and asked them to submit a weekly pain report for 12 weeks using a web-based pain mapping app. At baseline, all participants completed standardized e-questionnaires about demographics, primary pain site (cervical and thoracic pain or low back pain), disability, and pain catastrophizing. Six weeks into the study, patients were invited to participate in a battery of health care-related usability and acceptance (HUX) e-questionnaire and a self-developed technology-specific e-questionnaire to gain insight into the appropriateness and ease of use of the pain mapping app.

### Participants and Recruitment

The inclusion criteria were kept broad to obtain a general impression of the weekly differences in pain intensity, extent, and quality in patients with nonmalignant spinal referred pain. Men and women (aged 18-85 years) living in Denmark and able to communicate in Danish or English were recruited. Pregnant or breastfeeding women and people with drug abuse and addiction problems were not included. People with a cognitive deficiency, those who lack necessary computer skills (self-assessment or inability to create a password), and/or those who do not have regular internet access were excluded from the study.

In total, 2 recruitment strategies were designed. The traditional strategy consisted of recruiting patients referred to the Neurosurgical Department at Aalborg University Hospital (Denmark). The web-based recruitment strategy consisted of posting a 1-min video on social media platforms (Facebook, LinkedIn, and Twitter) with a call-to-action requesting those interested in taking part in the study to contact the researchers directly via email. Screening was carried out via email correspondence. Additional screening for all patients was carried out by phone interviews, where information about medical



history, diagnoses, and current treatment was collected. The patients' pain management was not monitored or affected during the study.

Patients received detailed information about the study via email, including technical information about registration to the pain mapping app and how to complete the digital pain reports. Subsequently, the researcher (MG) emailed the link to the demographic questionnaire related to their primary pain site (cervical and thoracic or low back). The links to the remaining e-questionnaires (disability and pain catastrophizing) were emailed individually on completion of the previous questionnaire. Each questionnaire took 2 to 3 min to complete.

Furthermore, 6 weeks after the first pain report was submitted, patients completed HUX e-questionnaires, such as the System Usability Scale (SUS) [32] and the modified Technology Acceptance Model (mTAM) questionnaire [39] and a self-developed technology-specific questionnaire.

### Baseline Patient Profile

Baseline e-questionnaires collected information about the patients' general demographics, walking distance ability, health and social information, disability, and pain beliefs. A maximum of 3 reminders were emailed to the patients, each 1 week apart. If a questionnaire was not completed following the 3 reminders, the patient received a final direct email reminder. Questionnaires were delivered using the SurveyXact software (Ramboll). Permission was obtained to use and adapt all the paper-format questionnaires to the electronic versions used in this study.

The general information questionnaire, adapted from the Danish National Spine Register's (DaneSpine) basic information questionnaire for degenerative spinal disorders, was used to collect information regarding demographics, walking distance ability, health, and social status. Disability was measured by the Oswestry Disability Index (ODI) in patients whose primary pain site was low back and by the Neck Disability Index (NDI) in patients whose primary pain site was the cervical or thoracic spine. The ODI and the NDI have been used since 1990 and are recommended tools for evaluating self-rated disability in spinal disorders [40,41]. The Pain Catastrophizing Scale (PCS) was used to evaluate pain-related catastrophizing thoughts. Catastrophizing is a negative anticipatory response associated with higher pain intensity [42,43].

### Quantification of Pain Intensity, Extent, and Distribution Consistency

Navigate Pain (Aglance Solutions) allows for the EMA of the temporal development of pain intensity and extent, thereby facilitating an objective and easy visualization of pain changes over time [44,45]. The web-based digital pain mapping app allows users to complete individual pain reports. The pain mapping app had a zoom feature and the option of moving the body chart on the screen to facilitate the visualization and capturing of pain. In each pain report, users indicated their pain area and location on a pseudo-three-dimensional body chart avatar in different views (anterior, posterior, lateral right, and lateral left) and provided a usual and a current pain intensity rating. Usual pain intensity was defined as the pain felt most of

the time, whereas current pain intensity was defined as the pain experienced at the time of reporting.

Patients received an email with a link to create a password for accessing the pain mapping app free of charge. Patients used a computer mouse or a touch screen device (ie, smartphone or tablet) to report the distribution of pain and discomfort on a male or female body chart. Patients selected among 11 color-coded pain and discomfort quality descriptors: tingling, throbbing, stabbing, dull aching, numbness, itchy, electric, cold, burning, other, and the general descriptor *pain*. The number of pixels was extracted from the drawn areas in the body charts, including the different views, to quantify the total pain and discomfort extent (total extent). Only one pain report was used to determine the extent and distribution of pain and discomfort each week. If patients submitted more than one weekly pain report, the first pain report following the weekly reminder or the report closest to a 7-day interval was selected. Weekly pain intensity ratings were calculated using all reports for each week.

Patients rated their overall usual and current pain intensity using 2 electronic Color Analogue Scales (eCASs). The eCAS is a colored line (green, yellow, and red) accompanied by the words *no pain*, *moderate pain*, and *severe pain* [46,47].

The consistency of pain distribution may assist in a more objective decision-making process. To assess the consistency of pain and discomfort distribution, 1 pain drawing in the posterior view for each of the 12 weeks was extracted from the weekly pain reports. The similarity of degree of distribution among the weekly pain drawings was calculated and expressed using the Jaccard similarity coefficient, also known as the Jaccard index [48]. The Jaccard indices were calculated between consecutive pain drawings (week to week) during the 12-week period. A higher Jaccard index represents a greater pixel overlap between drawings and is a proxy measure to determine the degree of changes in the distribution (ie, combination of the location and area) of pain and discomfort between weekly pain reports [11,12,48]. Owing to the inconsistent use of the anterior and lateral body chart views, only the posterior view was used to calculate the Jaccard index. The posterior view was used in all the weekly pain reports and was found to be the most relevant view to capture changes in spinally referred pain.

### Assessment of Pain and Discomfort Quality Descriptors

Pain and discomfort quality descriptors are clinically useful during the differential diagnosis process, especially for neuropathic pain [49-52]. The frequency of pain and discomfort quality descriptor selection was calculated by normalizing the number of times a quality descriptor was selected. Therefore, a selected descriptor was only accounted for once weekly for each patient.

### Health Care–Related Usability, Acceptance, and Technology-Specific Questionnaires

To explore the barriers of use, patients who had completed at least one pain report at 6 weeks were deemed as users and invited to complete a battery of electronic questionnaires, including HUX questionnaires and a technology-specific questionnaire. The HUX questionnaires consisted of SUS and

mTAM questionnaires to assess the usability and acceptance of the pain mapping app among patients.

The SUS is a simple 10-statement questionnaire evaluating the user experience before any discussion with the researcher. There are alternating positive and negative statements. The version used here replaced the term *the product* to the name of the digital body mapping software *Navigate Pain*. An SUS score over 68 is considered above average [32]. The Technology Acceptance Model (TAM) [39] was modified (mTAM) with the term *digital body chart* to describe the technology. The mTAM consisted of a total of 9 statements—4 of these rated perceived usefulness and 5 rated perceived ease of use. Both questionnaires use a 5-point Likert scale, ranging from 1 for *strongly disagree* to 5 for *strongly agree*.

A Navigate Pain Specific (NPS) questionnaire was developed to examine user behavior, particularly accessibility and ease of access. Patients were asked to answer 5 questions. Of these questions, 2 questions assessed point of access to the mapping app (computer or laptop, tablet, or smartphone) and the access pathway to the pain mapping app (reminder email or direct URL link). One question used a 5-point Likert scale, ranging from 1 for *very helpful* to 5 for *not at all helpful*, to explore the helpfulness of the weekly pain report reminder. The last 2 questions also used a 5-point Likert scale, ranging from 1 for *very easy* to 5 for *very difficult* to assess the registration process and the general patients' perspective of ease of use.

### Differences Between Regular and Nonregular Users

Patient characteristics were assessed in relation to pain reporting compliance. Regular users completed weekly pain reports, whereas nonregular users did not fulfill this criterion. In the NPS questionnaire, nonregular users received an additional question asking for reasons for the lack of regular pain reporting, such as "I didn't have time," "I forgot to do it," "I am in too much pain to do it," "I don't have pain," "I'm not interested in using Navigate Pain any more," and "Other."

### Statistical Analysis

Descriptive statistics were used to describe demographics and baseline characteristics and quality descriptor selection. Data distribution was assessed with histograms and quantile-quantile plots for pain intensity ratings, total extent (pixels), and pain distribution (Jaccard index). For nonparametric data, Friedman tests were used to assess the pain intensity ratings (usual and current) and the total extent between weeks. A Wilcoxon signed-rank test was used to compare the differences between current and usual pain intensity ratings. The consistency of pain distribution was determined by comparing the Jaccard indices

each week using a Friedman test. A logistic multiple regression analysis was carried out to explore the influence of reporting compliance on pain distribution consistency. A chi-square test was used to assess differences in weekly pain and discomfort quality descriptor selection during the 12 weeks.

Spearman correlation was used to determine the associations between the pain intensity ratings and the total extent for each patient and the baseline disability and PCS scores.

Mann-Whitney *U* tests were used to examine the differences in baseline characteristics between regular users and nonregular users, such as age, gender, primary pain site, usual and current pain intensity, total extent, disability, PCS scores, and HUX and NPS questionnaire results. A chi-square test was used to explore recruitment differences between regular and nonregular users. A logistic multiple regression analysis was carried out to explore the probability of better reporting compliance in relation to the severity of the pain symptoms. The assessment of pain severity included current pain intensity ratings, extent of pain quality (posterior view) from the first submitted pain report, and baseline disability (ODI or NDI) scores.

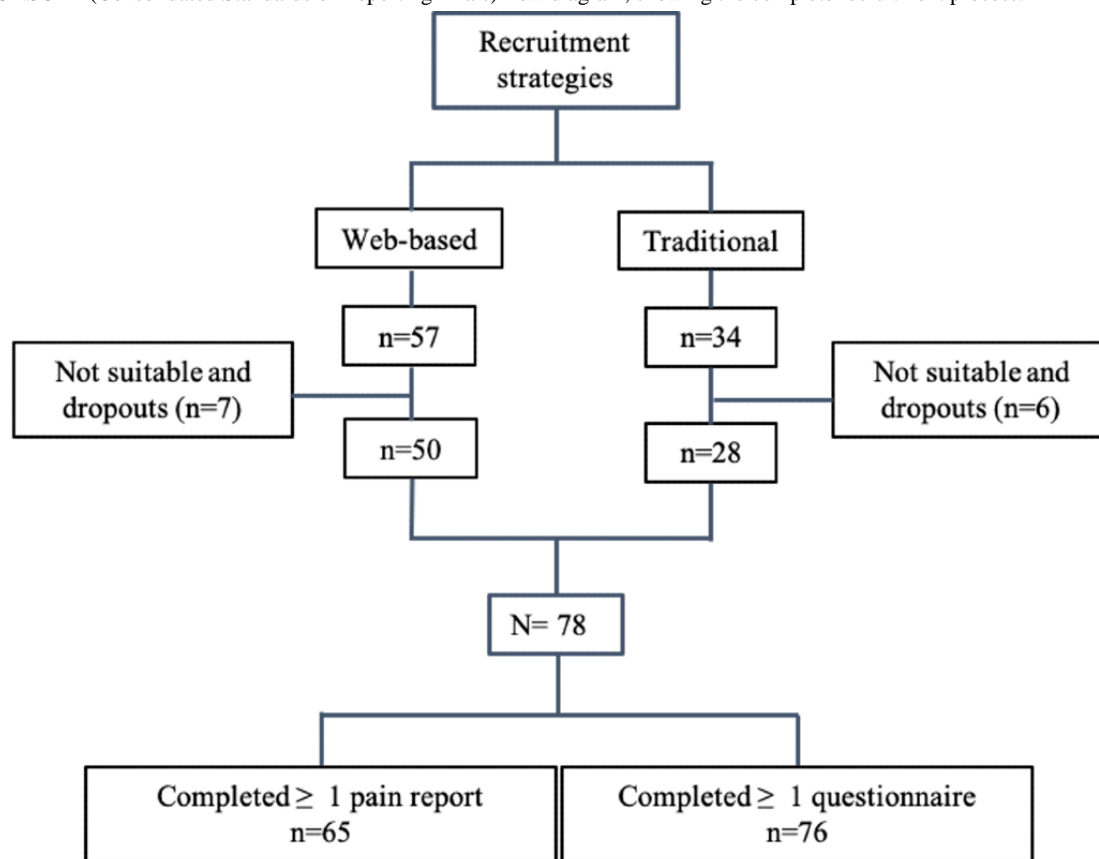
Statistical analyses were performed using SPSS 25 (SPSS Statistics, 2018). The Jaccard indices were calculated using MATLAB R2017b (The MathWorks, Inc). Correlation coefficients, means, SD values, median, and IQR were reported where relevant. *P* values of less than .05 were considered statistically significant, and Bonferroni adjustment was used in multiple analyses.

In Denmark, observational studies using surveys and questionnaires are not required to obtain ethical approval but are required to be registered with the Danish Data Protection Agency (journal numbers 2017-899/10-0159 and 2017-509-00011). This study adhered to ethical data privacy and storage General Data Protection Regulation requirements and was conducted according to the Declaration of Helsinki. The study is registered with ClinicalTrials.gov (NCT03926364).

## Results

### Baseline Patient Profile

A total of 91 patients (mean age 51.8, SD 13.5 years; 53 women) were willing to participate in the study. Recruitment methods included a traditional in-house strategy and a web-based strategy. Following screening and early dropouts, a total of 78 patients were recruited to participate in the study (Figure 1). Overall, 92% (12/13) of the early dropouts were due to technical difficulties with creating a password that limited access to the pain mapping app.

**Figure 1.** CONSORT (Consolidated Standards of Reporting Trials) flow diagram, showing the complete recruitment process.

Overall, 64% (50/78) of the participants were recruited using the web-based strategy. Only an age difference was identified between the patients recruited through the web-based strategy (mean age 48.7, SD 12.1 years) and the traditional in-house strategy (mean age 59.2, SD 13.4 years;  $P<.001$ ).

A total of 3863 pain reports were submitted (mean 59, SD 66, per patient). All the patients recruited for this study had pain for longer than 6 months, with 79% (60/76 completed questionnaires) of patients having pain for over 12 months (Table 1), thus fulfilling the criteria of chronic or persistent pain. Walking ability distance, disability, and catastrophizing scores at baseline are shown in Table 2.

**Table 1.** Self-reported patients' baseline characteristics.<sup>a</sup>

Demographics	Values	
	n (%)	Mean (SD)
Age (years)	76 (100)	51.78 (13)
<b>Gender</b>		
Male	24 (32)	N/A <sup>b</sup>
Female	52 (68)	N/A
<b>Primary pain site</b>		
Cervical and thoracic	17 (22)	N/A
Low back	59 (78)	N/A
BMI (kg/m <sup>2</sup> )	76 (100)	27.69 (5.3)
Regular smokers	10 (12)	N/A
<b>Regular alcohol intake</b>	21 (27)	N/A
Units per week	N/A	8.3 (6.3)
<b>Current social history</b>		
Retired due to spinal pain	19 (25)	N/A
Currently on sick leave	19 (25)	N/A
<b>Pain duration</b>		
Between 3-12 months	16 (21)	N/A
Between 12-24 months	9 (12)	N/A
More than 24 months	51 (67)	N/A
<b>Regular analgesia intake</b>	43 (75)	N/A
Morphine	22 (39)	N/A
<b>Self-reported diagnoses</b>		
Number of patients	61 (80)	N/A
Discus prolapse or protrusion	28 (63)	N/A
Degenerative changes	14 (32)	N/A
Nonspecific	14 (32)	N/A
Spinal stenosis	13 (30)	N/A
Modic changes	10 (23)	N/A
Spondylosis	7 (9)	N/A
Scoliosis	6 (8)	N/A
Spondylolisthesis	2 (3)	N/A
Hypermobility	1 (1)	N/A
Chronic pelvic pain	1 (1)	N/A
<b>Past medical history</b>		
Cardiac condition	4 (5)	N/A
Neurological condition	8 (10)	N/A
Cancer	0 (0)	N/A
Other (painful condition)	20 (26)	N/A
Other (affecting mobility)	17 (22)	N/A
<b>Previous spinal surgery</b>		
Once	11 (15)	N/A

Demographics	Values	
	n (%)	Mean (SD)
Twice	5 (7)	N/A
More than twice	4 (5)	N/A

<sup>a</sup>The self-reported diagnoses include all the diagnoses reported by the patients.

<sup>b</sup>N/A: not applicable.

**Table 2.** Baseline walking distance ability, disability, and catastrophizing scores.

Walking distance and disability	Values	
	n (%)	Mean (SD)
<b>Walking distance ability</b>		
<100 m	9 (12)	N/A <sup>a</sup>
100-150 m	12 (16)	N/A
0.5-1 km	7 (10)	N/A
More than 1 km	48 (62)	N/A
<b>Oswestry Disability Index</b>	46 (78)	35.16 (15.9)
Minimal disability	11 (24)	N/A
Moderate disability	19 (41)	N/A
Severe disability	15 (33)	N/A
Crippled	1 (2)	N/A
<b>Neck Disability Index</b>	19 (100)	34.7 (19.9)
No disability	0 (0)	N/A
Mild disability	3 (16)	N/A
Moderate disability	4 (21)	N/A
Severe disability	3 (16)	N/A
Complete disability	9 (47)	N/A
<b>Catastrophizing beliefs</b>		
<b>Pain Catastrophizing Scale</b>	51 (67)	21.92 (12.3)
Rumination	N/A	7.4 (4.7)
Magnification	N/A	4.11 (3.06)
Helplessness	N/A	10.41 (12.3)
Total score >30	15 (20)	N/A

<sup>a</sup>N/A: not applicable.

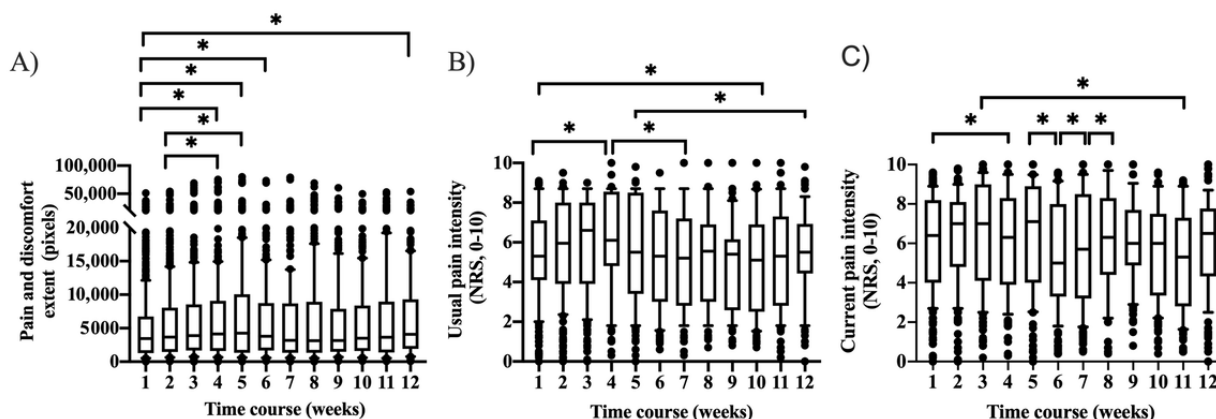
### Quantification of Pain Intensity, Total Extent, and Distribution Consistency

Current pain intensity ratings (median 6.3, IQR 4.5) were greater than usual pain intensity ratings (median 5.4, IQR 4.0;  $Z=-18.0$ ;  $P<.001$ ) when compared over the 12 weeks. Furthermore, fluctuations were identified in terms of usual pain intensity

ratings ( $\chi^2_{11}=145.3$ ;  $P<.001$ ), current pain intensity ratings ( $\chi^2_{11}=105.7$ ; same  $P<.001$ ), and total extent ( $\chi^2_{11}=48.7$ ;  $P<.001$ ) over the 12 weeks (Figure 2). Comparisons of subsequent pain drawings showed similar Jaccard indices ( $P=.52$ ) for the study cohort. The logistic regression model was not statistically significant ( $\chi^2_{11}=8.1$ ;  $P=.70$ ).



**Figure 2.** Median (IQR) pain and discomfort extent (A) and usual (B) and current (C) pain intensity ratings, increased and decreased as a group ( $n=65$ ), in comparison with baseline and throughout the 12 weeks. The lower and upper quartiles, representing observations between the 25th and 75th percentile range, show the median for a month. The whiskers are drawn down to the 10th percentile and up to the 90th. Points below and above the whiskers are drawn as individual dots.  $*P<.001$  adjusted for multiple comparisons. NRS: numerical rating scale.

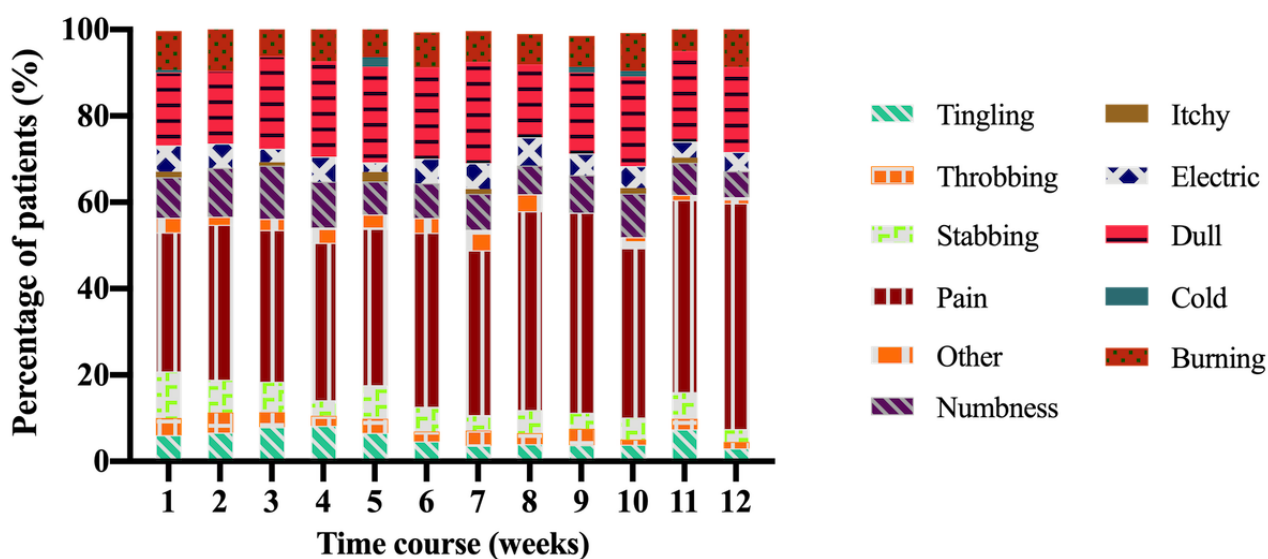


### Frequency of Pain and Discomfort Descriptors

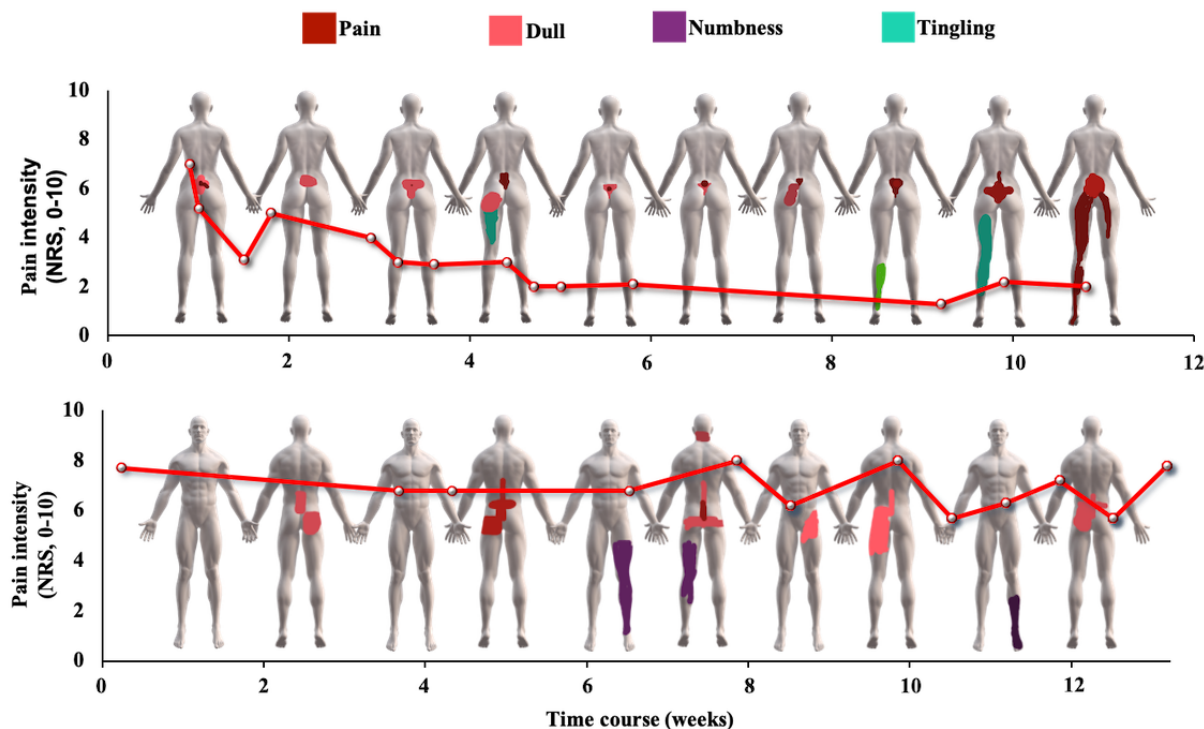
A chi-square test revealed that the selection of quality descriptors remained stable over the 12 weeks ( $P=.99$ ; Figure

3). However, individual variations in pain and discomfort quality descriptor selection were observed (Figure 4).

**Figure 3.** The most frequently selected pain and discomfort quality descriptors over 12 weeks were pain (39%) and dull (20%). The qualities numbness (9%), burning (8%), and stabbing (6%) were also chosen frequently. The least frequently selected quality descriptors were throbbing (3.2%), other (2.8%), itchy (0.8%), and cold (0.4%).



**Figure 4.** Examples of individual pain reports illustrating fluctuations in pain and discomfort intensity, total extent, and quality, spanning 12 weeks. The red line represents the weekly average current pain intensity rating. The pain drawings were selected every second week to capture the general overview of the changes in pain and discomfort quality descriptors selection and the fluctuations in pain and discomfort extent over the 12 weeks. NRS: numerical rating scale.



### Relationship Among Pain Intensity, Total Extent, Disability, and Pain Catastrophizing

Usual and current pain intensities were associated with the total extent ( $r=0.23$  and  $r=0.25$ , respectively;  $R^2=3\%$ ;  $P<.001$ ). Disability, as rated by the ODI and NDI, was not associated with the current and usual pain intensities ( $P=.29$  and  $P=.09$ , respectively), total extent ( $P=.31$ ), and pain catastrophizing ( $P=.83$ ). Similarly, pain catastrophizing was not associated with usual and current pain intensities ( $P=.89$  and  $P=.71$ , respectively) and the total extent ( $P=.29$ ).

### Differences Between Regular and Nonregular Users

Regular users ( $n=35$ ; mean age 48.7, SD 11.2 years) and nonregular users ( $n=30$ ; mean age 55.8, SD 15.3 years) differed in age ( $P<.001$ ). Interestingly, nonregular users had more intense current pain (median 6.4, IQR 3.2) than the regular users (median 6.1, IQR 5.1;  $U=3306.7$ ;  $P<.001$ ). However, regular users had a greater total extent (regular users 4063, IQR 8075.5; nonregular users 3221, IQR 4925.0;  $U=2,775,320.5$ ;  $P<.001$ ). In addition, 80% (28/35) of the regular users were recruited through the web, as opposed to 56% (17/30) of the nonregular ( $\chi^2_2=28.9$ ;  $P<.001$ ). Current pain intensity, pain extent, and disability did not influence reporting compliance between regular and nonregular users ( $P=.96$ ).

### HUX Analyses

HUX questionnaires were completed by 94% (33/35) of the regular and 73% (22/30) of the nonregular users. The total mean SUS score was 70, giving it a marginal usability score [53] with no difference between regular and nonregular users ( $P=.45$ ).

Regular users rated acceptable usability [54] with a mean SUS score of 71.5 points, whereas nonregular users rated marginal usability [53] with a mean SUS score of 68 points. There were no differences in mTAM scoring between the regular and nonregular users, with both groups having similar acceptance scores in each of the questionnaire statements ( $P$  value range for each statement=.12-.97).

Of the regular users, 49% (17/35) accessed the pain mapping app from a computer or laptop and did not need the weekly reminder to submit the pain reports; nonregular users accessed the pain mapping app using a variety of devices, and 82% (20/30) depended on the weekly reminder to submit the pain reports. Furthermore, 45% (10/30) and 41% (9/30) of the nonregular users reported forgetfulness and *other* as the reasons for poor reporting compliance. The remaining 13% of nonregular users reported too much (2/30) or no pain (1/30). None of the patients selected *I am not interested anymore* or *I did not have time*.

Regular users rated the general use of the pain mapping app as easy or very easy (15/35, 45% and 8/30, 27%, respectively). Similarly, nonregular users rated the general use of the pain mapping app as *easy* and *very easy* (11/30, 50% and 4/30, 18%, respectively). Only 6% (6/35) of the regular and 9% (2/30) of the nonregular users rated the general use of the pain mapping app as *difficult*.

## Discussion

### Principal Findings

This is the first pragmatic observational study using a web-based app to map, track, and quantify pain and discomfort remotely in patients with nonmalignant spinal referred pain over a sustained period. The results show fluctuations in current and usual pain intensities over 12 weeks for the group of patients recruited for this study. In addition, the results show fluctuations in pain and discomfort extent (total extent) on a group level. However, the pain drawings and the quality descriptor selection remained consistent on a group level. HUX scores were similar between regular and nonregular users. The regular users were generally younger, had a greater total extent, and accessed the app differently than the nonregular users.

Fluctuations in current and usual pain intensities and in total extent occurred over the 12-week observation period, suggesting that spatiotemporal patterns of chronic spinal referred pain may increase and decrease over time. These fluctuations may be related to the heterogeneity of the study cohort in which participants had different primary pain sites, differences in pain management, and differences in reporting compliance. Thus, there is no rationale for why the group pain intensity and total extent varied on any given week over the 12-week observation period. An interesting observation in this study was that current pain intensity ratings were greater than usual pain intensity ratings, suggesting that patients completed pain reports when their pain was more intense than usual. The reasons for this reporting behavior are unclear and should be further explored. However, the difference between current and usual pain intensity ratings was small ( $<1$  of 10) and may not be clinically meaningful [54].

The selection of pain and discomfort quality descriptors remained consistent over time, which can be expected for a group with chronic pain. Interestingly, pain distribution also remained consistent. In this study, pain distribution consistency was calculated by comparing the similarity of consecutive weekly pain drawings. Therefore, the study's results do not imply that pain distribution was consistent throughout the 12-week study period. It is possible that fluctuations in pain distribution may occur over longer periods ([Multimedia Appendix 1](#)). The Jaccard index has primarily been used to assess the ability of patients to reliably draw and redraw pain areas [48,55]. Therefore, the Jaccard index may only be appropriate to assess distribution consistency over shorter periods.

The study's findings align with previous studies [56-59] showing a weak and positive association between the pain intensities (usual and current) and the total extent. These results suggest that pain extent is related to pain intensity but do not explain the intensity variance. Therefore, capturing changes in clinical pain extent and distribution may add additional value to clinical interpretation.

The relationships of pain intensity and extent with disability and catastrophizing scores are contradictory. Similar to our study, studies on chronic low back pain [60,61] and knee

osteoarthritis [59] have shown that pain intensity and total extent were not associated with disability and pain catastrophizing. However, other studies have shown a positive relationship of the total extent with disability scores [11,62-66] and pain catastrophizing [43,62,67-69]. The relationship between catastrophizing and disability scores has been widely described in musculoskeletal shoulder and low back pain [70-72]. However, patients in our study reported high levels of disability but not high levels of catastrophizing, showing no association between disability and catastrophizing at baseline.

The HUX results demonstrated that the pain mapping app was considered a good pain communication tool by patients and therefore may be relevant for clinical practice. However, the HUX assessment failed to identify compliance differences, as represented by the similar results between regular and nonregular users.

Most of the regular users were recruited using the web-based strategy. Patients recruited through the web were younger (approximately by 3 years) and may already be more technically competent [73] than those recruited using the traditional strategy. A limited number of studies have assessed users' characteristics influencing the use and acceptance of pain technology [74]. Overall, regular users were younger (approximately by 7 years) and reported a larger pain extent than nonregular users, suggesting that reporting compliance of the pain mapping app may be based on an inherent need to communicate the pain extent [75]. Interestingly, the total extent differences between regular and nonregular users (approximately 850 pixels) represents a very small area on a pain drawing and may be insufficient to be clinically relevant. In a short period, the 7-year age gap between regular and nonregular users will become irrelevant as technical savviness increases among older adults [76]. Therefore, it is possible that in the future, only relevant baseline differences influencing compliance may be based on the recruitment method.

### Limitations

EMA can induce bias because of the lack of standardization of pain reporting in terms of context (location and environment), timing, and frequency [15,77]. This study lacks contextual information about, for example, type or change in pain management and activity levels. Therefore, it is not known whether the fluctuations in total extent reporting may be due to a change in experienced pain or a change in the number of pain report submissions over time. Furthermore, fluctuations could also be due to a change in the drawing skills, as the patients' drawing confidence may have increased with repetition or even whether patients used different accessing devices affecting their technical skills.

Technical limitations influenced the total extent results and the distribution consistency index. First, the total extent may be overestimated, as it is the sum of all the quality descriptors used in a pain report. Second, only the Jaccard index was used as a measure of distribution consistency. Using one single index, the Jaccard index, to measure pain distribution consistency, carries potential risks for misinterpretation, as it will miss similarities between 2 pain drawings with equal pain areas but

different locations, representing a small analytical variation, but with clinical implications.

Technical and interpretation barriers of use were identified. Technical barriers caused most of the early dropouts, likely because of using an old version of a browser, and may have biased the usability ratings. Interpretation barriers led to different drawing behaviors despite receiving the same instructions, highlighting different ways of understanding the provided pain reporting instructions or perceiving the individual pain experience ([Multimedia Appendix 2](#)).

### Future Perspectives

The digital pain biomarkers acquired from the app can assist in a more objective diagnostic process and monitor the outcomes following treatment. An example of outcome monitoring could be useful following spinal surgery or a conservative approach, such as exercise therapy, where the pain mapping app can be used to monitor changes in pain intensity and referred arm and leg pain distribution [78]. New metrics for assessing consistency

over a longer period may prove clinically valuable as fluctuating and stable pain reports may require different pain management approaches.

The usefulness and advantages of digital pain mapping to track digital pain biomarkers, combined with machine learning, have already revealed spatiotemporal patterns of pain and discomfort [10]. These patterns have the potential to become a game changer and may be able to predict those patients more likely to respond to treatments or assist in the prognosis of pain conditions [79].

### Conclusions

This is the first study to remotely track pain intensity and distribution and examine barriers of use and compliance using a digital mapping pain app for a sustained period. Differences in age and pain distribution may influence reporting behavior and compliance and recruitment strategies that may play a role in the success of future web-based studies.

### Acknowledgments

The authors thank Aalborg University (Talent Management Programme, 2016) and Tryg Fonden (109647) for providing funding. SB is part of the Center for Neuroplasticity and Pain, which is supported by the Danish National Research Foundation (DNRF121). These funding sources were not involved in the design, data collection, data analysis, and manuscript preparation and/or publication decisions.

### Conflicts of Interest

SB is the codeveloper of the software app Navigate Pain, which was used to collect the pain drawings. SB has company holdings in Aglance Solutions ApS, which licenses a web application of Navigate Pain. Other authors report no conflicts of interest.

#### Multimedia Appendix 1

Jaccard indices for each individual patient (n=59) with nonmalignant chronic spinal referred pain in cervical or thoracic and low back (gray) and the group average (black) showing the degree of similarity between 2 consecutive weekly pain and discomfort drawings over the 12 weeks.

[[PNG File , 164 KB - jmir\\_v22i10e21475\\_app1.png](#)]

#### Multimedia Appendix 2

Example of individual pain drawings showing different drawing behaviors following the delivery of the same instructions.

[[PNG File , 220 KB - jmir\\_v22i10e21475\\_app2.png](#)]

### References

1. Izzo R, Papolizio T, D'Aprile P, Muto M. Spinal pain. *Eur J Radiol* 2015 May;84(5):746-756. [doi: [10.1016/j.ejrad.2015.01.018](#)] [Medline: [25824642](#)]
2. Ferreira ML, de Luca K. Spinal pain and its impact on older people. *Best Pract Res Clin Rheumatol* 2017 Apr;31(2):192-202. [doi: [10.1016/j.berh.2017.08.006](#)] [Medline: [29224696](#)]
3. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC Musculoskelet Disord* 2010 Jul 2;11:144 [FREE Full text] [doi: [10.1186/1471-2474-11-144](#)] [Medline: [20598124](#)]
4. Hurwitz EL, Randhawa K, Yu H, Côté P, Haldeman S. The global spine care initiative: a summary of the global burden of low back and neck pain studies. *Eur Spine J* 2018 Sep;27(Suppl 6):796-801. [doi: [10.1007/s00586-017-5432-9](#)] [Medline: [29480409](#)]
5. Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. *Arthritis Rheum* 2012 Jun;64(6):2028-2037 [FREE Full text] [doi: [10.1002/art.34347](#)] [Medline: [22231424](#)]
6. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. *Lancet* 2018 Jun 9;391(10137):2356-2367. [doi: [10.1016/S0140-6736\(18\)30480-X](#)] [Medline: [29573870](#)]



7. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: [10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)] [Medline: [30496104](https://pubmed.ncbi.nlm.nih.gov/30496104/)]
8. IASP Terminology. IASP. 2017. URL: <https://www.iasp-pain.org/terminology?navItemNumber=576> [accessed 2019-12-10]
9. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005 Jan;113(1-2):9-19. [doi: [10.1016/j.pain.2004.09.012](https://doi.org/10.1016/j.pain.2004.09.012)] [Medline: [15621359](https://pubmed.ncbi.nlm.nih.gov/15621359/)]
10. Boudreau SA, Royo AC, Matthews M, Graven-Nielsen T, Kamavuako EN, Slabaugh G, et al. Distinct patterns of variation in the distribution of knee pain. *Sci Rep* 2018 Nov 8;8(1):16522 [FREE Full text] [doi: [10.1038/s41598-018-34950-2](https://doi.org/10.1038/s41598-018-34950-2)] [Medline: [30410031](https://pubmed.ncbi.nlm.nih.gov/30410031/)]
11. Cruder C, Falla D, Mangili F, Azzimonti L, Araújo LS, Williamon A, et al. Profiling the location and extent of musicians' pain using digital pain drawings. *Pain Pract* 2018 Jan;18(1):53-66 [FREE Full text] [doi: [10.1111/papr.12581](https://doi.org/10.1111/papr.12581)] [Medline: [28466572](https://pubmed.ncbi.nlm.nih.gov/28466572/)]
12. Shaballout N, Aloumar A, Neubert T, Dusch M, Beissner F. Digital pain drawings can improve doctors' understanding of acute pain patients: survey and pain drawing analysis. *JMIR Mhealth Uhealth* 2019 Jan 10;7(1):e11412 [FREE Full text] [doi: [10.2196/11412](https://doi.org/10.2196/11412)] [Medline: [30632970](https://pubmed.ncbi.nlm.nih.gov/30632970/)]
13. Barbero M, Moresi F, Leoni D, Gatti R, Egloff M, Falla D. Test-retest reliability of pain extent and pain location using a novel method for pain drawing analysis. *Eur J Pain* 2015 Sep;19(8):1129-1138. [doi: [10.1002/ejp.636](https://doi.org/10.1002/ejp.636)] [Medline: [25565607](https://pubmed.ncbi.nlm.nih.gov/25565607/)]
14. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008;4:1-32. [doi: [10.1146/annurev.clinpsy.3.022806.091415](https://doi.org/10.1146/annurev.clinpsy.3.022806.091415)] [Medline: [18509902](https://pubmed.ncbi.nlm.nih.gov/18509902/)]
15. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological momentary assessment methodology in chronic pain research: a systematic review. *J Pain* 2018 Jul;19(7):699-716 [FREE Full text] [doi: [10.1016/j.jpain.2018.01.006](https://doi.org/10.1016/j.jpain.2018.01.006)] [Medline: [29371113](https://pubmed.ncbi.nlm.nih.gov/29371113/)]
16. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. *J Med Internet Res* 2019 Feb 5;21(2):e11398 [FREE Full text] [doi: [10.2196/11398](https://doi.org/10.2196/11398)] [Medline: [30720437](https://pubmed.ncbi.nlm.nih.gov/30720437/)]
17. Palmer H. Pain charts; a description of a technique whereby functional pain may be diagnosed from organic pain. *N Z Med J* 1949 Apr;48(264):187-213. [Medline: [18138692](https://pubmed.ncbi.nlm.nih.gov/18138692/)]
18. Toomingas A. Characteristics of pain drawings in the neck-shoulder region among the working population. *Int Arch Occup Environ Health* 1999 Mar;72(2):98-106. [doi: [10.1007/s004200050344](https://doi.org/10.1007/s004200050344)] [Medline: [10197481](https://pubmed.ncbi.nlm.nih.gov/10197481/)]
19. Brismar H, Vucetic N, Svensson O. Pain patterns in lumbar disc hernia Drawings compared to surgical findings in 159 patients. *Acta Orthop Scand* 1996 Oct;67(5):470-472. [doi: [10.3109/17453679608996670](https://doi.org/10.3109/17453679608996670)] [Medline: [8948252](https://pubmed.ncbi.nlm.nih.gov/8948252/)]
20. Mann NH, Brown MD, Enger I. Statistical diagnosis of lumbar spine disorders using computerized patient pain drawings. *Comput Biol Med* 1991;21(6):383-397. [doi: [10.1016/0010-4825\(91\)90040-g](https://doi.org/10.1016/0010-4825(91)90040-g)] [Medline: [1838724](https://pubmed.ncbi.nlm.nih.gov/1838724/)]
21. Ohnmeiss DD, Vanharanta H, Ekholm J. Relation between pain location and disc pathology: a study of pain drawings and CT/discography. *Clin J Pain* 1999 Sep;15(3):210-217. [doi: [10.1097/00002508-199909000-00008](https://doi.org/10.1097/00002508-199909000-00008)] [Medline: [10524474](https://pubmed.ncbi.nlm.nih.gov/10524474/)]
22. Landmark T, Dale O, Romundstad P, Woodhouse A, Kaasa S, Borchgrevink PC. Development and course of chronic pain over 4 years in the general population: the HUNT pain study. *Eur J Pain* 2018 Oct;22(9):1606-1616. [doi: [10.1002/ejp.1243](https://doi.org/10.1002/ejp.1243)] [Medline: [29754398](https://pubmed.ncbi.nlm.nih.gov/29754398/)]
23. Vasseljen O, Woodhouse A, Bjørngaard JH, Leivseth L. Natural course of acute neck and low back pain in the general population: the HUNT study. *Pain* 2013 Aug;154(8):1237-1244. [doi: [10.1016/j.pain.2013.03.032](https://doi.org/10.1016/j.pain.2013.03.032)] [Medline: [23664654](https://pubmed.ncbi.nlm.nih.gov/23664654/)]
24. Artus M, van der Windt D, Jordan KP, Croft PR. The clinical course of low back pain: a meta-analysis comparing outcomes in randomised clinical trials (RCTs) and observational studies. *BMC Musculoskelet Disord* 2014 Mar 7;15:68 [FREE Full text] [doi: [10.1186/1471-2474-15-68](https://doi.org/10.1186/1471-2474-15-68)] [Medline: [24607083](https://pubmed.ncbi.nlm.nih.gov/24607083/)]
25. Vos CJ, Verhagen AP, Passchier J, Koes BW. Clinical course and prognostic factors in acute neck pain: an inception cohort study in general practice. *Pain Med* 2008;9(5):572-580. [doi: [10.1111/j.1526-4637.2008.00456.x](https://doi.org/10.1111/j.1526-4637.2008.00456.x)] [Medline: [18565009](https://pubmed.ncbi.nlm.nih.gov/18565009/)]
26. Enthoven P, Skargren E, Oberg B. Clinical course in patients seeking primary care for back or neck pain: a prospective 5-year follow-up of outcome and health care consumption with subgroup analysis. *Spine (Phila Pa 1976)* 2004 Nov 1;29(21):2458-2465. [doi: [10.1097/01.brs.0000143025.84471.79](https://doi.org/10.1097/01.brs.0000143025.84471.79)] [Medline: [15507811](https://pubmed.ncbi.nlm.nih.gov/15507811/)]
27. Artus M, van der Windt DA, Jordan KP, Hay EM. Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: a systematic review of randomized clinical trials. *Rheumatology (Oxford)* 2010 Dec;49(12):2346-2356. [doi: [10.1093/rheumatology/keq245](https://doi.org/10.1093/rheumatology/keq245)] [Medline: [20713495](https://pubmed.ncbi.nlm.nih.gov/20713495/)]
28. Tesarz J, Gerhardt A, Hartmann M, Kohlmann T, Eich W. The course of the spatial extent of pain in nonspecific chronic back pain: a prospective population-based cohort study with clinical evaluation. *Clin J Pain* 2016 Jul;32(7):580-587. [doi: [10.1097/AJP.0000000000000308](https://doi.org/10.1097/AJP.0000000000000308)] [Medline: [26491937](https://pubmed.ncbi.nlm.nih.gov/26491937/)]
29. Gerhardt A, Hartmann M, Blumenstiel K, Tesarz J, Eich W. The prevalence rate and the role of the spatial extent of pain in nonspecific chronic back pain--a population-based study in the south-west of Germany. *Pain Med* 2014 Jul;15(7):1200-1210. [doi: [10.1111/pme.12286](https://doi.org/10.1111/pme.12286)] [Medline: [24341931](https://pubmed.ncbi.nlm.nih.gov/24341931/)]



30. Suso-Ribera C, Castilla D, Zaragoza I, Ribera-Canudas MV, Botella C, García-Palacios A. Validity, reliability, feasibility, and usefulness of pain monitor: a multidimensional smartphone app for daily monitoring of adults with heterogeneous chronic pain. *Clin J Pain* 2018 Oct;34(10):900-908. [doi: [10.1097/AJP.0000000000000618](https://doi.org/10.1097/AJP.0000000000000618)] [Medline: [29659375](https://pubmed.ncbi.nlm.nih.gov/29659375/)]
31. Aaron LA, Mancil L, Turner JA, Sawchuk CN, Klein KM. Reasons for missing interviews in the daily electronic assessment of pain, mood, and stress. *Pain* 2004 Jun;109(3):389-398. [doi: [10.1016/j.pain.2004.02.014](https://doi.org/10.1016/j.pain.2004.02.014)] [Medline: [15157700](https://pubmed.ncbi.nlm.nih.gov/15157700/)]
32. Brooke J. SUS-a quick and dirty usability scale. In: *Usability Evaluation In Industry*. London, UK: Taylor & Francis Group; Jun 11, 1996.
33. Chaudhry BM. No gain without pain: using pain tracking mobile Apps. *Mhealth* 2016;2:27 [FREE Full text] [doi: [10.21037/mhealth.2016.06.03](https://doi.org/10.21037/mhealth.2016.06.03)] [Medline: [28293601](https://pubmed.ncbi.nlm.nih.gov/28293601/)]
34. Guillory J, Chang P, Henderson CR, Shengelia R, Lama S, Warmington M, et al. Piloting a text message-based social support intervention for patients with chronic pain: establishing feasibility and preliminary efficacy. *Clin J Pain* 2015 Jun;31(6):548-556 [FREE Full text] [doi: [10.1097/AJP.0000000000000193](https://doi.org/10.1097/AJP.0000000000000193)] [Medline: [25565587](https://pubmed.ncbi.nlm.nih.gov/25565587/)]
35. Lalloo C, Kumbhare D, Stinson JN, Henry JL. Pain-QuILT: clinical feasibility of a web-based visual pain assessment tool in adults with chronic pain. *J Med Internet Res* 2014 May 12;16(5):e127 [FREE Full text] [doi: [10.2196/jmir.3292](https://doi.org/10.2196/jmir.3292)] [Medline: [24819478](https://pubmed.ncbi.nlm.nih.gov/24819478/)]
36. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. 'There's a pain app for that': review of patient-targeted smartphone applications for pain management. *Clin J Pain* 2015 Jun;31(6):557-563. [doi: [10.1097/AJP.0000000000000171](https://doi.org/10.1097/AJP.0000000000000171)] [Medline: [25370138](https://pubmed.ncbi.nlm.nih.gov/25370138/)]
37. Barmettler G, Brawn J, Maleki N, Scrivani S, Burstein R, Becerra L, et al. A new electronic diary tool for mapping and tracking spatial and temporal head pain patterns in migraine. *Cephalalgia* 2015 Apr;35(5):417-425. [doi: [10.1177/0333102414545892](https://doi.org/10.1177/0333102414545892)] [Medline: [25143550](https://pubmed.ncbi.nlm.nih.gov/25143550/)]
38. de la Vega R, Roset R, Castarlenas E, Sánchez-Rodríguez E, Solé E, Miró J. Development and testing of painometer: a smartphone app to assess pain intensity. *J Pain* 2014 Oct;15(10):1001-1007. [doi: [10.1016/j.jpain.2014.04.009](https://doi.org/10.1016/j.jpain.2014.04.009)] [Medline: [24854065](https://pubmed.ncbi.nlm.nih.gov/24854065/)]
39. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q* 1989 Sep;13(3):319. [doi: [10.2307/249008](https://doi.org/10.2307/249008)]
40. Vernon H. The neck disability index: state-of-the-art, 1991-2008. *J Manipulative Physiol Ther* 2008 Sep;31(7):491-502. [doi: [10.1016/j.jmpt.2008.08.006](https://doi.org/10.1016/j.jmpt.2008.08.006)] [Medline: [18803999](https://pubmed.ncbi.nlm.nih.gov/18803999/)]
41. Fairbank JC, Pynsent PB. The Oswestry disability index. *Spine (Phila Pa 1976)* 2000 Nov 15;25(22):2940-52; discussion 2952. [doi: [10.1097/00007632-200011150-00017](https://doi.org/10.1097/00007632-200011150-00017)] [Medline: [11074683](https://pubmed.ncbi.nlm.nih.gov/11074683/)]
42. Khan RS, Ahmed K, Blakeway E, Skapinakis P, Nihoyannopoulos L, Macleod K, et al. Catastrophizing: a predictive factor for postoperative pain. *Am J Surg* 2011 Jan;201(1):122-131. [doi: [10.1016/j.amjsurg.2010.02.007](https://doi.org/10.1016/j.amjsurg.2010.02.007)] [Medline: [20832052](https://pubmed.ncbi.nlm.nih.gov/20832052/)]
43. Bierke S, Häner M, Petersen W. Influence of somatization and depressive symptoms on the course of pain within the first year after uncomplicated total knee replacement: a prospective study. *Int Orthop* 2016 Jul;40(7):1353-1360. [doi: [10.1007/s00264-015-3105-z](https://doi.org/10.1007/s00264-015-3105-z)] [Medline: [26820743](https://pubmed.ncbi.nlm.nih.gov/26820743/)]
44. Egsgaard LL, Christensen TS, Petersen IM, Brønnum DS, Boudreau SA. Do gender-specific and high-resolution three dimensional body charts facilitate the communication of pain for women? A quantitative and qualitative study. *JMIR Hum Factors* 2016 Jul 20;3(2):e19 [FREE Full text] [doi: [10.2196/humanfactors.5693](https://doi.org/10.2196/humanfactors.5693)] [Medline: [27440737](https://pubmed.ncbi.nlm.nih.gov/27440737/)]
45. Boudreau SA, Badsberg S, Christensen SW, Egsgaard LL. Digital pain drawings: assessing touch-screen technology and 3D body schemas. *Clin J Pain* 2016 Feb;32(2):139-145. [doi: [10.1097/AJP.0000000000000230](https://doi.org/10.1097/AJP.0000000000000230)] [Medline: [25756558](https://pubmed.ncbi.nlm.nih.gov/25756558/)]
46. Boonstra AM, Stewart RE, Köke AJ, Oosterwijk RF, Swaan JL, Schreurs KM, et al. Cut-off points for mild, moderate, and severe pain on the numeric rating scale for pain in patients with chronic musculoskeletal pain: variability and influence of sex and catastrophizing. *Front Psychol* 2016;7:1466 [FREE Full text] [doi: [10.3389/fpsyg.2016.01466](https://doi.org/10.3389/fpsyg.2016.01466)] [Medline: [27746750](https://pubmed.ncbi.nlm.nih.gov/27746750/)]
47. Tsze DS, Hirschfeld G, Dayan PS, Bulloch B, von Baeyer CL. Defining no pain, mild, moderate, and severe pain based on the faces pain scale-revised and color analog scale in children with acute pain. *Pediatr Emerg Care* 2018 Aug;34(8):537-544 [FREE Full text] [doi: [10.1097/PEC.0000000000000791](https://doi.org/10.1097/PEC.0000000000000791)] [Medline: [27228146](https://pubmed.ncbi.nlm.nih.gov/27228146/)]
48. Boudreau SA, Kamavuako EN, Rathleff MS. Distribution and symmetrical patellofemoral pain patterns as revealed by high-resolution 3D body mapping: a cross-sectional study. *BMC Musculoskelet Disord* 2017 Apr 18;18(1):160 [FREE Full text] [doi: [10.1186/s12891-017-1521-5](https://doi.org/10.1186/s12891-017-1521-5)] [Medline: [28420379](https://pubmed.ncbi.nlm.nih.gov/28420379/)]
49. Baron R, Binder A, Wasner G. Neuropathic pain: diagnosis, pathophysiological mechanisms, and treatment. *Lancet Neurol* 2010 Aug;9(8):807-819. [doi: [10.1016/S1474-4422\(10\)70143-5](https://doi.org/10.1016/S1474-4422(10)70143-5)] [Medline: [20650402](https://pubmed.ncbi.nlm.nih.gov/20650402/)]
50. Bouhassira D, Attal N, Fermanian J, Alchaar H, Gautron M, Masquelier E, et al. Development and validation of the Neuropathic Pain Symptom Inventory. *Pain* 2004 Apr;108(3):248-257. [doi: [10.1016/j.pain.2003.12.024](https://doi.org/10.1016/j.pain.2003.12.024)] [Medline: [15030944](https://pubmed.ncbi.nlm.nih.gov/15030944/)]
51. Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, et al. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain* 2005 Mar;114(1-2):29-36. [doi: [10.1016/j.pain.2004.12.010](https://doi.org/10.1016/j.pain.2004.12.010)] [Medline: [15733628](https://pubmed.ncbi.nlm.nih.gov/15733628/)]
52. van Hecke O, Kamerman PR, Attal N, Baron R, Bjornsdottir G, Bennett DL, et al. Neuropathic pain phenotyping by international consensus (NeuroPPIC) for genetic studies: a NeuPSIG systematic review, Delphi survey, and expert panel

- recommendations. *Pain* 2015 Nov;156(11):2337-2353 [FREE Full text] [doi: [10.1097/j.pain.0000000000000335](https://doi.org/10.1097/j.pain.0000000000000335)] [Medline: [26469320](https://pubmed.ncbi.nlm.nih.gov/26469320/)]
53. Sauro J. 5 Ways to Interpret a SUS Score. MeasuringU. URL: <https://measuringu.com/interpret-sus-score/> [accessed 2019-04-07]
  54. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001 Nov;94(2):149-158. [doi: [10.1016/s0304-3959\(01\)00349-9](https://doi.org/10.1016/s0304-3959(01)00349-9)] [Medline: [11690728](https://pubmed.ncbi.nlm.nih.gov/11690728/)]
  55. Leoni D, Falla D, Heitz C, Capra G, Clijsen R, Egloff M, et al. Test-retest reliability in reporting the pain induced by a pain provocation test: further validation of a novel approach for pain drawing acquisition and analysis. *Pain Pract* 2017 Feb;17(2):176-184. [doi: [10.1111/papr.12429](https://doi.org/10.1111/papr.12429)] [Medline: [26992099](https://pubmed.ncbi.nlm.nih.gov/26992099/)]
  56. Barbero M, Fernández-de-Las-Peñas C, Palacios-Ceña M, Cescon C, Falla D. Pain extent is associated with pain intensity but not with widespread pressure or thermal pain sensitivity in women with fibromyalgia syndrome. *Clin Rheumatol* 2017 Jun;36(6):1427-1432. [doi: [10.1007/s10067-017-3557-1](https://doi.org/10.1007/s10067-017-3557-1)] [Medline: [28161767](https://pubmed.ncbi.nlm.nih.gov/28161767/)]
  57. Miró J, de la Vega R, Tomé-Pires C, Sánchez-Rodríguez E, Castarlenas E, Jensen MP, et al. Pain extent and function in youth with physical disabilities. *J Pain Res* 2017;10:113-120 [FREE Full text] [doi: [10.2147/JPR.S121590](https://doi.org/10.2147/JPR.S121590)] [Medline: [28115871](https://pubmed.ncbi.nlm.nih.gov/28115871/)]
  58. Miró J, Gertz KJ, Carter GT, Jensen MP. Pain location and intensity impacts function in persons with myotonic dystrophy type 1 and facioscapulohumeral dystrophy with chronic pain. *Muscle Nerve* 2014 Jun;49(6):900-905 [FREE Full text] [doi: [10.1002/mus.24079](https://doi.org/10.1002/mus.24079)] [Medline: [24415580](https://pubmed.ncbi.nlm.nih.gov/24415580/)]
  59. Lluch Gribés E, Dueñas L, Barbero M, Falla D, Baert IA, Meeus M, et al. Expanded distribution of pain as a sign of central sensitization in individuals with symptomatic knee osteoarthritis. *Phys Ther* 2016 Aug;96(8):1196-1207. [doi: [10.2522/ptj.20150492](https://doi.org/10.2522/ptj.20150492)] [Medline: [26939604](https://pubmed.ncbi.nlm.nih.gov/26939604/)]
  60. Parker H, Wood PL, Main CJ. The use of the pain drawing as a screening measure to predict psychological distress in chronic low back pain. *Spine (Phila Pa 1976)* 1995 Jan 15;20(2):236-243. [doi: [10.1097/00007632-199501150-00022](https://doi.org/10.1097/00007632-199501150-00022)] [Medline: [7716631](https://pubmed.ncbi.nlm.nih.gov/7716631/)]
  61. Hildebrandt J, Franz CE, Choroba-Mehnen B, Temme M. The use of pain drawings in screening for psychological involvement in complaints of low-back pain. *Spine (Phila Pa 1976)* 1988 Jun;13(6):681-685. [Medline: [2972073](https://pubmed.ncbi.nlm.nih.gov/2972073/)]
  62. Dave AJ, Selzer F, Losina E, Klara KM, Collins JE, Usiskin I, et al. Is there an association between whole-body pain with osteoarthritis-related knee pain, pain catastrophizing, and mental health? *Clin Orthop Relat Res* 2015 Dec;473(12):3894-3902 [FREE Full text] [doi: [10.1007/s11999-015-4575-4](https://doi.org/10.1007/s11999-015-4575-4)] [Medline: [26443774](https://pubmed.ncbi.nlm.nih.gov/26443774/)]
  63. Bortsov AV, Platts-Mills TF, Peak DA, Jones JS, Swor RA, Domeier RM, et al. Pain distribution and predictors of widespread pain in the immediate aftermath of motor vehicle collision. *Eur J Pain* 2013 Sep;17(8):1243-1251 [FREE Full text] [doi: [10.1002/j.1532-2149.2013.00285.x](https://doi.org/10.1002/j.1532-2149.2013.00285.x)] [Medline: [23335385](https://pubmed.ncbi.nlm.nih.gov/23335385/)]
  64. Ris I, Barbero M, Falla D, Larsen MH, Kraft MN, Sjøgaard K, et al. Pain extent is more strongly associated with disability, psychological factors, and neck muscle function in people with non-traumatic versus traumatic chronic neck pain: a cross sectional study. *Eur J Phys Rehabil Med* 2019 Feb;55(1):71-78 [FREE Full text] [doi: [10.23736/S1973-9087.18.04977-8](https://doi.org/10.23736/S1973-9087.18.04977-8)] [Medline: [30156084](https://pubmed.ncbi.nlm.nih.gov/30156084/)]
  65. Prins MR, van der Wurff P, Groen GJ. Chronic low back pain patients with accompanying leg pain: the relationship between pain extent and pain intensity, disability and health status. *J Back Musculoskelet Rehabil* 2013;26(1):55-61. [doi: [10.3233/BMR-2012-00350](https://doi.org/10.3233/BMR-2012-00350)] [Medline: [23411649](https://pubmed.ncbi.nlm.nih.gov/23411649/)]
  66. Ohnmeiss DD, Vanharanta H, Estlander AM, Jämsén A. The relationship of disability (Oswestry) and pain drawings to functional testing. *Eur Spine J* 2000 Jun;9(3):208-212 [FREE Full text] [doi: [10.1007/s005869900118](https://doi.org/10.1007/s005869900118)] [Medline: [10905438](https://pubmed.ncbi.nlm.nih.gov/10905438/)]
  67. Reis F, Guimarães F, Nogueira LC, Meziat-Filho N, Sanchez TA, Wideman T. Association between pain drawing and psychological factors in musculoskeletal chronic pain: A systematic review. *Physiother Theory Pract* 2019 Jun;35(6):533-542. [doi: [10.1080/09593985.2018.1455122](https://doi.org/10.1080/09593985.2018.1455122)] [Medline: [29659315](https://pubmed.ncbi.nlm.nih.gov/29659315/)]
  68. Marshall PW, Schabrun S, Knox MF. Physical activity and the mediating effect of fear, depression, anxiety, and catastrophizing on pain related disability in people with chronic low back pain. *PLoS One* 2017;12(7):e0180788 [FREE Full text] [doi: [10.1371/journal.pone.0180788](https://doi.org/10.1371/journal.pone.0180788)] [Medline: [28686644](https://pubmed.ncbi.nlm.nih.gov/28686644/)]
  69. Willett MJ, Siebertz M, Petzke F, Erlenwein J, Rushton A, Soldini E, et al. The extent of pain is associated with signs of central sensitization in patients with hip osteoarthritis. *Pain Pract* 2020 Mar;20(3):277-288. [doi: [10.1111/papr.12851](https://doi.org/10.1111/papr.12851)] [Medline: [31665822](https://pubmed.ncbi.nlm.nih.gov/31665822/)]
  70. Martinez-Calderon J, Struyf F, Meeus M, Luque-Suarez A. The association between pain beliefs and pain intensity and/or disability in people with shoulder pain: A systematic review. *Musculoskelet Sci Pract* 2018 Oct;37:29-57. [doi: [10.1016/j.msksp.2018.06.010](https://doi.org/10.1016/j.msksp.2018.06.010)] [Medline: [29980139](https://pubmed.ncbi.nlm.nih.gov/29980139/)]
  71. Millere A, Kalnberza-Ribule Z, Mezals M, Nulle A, Millere I, Deklava L. Disability, pain catastrophizing and stress coping of patients with low back pain in rehabilitation practice in Latvia. *J Back Musculoskelet Rehabil* 2020;33(2):323-328. [doi: [10.3233/BMR-170945](https://doi.org/10.3233/BMR-170945)] [Medline: [31524136](https://pubmed.ncbi.nlm.nih.gov/31524136/)]

72. Wertli MM, Eugster R, Held U, Steurer J, Kofmehl R, Weiser S. Catastrophizing-a prognostic factor for outcome in patients with low back pain: a systematic review. *Spine J* 2014 Nov 1;14(11):2639-2657. [doi: [10.1016/j.spinee.2014.03.003](https://doi.org/10.1016/j.spinee.2014.03.003)] [Medline: [24607845](https://pubmed.ncbi.nlm.nih.gov/24607845/)]
73. Nijs J, Inghelbrecht E, Daenen L, Hachimi-Idrissi S, Hens L, Willems B, et al. Recruitment bias in chronic pain research: whiplash as a model. *Clin Rheumatol* 2011 Nov;30(11):1481-1489. [doi: [10.1007/s10067-011-1829-8](https://doi.org/10.1007/s10067-011-1829-8)] [Medline: [21853277](https://pubmed.ncbi.nlm.nih.gov/21853277/)]
74. Ranney ML, Duarte C, Baird J, Patry EJ, Green TC. Correlation of digital health use and chronic pain coping strategies. *Mhealth* 2016;2:35 [FREE Full text] [doi: [10.21037/mhealth.2016.08.05](https://doi.org/10.21037/mhealth.2016.08.05)] [Medline: [28293608](https://pubmed.ncbi.nlm.nih.gov/28293608/)]
75. Becker KL. Cyberhugs: creating a voice for chronic pain sufferers through technology. *Cyberpsychol Behav Soc Netw* 2013 Feb;16(2):123-126. [doi: [10.1089/cyber.2012.0361](https://doi.org/10.1089/cyber.2012.0361)] [Medline: [23276258](https://pubmed.ncbi.nlm.nih.gov/23276258/)]
76. Göransson C, Eriksson I, Ziegert K, Wengström Y, Langius-Eklöf A, Brovall M, et al. Testing an app for reporting health concerns-experiences from older people and home care nurses. *Int J Older People Nurs* 2018 Jun;13(2):e12181 [FREE Full text] [doi: [10.1111/opn.12181](https://doi.org/10.1111/opn.12181)] [Medline: [29210218](https://pubmed.ncbi.nlm.nih.gov/29210218/)]
77. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain* 2003 Jul;104(1-2):343-351. [doi: [10.1016/s0304-3959\(03\)00040-x](https://doi.org/10.1016/s0304-3959(03)00040-x)] [Medline: [12855344](https://pubmed.ncbi.nlm.nih.gov/12855344/)]
78. Verkerk K, Luijsterburg PA, Heymans MW, Ronchetti I, Pool-Goudzwaard AL, Miedema HS, et al. Prognosis and course of pain in patients with chronic non-specific low back pain: a 1-year follow-up cohort study. *Eur J Pain* 2015 Sep;19(8):1101-1110. [doi: [10.1002/ejp.633](https://doi.org/10.1002/ejp.633)] [Medline: [25565501](https://pubmed.ncbi.nlm.nih.gov/25565501/)]
79. Zhang C, Kjellström H, Ek CH, Bertilson BC. Diagnostic Prediction Using Discomfort Drawings with IBTM. In: Proceedings of the 1st Machine Learning for Healthcare Conference. 2016 Jul Presented at: MLR'16; August 19-20, 2016; Los Angeles, CA, USA p. 226-238 URL: <http://proceedings.mlr.press/v56/Zhang16.pdf>

## Abbreviations

**eCAS:** electronic Color Analogue Scale  
**EMA:** ecological momentary assessment  
**HUX:** health care-related usability and acceptance  
**mTAM:** modified Technology Acceptance Model  
**NDI:** Neck Disability Index  
**NPS:** Navigate Pain Specific  
**ODI:** Oswestry Disability Index  
**PCS:** Pain Catastrophizing Scale  
**SUS:** System Usability Scale  
**TAM:** Technology Acceptance Model

*Edited by G Eysenbach; submitted 16.06.20; peer-reviewed by N Shaballout, A Coda; comments to author 23.07.20; revised version received 25.08.20; accepted 31.08.20; published 26.10.20.*

### *Please cite as:*

Galve Villa M, S Palsson T, Cid Royo A, R Bjarkam C, Boudreau SA  
*Digital Pain Mapping and Tracking in Patients With Chronic Pain: Longitudinal Study*  
*J Med Internet Res* 2020;22(10):e21475  
URL: <http://www.jmir.org/2020/10/e21475/>  
doi:[10.2196/21475](https://doi.org/10.2196/21475)  
PMID:[33104012](https://pubmed.ncbi.nlm.nih.gov/33104012/)

©Maria Galve Villa, Thorvaldur S Palsson, Albert Cid Royo, Carsten R Bjarkam, Shellie A Boudreau. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 26.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

---

Publisher:  
JMIR Publications  
130 Queens Quay East.  
Toronto, ON, M5A 3Y5  
Phone: (+1) 416-583-2040  
Email: [support@jmir.org](mailto:support@jmir.org)

---

<https://www.jmirpublications.com/>